Subtitling in immersive media:
A user-centered study

Belén Agulló García

Universitat Autònoma de Barcelona

Departament de Traducció i d'Interpretació i d'Estudis de l'Àsia Oriental
Doctorat en Traducció i Estudis Interculturals

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

Subtitling in immersive media: A user-centered study

PhD Thesis presented by:

Belén Agulló García

Supervised by:

Dr. Anna Matamala

Dr. José Ramón Calvo-Ferrer

2020

Als meus pares, Paqui i Pepe

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisors, Anna Matamala, PhD, and José-Ramón Calvo-Ferrer, PhD. This dissertation would not have been possible without you. Thank you, Anna, for sharing your knowledge, for your constant feedback, for believing in me, for your patience and your guidance during the entire process. Thank you, Ramón, for your invaluable feedback and advice, and for always being there to support and guide me through every crucial career decision. *Gràcies. Gracias*.

I cannot go on without thanking the Immersive Accessibility project, funded by the European Union as part of the H2020 framework (under grant agreement No 761974), and all the consortium members who have provided me with the resources, knowledge, and support to develop this dissertation. I would like to acknowledge Mario Montagud, PhD, Isaac Fraile, Einar Meyerson, Francesc Mas, Jordi Mata, Enric Torres, and Àlex Soler for their help with the different tests within this dissertation. Also, special thanks to the content creators of the stimuli used in the tests: RBB, ARTE and Jaunt VR.

I would like to extend my appreciation to Pilar Orero, PhD, for welcoming me into the Transmedia Catalonia research group and for sharing her inestimable knowledge and experience.

Mis más profundos agradecimientos también van dedicados a la Asociación Provincial de Personas con Discapacidad Auditiva y sus Familias (APANAH) con sede en la ciudad de Elda (Alicante), quienes me prestaron sus instalaciones y su apoyo incondicional para realizar el estudio de recepción con personas con pérdida auditiva. Y, además, con el mejor humor y disposición del mundo. Mil gracias a Anto, Rafa, Vicente, Graciela, Isa y todas las demás. Y, por supuesto, mil gracias a todos los participantes de los tests sin los cuales no solo no hubiera sido posible realizar el estudio, sino que además carecería de sentido.

I would like to acknowledge Sylvia Rothe for contributing her knowledge in relation to subtitles in immersive media, for her feedback during the course of my dissertation, and for sharing her VR head tracking tool (CVR Analyzer) with me. *Danke schön!*

I also wish to take this opportunity to thank Carme Mangiron, PhD, who first inspired me to pursue my career in the games localization industry and then inspired me, once again, to begin my career in academia. *Gràcies, Carme!*

Me gustaría darles las gracias a mis compis del MRA por siempre mostrarme su apoyo y compartir su buen humor durante las interminables horas de trabajo en la oficina. Gracias Anita, Blanca, Dani, Gonzalo, Irene Hermosa, Jara y Joan. Pero mi más sentido agradecimiento va para mi compañera, Irene Tor Carroggio, quien no solo me ha prestado una ayuda incondicional en la tesis aportándome sus conocimientos sobre estadística, sino que además ha sido una amiga inmejorable durante estos años de doctorado. Gràcies, Irene!

Me gustaría también dar las gracias a mi familia y amigos. En especial, quiero agradecerles a mis padres todo el apoyo que me han prestado siempre, sin dudar ni un solo momento de mis decisiones, y sin los cuales no estaría donde estoy hoy. Gràcies, pares!

Muchas gracias, David, por ser mi faro en medio de la tempestad, mis aguas tranquilas a las que siempre volver a fondear.

Contents

Index of Figures and Tables

Index of Figures

Index of Tables

Abbreviation and Acronym Glossary

AD	Audio description
AR	Augmented reality
AS	Accessibility Studies
ASL	American Sign Language
ASR	Automatic speech recognition
AVT	Audiovisual translation
BBC	British Broadcasting Corporation
BITRA	Bibliography on translation and interpreting
CAT	Computer-aided translation
CCMA	Catalan Media Corporation
CFoV	Comfortable field of view
CVR	Cinematic virtual reality
EBU	European Broadcasting Union
FoV	Field of view
H	Hypothesis
HMD	Head-mounted display
HoH	Hard of hearing
ImAc	Immersive Accessibility
IPQ	Igroup Presence Questionnaire
MA	Media accessibility
MT	Machine translation
NYT	New York Times
SDH	Subtitles for the deaf and hard-of-hearing
SL	Sign language
SUS	System Usability Scale
TM	Translation memory
TS	Translation Studies
UCT	User-centered translation
UN	United Nations
VE	Virtual environment
VR	Virtual reality

Chapter 1. Introduction

1.1. Introduction

Audiovisual translation (AVT) has become an established field in Translation Studies (TS), gaining special attention in recent years. Since the 1990s, research studies in AVT have become widespread within TS (Orero, 2004; Díaz-Cintas, 2005; Remael, Orero, & Carroll, 2014; Matamala, 2019). The main AVT modes – namely dubbing (Chaume, 2013) and subtitling (Díaz-Cintas & Remael, 2007) – have been widely studied and analyzed. Other AVT modes such as voice-over (Franco, Matamala, & Orero, 2010), subtitling for the deaf and hard-of-hearing (SDH) (Neves, 2005; Matamala & Orero, 2010; Romero-Fresco, 2015), or audio description (AD) (Maszerowska, Matamala, & Orero 2014; Fryer, 2016) have also emerged in the last few decades. Studies in the AVT field have been descriptive, like some of the abovementioned work, and also experimental, including reception studies (Di Giovanni & Gambier, 2018; Orero et al., 2018).

The main area of study of this dissertation is subtitling. In particular, the key question raised here is how to present subtitles in immersive media such as virtual reality (VR) or 360° content in a suitable way for end users, without disrupting their sense of presence. This is a new area of study within AVT and an underresearched topic that has not been tackled so far. Until today, subtitling has been studied from different angles and fields. For example, psychology (D'Ydewalle, Pollet, & van Rensbergen, 1987; D'Ydewalle, Praet, Verfaille, & van Rensbergen, 1991), engineering (Lambooij, Murdoch, Ijsselsteijn, & Heynderickx, 2013; Hu, Kautz, Yu, & Wang, 2015; Hughes, Armstrong, Jones, & Crabb, 2015) and, of course, AVT (Perego, Del Missier, Porta, & Mosconi, 2010; Szarkowska, Krejtz, Kłyszejko, & Wieczorek, 2011; Szarkowska, Krejtz, Pilipczuk, Dutka, & Kruger, 2016). Within AVT studies in particular, different parameters of subtitling have been addressed including reading speed (Szarkowska & Gerber-Morón, 2018), formal aspects such as characters per line or lines per subtitle, boxes, borders or shadows (Arnáiz-Uzquiza, 2012), line breaks (Perego et al., 2010; Gerber-Morón, Szarkowska, & Woll, 2018), subtitle editing (condensed vs verbatim) (Romero-Fresco, 2009), position of subtitles (Bartoll & Martínez-Tejerina, 2010), and representation of non-speech information (Civera & Orero, 2010).

In 2008, the United Nations (UN) issued the Convention on the Rights of Persons with Disabilities, which is currently ratified by 180 countries (United Nations, n/d). 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 access services, such as SDH or AD, in media content. SDH has become an interesting topic of research for AVT scholars due to its social relevance (Neves, 2005; Matamala & Orero, 2010; Arnáiz-Uzquiza, 2012; Romero-Fresco, 2015).

This dissertation considers subtitling to be a universal form of access to audiovisual content, regardless of the abilities of the viewers, as will be further developed in the theoretical framework (section 1.3). Traditionally, there has been a difference between subtitling (mainly interlingual and aimed at hearing viewers) and SDH (mainly intralingual, including paralinguistic information, and aimed at persons with hearing loss). The industry and academia have emphasized the difference between subtitles for the hearing and subtitles for persons with hearing loss. Therefore, it is not possible to tackle this topic without understanding this distinction between the terms of subtitling and SDH. However, from a theoretical point of view, we aimed at a universal definition of subtitling, as will be further detailed in section 1.3.

Subtitling has been widely researched in its current context. Subtitling studies have been carried out for different media such as television (Jensema, McCann, & Ramsey, 1996; Jensema, 1998; Utray Delgado, Ruiz Mezcua, & Moreiro, 2010; Pedersen, 2011; Cambra, Silvestre, & Leal, 2015), live events (Vervecken, 2012; Oncins, Orero, Lopes, & Serrano, 2013; Eardley-Weaver, 2015), video games (Mangiron, 2013, 2016) and, more recently, video-on-demand (VoD) platforms (Pedersen, 2018). However, both subtitling users and audiovisual media are changing. For example, younger audiences are using subtitles to help them multitask while watching audiovisual content without missing the plot (Ulanoff, 2019). This adds an additional layer of complexity to the already heterogeneous group of subtitling target audiences. Audiovisual content is also being disrupted by new distribution channels. Specifically, VoD platforms have created a new scenario for the distribution of audiovisual content. One can choose what content to watch, where to watch it, on which device (TV set, laptop, tablet, smartphone), with whom (regardless of the language or capabilities of the viewers), in which language, or with which AVT modes (dubbed, subtitled, with SDH, with AD). Advanced customization options for AVT modes offered on new platforms are a sign that viewers' preferences are becoming more diverse, and both the industry and academia need to cater for evolving user needs and viewing behaviors.

The new distribution model introduced by VoD platforms is not the only way in which technology affects how we consume audiovisual content. New formats such as immersive content are changing how audiovisual products are conceived, resulting in more interactive experiences that make viewers feel transported to the virtual world. Immersive media such as VR content have recently emerged and their potential for transforming the way entertainment is consumed has sparked the interest of the industry and researchers (EBU, 2017a; Sheikh, Brown, Watson, & Evans, 2017; Dooley, 2017; Mateer, 2017; Gödde, Gabler, Siegmund, & Braun, 2018). Major technology companies have developed VR hardware such as Sony's PlayStation VR, Facebook's Oculus Rift or HTC's Vive. Nowadays, immersive content such as 360° videos can be easily accessed through basically any smartphone and VR glasses such as Samsung Gear VR, Google Daydream or even Google Cardboard. 360° videos are real images recorded with special camera sets that are then stitched together forming a 360° sphere. These videos can be watched with special VR headsets where the viewer is at the center of the action and can look around to see the different parts of the scene.

In the entertainment industry, VR content can mainly be found in video games. The interactive nature of video games has favored the implementation of VR technology to enhance the gaming experience. According to Statista (2018), VR video gaming sales revenue will grow to 22.9 billion U.S. dollars by the end of 2020. The filmmaking industry has not been so prolific in developing VR experiences, although interest has arisen. Film festivals such as Sundance[1] or Berlinale[2] include 360° short films among their showcase selection. The International Fantastic Film Festival of Sitges has also included a competition dedicated to 360° content[3]. On the broadcasters' side, The New York Times (NYT) and the British Broadcasting Corporation (BBC) have spurred the creation of 360° content to deliver a new narrative in which news is told from a different perspective, as will be further discussed in Chapter 3.

Immersive content developed for these platforms poses challenges for storytelling that have not been explored so far. Traditional storytelling needs to be updated for this

---

[1] https://www.theverge.com/2019/2/7/18201568/sundance-2019-vr-ar-virtual-augmented-reality-new-frontier (consulted on 03.23.2020)
[2] https://www.efm-berlinale.de/en/screenings/vr-screenings/vr-screenings.html (consulted on 03.23.2020)
[3] https://sitgesfilmfestival.com/eng/noticies?id=1003531 (consulted on 03.23.2020)

new medium, and plots are not necessarily linear anymore. According to the European Broadcasting Union (EBU), 360°/immersive content is valuable if it brings added value compared to traditional storytelling (EBU, 2017a). For example, creating a VR experience where the user is exposed to an earthquake or a war situation which could not otherwise be experienced in real life, or stories that bring users to scenarios that they cannot reach, such as a space expedition to Mars. Content creators are still learning how to design engaging and immersive content that enhances the user experience.

Nowadays, immersive content is reaching a wider audience. However, this content has not yet been made accessible and some potential users are being neglected. To fill that gap, the Immersive Accessibility (ImAc)[4] project was born. This dissertation is built in relation to that project, funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement number 761974. The goal for this project is to integrate accessibility in emerging immersive media, such as 360° videos. The project defends the idea that accessibility must be considered from the beginning, when designing a product, instead of being an afterthought. The aim of the ImAc project is, therefore, to provide accessibility services such as SDH, AD, audio subtitling or sign language interpreting for these immersive environments and to ensure that every user, regardless of their age, sex, language or capabilities, can access them. The specific focus of this dissertation is subtitling in immersive environments, particularly 360° videos. The dissertation explores how to include subtitles in immersive content from a user-centered approach.

Public statistics on the VR market and penetration are scarce, although some information can be found online. According to eMarketer[5], VR users in the US would grow to 36.7 million by 2018, meaning a 60 percent increase compared to 2017. No public data are available confirming or rejecting that prediction. Statista (2019) forecasts a significant increase in the augmented reality (AR) and VR market, from 16.8 billion U.S. dollars in 2019 to 160 billion U.S. dollars in 2023. Therefore, trends point to an increase in demand and usage of this technology. Moreover, the applications of this technology transcend media entertainment. It is also used in other industries, such as education (Virvou & Katsionis, 2008; Chen, 2016), health (Haluck & Krummel, 2000; Rothbaum

[4] http://www.imac-project.eu/ (consulted on 03.23.2020)
[5] https://mediakix.com/blog/virtual-reality-statistics-user-demographics-growth/ (consulted on 03.23.2020)

et al., 1999; Freeman et al., 2017), or museums (Wojciechowski, Walczak, White, & Cellary, 2004; Walczak, Cellary, & White, 2006; Jung, tom Dieck, Lee, & Chung, 2016).

According to the World Health Organization, around 466 million people worldwide have hearing loss, rising to more than 900 million by 2050 (World Health Organization, n/d). The social justification for this dissertation is, therefore, clear. The adoption of VR technologies and consequently immersive content is on the rise. A considerable number of users with hearing loss must have the possibility to access this type of content, and subtitles are an instrument that can help to enable that access.

The development of this dissertation has been possible thanks to collaboration with the ImAc project team. Implement subtitles in immersive media is highly technology-dependent at the moment and requires the support of a technical team to develop: a) a subtitle editor with which to create subtitles for this medium; b) a player to display 360° videos with subtitles, and c) 360° videos to be used as stimuli. Therefore, collaboration among different disciplines (content creators, engineers, and AVT researchers) was needed to conduct a reception study testing different subtitling strategies in immersive content with end users. Studies had been carried out on subtitling in immersive media, but only from a technical perspective in the field of engineering (Brown et al., 2017; Rothe, Tran, & Hussman, 2018). These studies explored the technical aspects of implementing subtitles in 360° videos rather than specific subtitling features, such as character identification or directions. It was, therefore, important to undertake this study along with AVT researchers who could provide new insights to achieve an optimal solution for viewers.

This multidisciplinary team brought about a unique opportunity that provided an optimal scenario to carry out reception tests for subtitles in immersive environments. To the best of our knowledge, previous studies on subtitling in immersive media are non-existent in the field of AVT studies. And studies regarding subtitles aimed at persons with hearing loss in immersive media are non-existent in any field. Moreover, a standard solution for subtitles in immersive environments is not available yet. It is, therefore, important to fill that gap by finding a solution for subtitling immersive content from a user-centered perspective. The established standards and guidelines for subtitles must be revisited and updated to comply with the attributes of immersive media and cater to the needs of the users. With that in mind, a user-centric methodology has been applied. End users have been integrated in the development from the beginning, using research tools such as focus groups and iterative usability tests to find the most suitable subtitling solution. A descriptive analysis of state-of-the-art and current commercial solutions has also been carried out to support the study. The methodology will be further discussed in section 1.4.

This dissertation, which is presented as a compendium of publications, has adopted a mixed-method approach. It integrates descriptive studies (including corpus analysis) and experimental design (including reception studies with professional subtitlers, hearing users, and users with hearing loss). This method has contributed to overcoming the hurdles encountered due to the lack of previous studies and standard commercial solutions. Involving users from the beginning has proven to be a powerful tool when trying to implement AVT modes in a new medium. The development and results of this dissertation will be discussed in the following chapters.

1.2. Objectives and hypotheses

The main goals of this dissertation are:

1. To identify challenges and subtitling strategies for immersive content.

In order to achieve the first main goal, the following specific objectives are defined:

1.1. Gathering feedback from subtitle users about their expectations and recommendations regarding the consumption of subtitles in immersive content (360° videos).

1.2. Describing the state of the art of immersive technology in regard to media applications.

2. To identify professional subtitlers' needs for immersive content.

In order to achieve the second main goal, the following specific objectives are defined:

2.2. Gathering feedback from professional subtitlers about their expectations and recommendations regarding the creation of subtitles for immersive content (360° videos).

2.3. Testing a 360° content subtitle editor prototype with professional subtitlers in order to gather their feedback in regard to the needs for subtitling 360° videos.

3. **To evaluate different subtitling strategies in immersive content (360° videos).** Specifically:
   - Comparing levels of presence and preferences among subtitle users (both hearing and with hearing loss) when the position of subtitles is either (a) always visible or (b) fixed position every 120º.
   - Comparing levels of presence and preferences among subtitle users (both hearing and with hearing loss) when directions are transferred using either (a) arrows or (b) a radar.

The following hypotheses (H) have been established in regard to the third main goal:

H1: Users report higher levels of presence when viewing subtitles in always-visible mode compared to fixed-position mode. Always-visible mode is also their preferred option.

H2: Users report higher levels of presence when visualizing directions with arrows compared to radar. Arrows are the preferred option.

Always-visible subtitles are expected to make viewers feel more present in the virtual world and users are thought to prefer this strategy because its behavior is similar to that of subtitles on a traditional screen. Therefore, this type of subtitles proves more intuitive, which is one of the preconditions for usability. Similarly, arrows are expected to achieve higher levels of presence because they are less intrusive to the experience, as they only appear when the speaker is out of sight.

The main goals and subsequent specific objectives are addressed in different articles. Objectives 1.1 and 2.1 are tackled in the article *Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group* (published in *The Journal of Specialised Translation*) presented in Chapter 2. Objective 1.2 is approached in the article *(Sub)titles in cinematic virtual reality: a descriptive study* (to be published in *Onomazéin*) presented in Chapter 3. Objective 2.2 is tackled in the article *Technology for subtitling: a 360-degree turn* (to be published in *Hermeneus*) presented in Chapter 4.

The third main goal is addressed in the article *Subtitles in virtual reality: a reception study* (to be published in *Íkala*) presented in Chapter 5, and is further developed in Chapter 7, along with the conclusions of the dissertation. The structure is summarized in Figure 1.

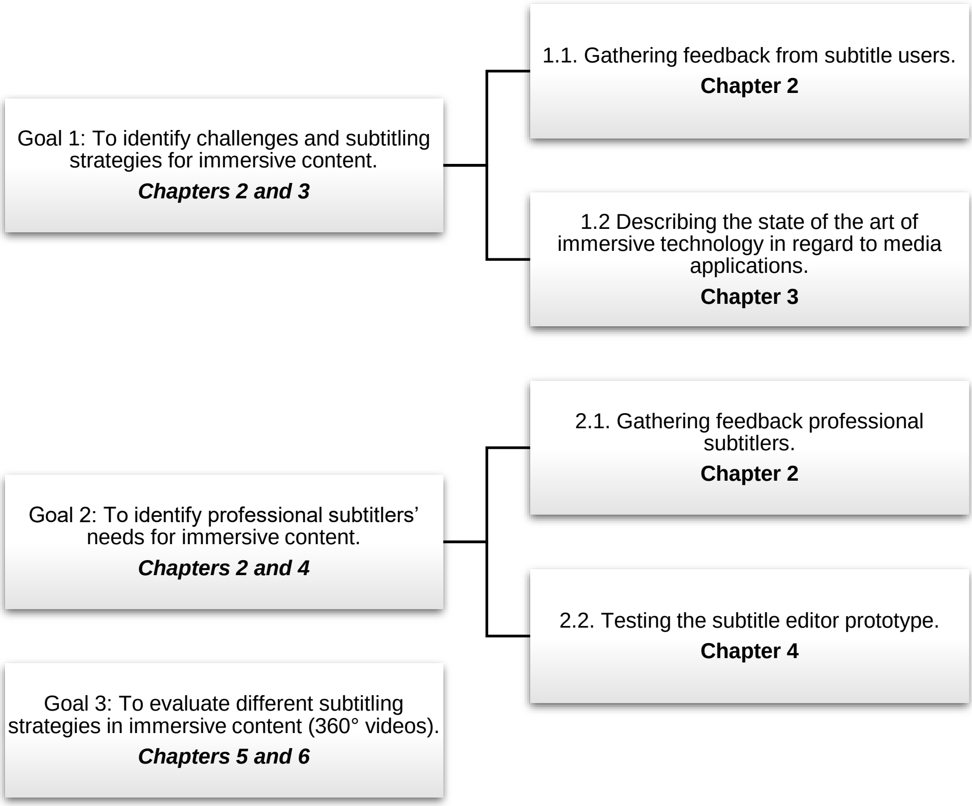


Figure 1. Summary of different objectives and corresponding chapters.

1.3. Theoretical framework

This research is mainly grounded on the concepts and definitions found in AVT studies on subtitling, media accessibility (MA) and SDH. However, the research has adopted a multidisciplinary approach because, due to the nature of the study, it has proven necessary to borrow concepts from outside TS, such as usability and presence. Moreover, this research follows the User-Centered Translation (UCT) model by Suojanen, Koskinen, and Tuominen (2015). According to the authors, "[i]n user-centered translation, information about users is gathered iteratively throughout the process and through different methods, and this information is used to create a usable translation." (Suojanen et al., 2015, p. 5) The user-centered approach adopted for this Ph.D. dissertation is based on the UCT model.

In the following subsections, we will define and place this dissertation within the corresponding theoretical framework for each concept, including subtitling, MA, SDH, the UCT model and presence (a key concept for this study due to the immersive nature of the product under research, namely 360° content).

1.3.1. AVT studies: Subtitling

Subtitling within AVT studies has been a prolific area of research, as explained in section 1.1. Apart from the studies cited in said section, there is also descriptive research on matters such as subtitling theory (Gottlieb, 1994; Guillot, 2018), corpus-based studies (Kalantzi, 2008; Mattsson, 2009; Chong & Wang, 2014; Arias Badia, 2017; Bruti & Zanotti, 2018), or case studies such as representing orality or dialects in subtitling (Guillot, 2008; Longo, 2009; De Meao, 2012), to name but a few. Reception studies (Gottlieb, 1995; Antonini, 2005; Cavaliere, 2008; Caffrey, 2009) have also abounded in this discipline.

Subtitling has been defined in different ways. Depending on where the emphasis lies, its definitions can be classified into three categories: product-oriented, process-oriented and user-oriented. Product-oriented definitions are more common and highlight the formal characteristics of subtitles such as position, number of lines or content. For example, this is the definition by Díaz-Cintas and Remael (2007, p. 8):

> a translation practice that consists of presenting a written text, generally on the lower part of the screen, that endeavours to recount the original dialogue of the speakers, as well as the discursive elements that appear in the image (letters, inserts, graffiti, inscriptions, placards and the like) and the information that is contained on the soundtrack (songs, voices off).

Previously, Gottlieb (1994, pp. 101-102) had also defined subtitling using a product-oriented approach:

> Subtitling is an amphibian: it flows with the current of speech, defining the pace of reception; it jumps at regular intervals, allowing a new text chunk to be read; and flying over the audiovisual landscape, it does not mingle with the human voices of that landscape: instead it provides the audience with a bird's-eye view of the scenery.
>
> […]
>
> Subtitling is an overt type of translation, retaining the original version, thus laying itself bare to criticism from everybody with the slightest knowledge of the source language. At the same time, subtitling is fragmentary in that it only represents the

> lexical and the syntactic features of the dialog. The prosodic features are not truly represented in subtitles: exclamation marks, italics, etc. are only faint echoes of the certain ring that intonation gives the wording of the dialog. Added to this, subtitling has to do without well-known literary and dramatic devices such as stage direction, author's remarks, footnotes, etc. The audience has to turn to the original acoustic and visual clues in trying to grasp the meaning behind the words of the subtitles.

However, in this definition, Gottlieb introduces viewers as an active part of the subtitling practice when he points out that subtitling lays "itself bare to criticism from everybody with the slightest knowledge of the source language." (p. 102) The author also refers to the viewers, explaining that they have to make an active effort to understand the subtitles.

Pérez-González (2014, pp. 15-16) also seeks to define subtitles with a more generalist yet still product-oriented approach: "snippets of written text superimposed on visual footage that convey a target language version of the source speech." This definition is less restrictive than that given by Díaz-Cintas and Remael (2007) (in which they specify that the text is usually located at the bottom of the screen) and leaves room for more creative approaches to subtitling or to different media where there is no bottom of the screen, as is the case with VR content. Voices have previously criticized the constraints of subtitling, defending more creative and unrestricted approaches to this practice (Nornes, 1999; McClarty, 2012).

Process-oriented definitions emphasize the transformation that the product undergoes from its original version to the subtitled version. Such is the case of the definition by Gambier (2003, p. 172):

> Interlingual subtitling (or open caption) involves moving from the oral dialogue to one/two written lines and from one language to another, sometimes to two other languages (bilingual subtitling, as in Finland, Belgium, Israel, etc.). The task can be carried out by the same person (translating and spotting) or by a translator and a technician spotting and timing the subtitles. It can also be done with or without a dialogue list (post-production script). Subtitling does not follow the same process or division of labour everywhere; the responsibility of the translator is not,

> therefore, the same in different countries and different TV broadcasting companies.

User-oriented definitions stress the importance of viewers. In these definitions, references to the users, if not made directly, can be implicitly understood. Karamitroglou (1998, n. p.) refers to the practice of subtitling as follows:

> The general practice of the production and layout of TV subtitles should be guided by the aim to provide maximum appreciation and comprehension of the target film as a whole by maximising the legibility and readability of the inserted subtitled text.

In this definition, the author emphasizes the importance of legibility and readability of the subtitles. These concepts are intrinsically related to usability and, therefore, to users, as will be explained in subsection 1.3.4.

In de Linde and Kay (1999, p. 2), a definition is not found *per se*. However, an attempt to clarify the differences between intralingual and interlingual subtitling can be taken as a way of defining these concepts by contrasting the two. The text reads as follows:

> […] the differences between intralingual and interlingual subtitling stem principally from their target viewers: deaf and hard-of-hearing people and non-native language users, respectively. Their different requirements impose different objectives on the two forms of subtitling. The objective in intralingual subtitling is to substitute the dialogue and other important features of a soundtrack by a written representation, while in interlingual subtitling the objective is to achieve something approaching translation equivalence.

In this explanation, users are the defining concept of subtitling, as well as the objective or function thereof.

The abovementioned definitions reflect the complexity and the multiple factors to be considered when approaching the practice of subtitling from a theoretical point of view. In the context of the present study, we believe that an updated definition is needed in order to cater for new realities, specifically: a) evolving users of subtitling content with different needs and preferences; and b) new media that differ from traditional formats such as TV or cinema. As already mentioned in section 1.1, user preferences are becoming more diverse and the dichotomy found in de Linde and Kay (1999, p. 2) between “deaf and hard-of-hearing people and non-native language users" no longer reflects the reality of subtitle users.

Users are heterogeneous, and they cannot be easily classified. Some viewers with hearing loss whose hearing aids provide them with enough sound quality prefer to watch content without subtitles (as reported by some users in the reception study carried out within this dissertation). Some users with complete hearing loss prefer sign language interpreting rather than subtitles (Arnáiz-Uzquiza, 2012). Also, users with hearing loss can consume interlingual subtitles to watch a film in a different language. Non-native speakers can employ interlingual subtitles, but also intralingual subtitles (allegedly aimed at users with hearing loss) in order to learn how words are written in that language, for example. As discussed in the introduction, young viewers are using subtitles as a support for their multitasking lifestyle (Ulanoff, 2018). Therefore, we purpose to move away from the differentiation between subtitling for hearing viewers and subtitling for viewers with hearing loss because we believe it no longer reflects the reality of the actual users of subtitles. Subtitling is for every user who needs support in understanding audiovisual content.

Audiovisual media and consumption devices have also undergone significant changes in recent years. Viewers now have more options to consume audiovisual content on handheld devices such as laptops, tablets or smartphones, apart from the traditional TV in the living room or the cinema. Moreover, new media such as VR or AR, as well as video games, have changed how users interact with audiovisual content. Therefore, the multimodality linked to audiovisual content now needs to factor in a new aspect: interactivity. Again, the point is that media are ever-changing, and subtitling definitions cannot be restrictive in that sense because they run the risk of becoming outdated.

Based on the above, we suggest the following updated definition for subtitling, mainly inspired by the definitions by Pérez-González (2014) and Karamitroglou (1998) and based on the UCT model (Suojanen et al., 2015):

> 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.

In this definition the emphasis is visibly placed on the users and the usability of the subtitles, rather than prescribing formal aspects of the process of subtitling, which can vary depending on many factors (target audience, industry standards, platforms, media, subtitling tools, etc.). Moreover, specifying the abilities or disabilities of users is avoided. We consider that differentiation to be outdated based on previously stated reasons related to evolving subtitle users. Allusions to the intra- or interlinguistic nature of subtitles are also consciously omitted to reach a more universal, user-centered definition of subtitling.

1.3.2. AVT studies: Media Accessibility

Gambier (2003) addressed the concept of accessibility in AVT studies when describing the main features of screen translation, saying that "[t]he key word in screen translation is now accessibility." (p. 179) Previously in the article, the author introduces the different AVT modes, classifying them as dominant types (mainly subtitling, dubbing, and voice-over, among others) and challenging types (including SDH and AD). The author argues that the concept of accessibility encompasses different features such as acceptability, legibility, readability, synchronicity, relevance, and domestication strategies (Gambier, 2003). Moreover, the author emphasizes the importance of considering viewers when translating audiovisual content and of carrying out reception studies, promoting a user-centered approach to AVT studies.

Díaz-Cintas (2005) defended the idea that AVT modes such as SDH and AD were indeed considered part of TS. The author referred to the translation taxonomy proposed by Jakobson (1959). The three categories defined by Jakobson were intralingual (or rewording), interlingual (or proper translation) and intersemiotic (or transmutation) (Jakobson, 1959). SDH would fall under the category of intralingual, while AD would be considered intersemiotic (Díaz-Cintas, 2005). According to Díaz-Cintas (2005, p. 4), "[a]ccessibility in our field means making an audiovisual programme available to people that otherwise could not have access to it, always bearing in mind a degree of impairment on the part of the receiver." However, the author went further in the definition of accessibility arguing that "in essence, to lip-sync, to subtitle or to voice-over a programme shares as much the idea of accessibility as SDH or AD. Only the intended audiences are different." (Díaz-Cintas, 2005, p. 4) To sum up, he points out that:

> Whether the hurdle is a language or a sensorial barrier, the aim of the translation process is exactly the same: to facilitate the access to an otherwise hermetic source

of information and entertainment. In this way, accessibility becomes a common denominator that underpins these practices. (Díaz-Cintas, 2005, p. 4)

The previous subtitling definition suggested in this Ph.D. dissertation (§1.3.1) includes this principle, as the main goal of subtitling entails "supporting users in understanding any type of audiovisual content." This can be taken as either linguistic or sensorial support.

As worded by Orero and Matamala (2007, p. 269), "[m]edia accessibility is a growing subfield of Audiovisual Translation Studies […]." In the same article, the authors also define accessibility as both linguistic and sensorial, reinforcing previous definitions by Gambier (2003) and Díaz-Cintas (2005). More recently, Greco (2016, p. 23) has defined MA as a "set of theories, practices, services, technologies and instruments providing access to audiovisual content for people that cannot, or cannot properly, access the content in its original form." This definition is based on the definition by Szarkowska, Krejtz, Krejtz, & Duchowski (2013). Greco also points out that the instruments pertaining to MA such as subtitling or AD "stem from or involve audiovisual translation (AVT), because AVT is the field where MA has been developing as a research discipline for the last decade." (Greco, 2016, p. 23) The same author delves deeper and provides three different accounts of MA, as can be seen in Figure 2 (Greco, 2019):

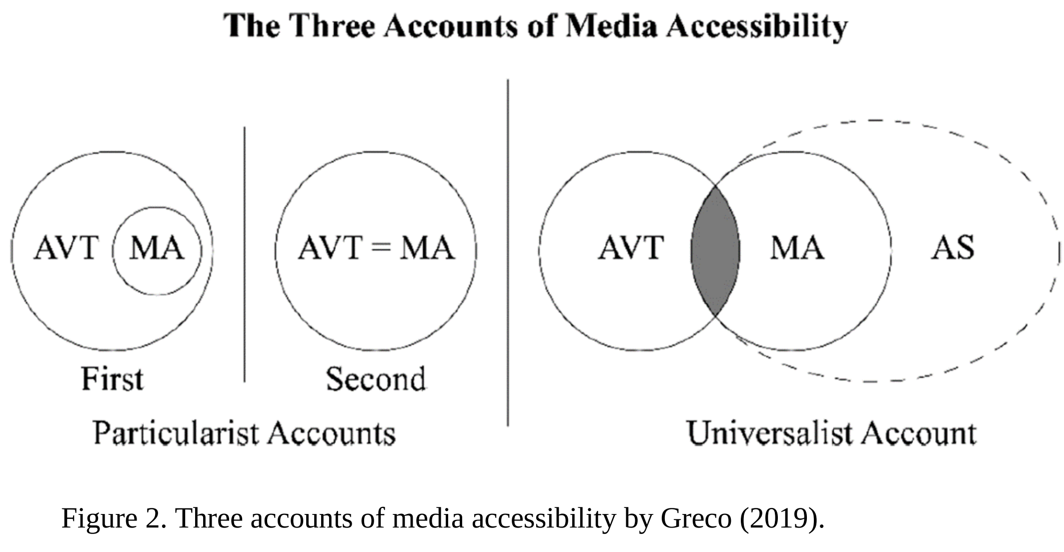


Figure 2. Three accounts of media accessibility by Greco (2019).

According to the author, there are three different accounts to explain the relationship between AVT and MA. In the first account, scholars understood MA as a subfield of AVT that mainly included the study of SDH and AD. These modalities were intended for a specific group of persons with sensorial disabilities. In the second account, scholars argued that MA addressed not only sensorial but also linguistic barriers, but still from a particularist perspective. That is, accessibility is addressed to a specific type of users who need additional services to access audiovisual content. The third account, which is provided by the author himself, is the most recent and presents a universalist approach. The author defines MA as "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." (Greco, 2019) According to the author:

> The universalist definition does not limit MA to any specific group but rather, focuses on the functional processes involved in the interaction between users' specificities, the particular contexts within which they act or are placed, and the means to address those specificities in such contexts.

This is an oversimplification of the different MA accounts, as stated by the author, so we encourage the reader to refer to the source for more details (Greco, 2019).

Matamala (2019) argues that we should allude to the terms of audiovisual accessibility and audiovisual translation. It is important to highlight that we are referring to audiovisual accessibility and not media accessibility, because the specific focus of this study is access to audiovisual content in particular, as opposed to any type of media. The author understands that there are different ways of approaching the same subject (in this case, how to access audiovisual content), and we can look at it from the perspective of either translation or accessibility. But the common denominator is that the subject under research is audiovisual content. Independently, there is an ongoing, healthy and necessary debate regarding the position of MA within the different research fields and it is important to keep communication channels open between the different disciplines.

Scholars do not seem to agree on a single definition of MA as it is a relatively new concept that is still under development. The question posed by this dissertation is how to present subtitles in immersive content without disrupting the experience. We believe that this question belongs to the field of AVT studies given that aesthetic is one of the aspects studied in subtitling (Arnáiz-Uzquiza, 2012). Therefore, our focus of study is a matter of translation rather than accessibility, and so we are framing our study in AVT studies rather than MA studies.

Additionally, we believe that the differentiation between subtitling (allegedly for the hearing) and SDH (allegedly for persons with hearing loss) is outdated. We understand that there are several layers of content in the AVT mode of subtitling, including the translation of the script (either intralinguistc or interlinguistic) and the inclusion or not of extralinguistic features such as character identification, sounds, etc. These options could be customized with the use of technology depending on viewers' needs and preferences, as will be suggested in the next subsection 1.3.3. The attributes that are desirable for each translation mode are based on the viewers' preferences and needs, which can be customizable. Hence the need to apply a user-centered approach in AVT Studies.

1.3.3. AVT studies: SDH

AVT scholars have argued that AVT modes help users overcome barriers when accessing audiovisual content (be they linguistic or sensorial barriers). According to this understanding, subtitling could be aimed at any type of audience. However, both scholars and the industry have traditionally differentiated between subtitles (for hearing viewers) and SDH (for viewers with hearing loss). From a theoretical point of view, we are aiming for a universal definition of subtitling as stated in subsection 1.3.1. However, it would be an oversight to exclude the concept of SDH from the present study, as it has been a key concept in its development.

Usually, as stated before, subtitles for hearing viewers have been referred to as interlinguistic, and SDH as intralinguistic. Gambier (2003, p. 174) defines SDH as follows: "[i]ntralingual subtitling (or closed caption) is done for the benefit of the deaf and hard of hearing (HH), or sometimes to help migrants, for instance, to learn or to improve their command of their new language." In this definition, the author refers to the target audience as a definitory trait of this type of subtitling, as well as stating that they are "intralingual."

Díaz-Cintas (2003, p. 199-200) also refers to this differentiation between subtitling for hearing users and SDH:

> Traditionally, and grosso modo, two types of subtitles have come to the fore: interlingual subtitles, which imply transfer from a source language (SL) to a TL and intralingual subtitles (also known as captioning) where there is no change of language and, as far as television is concerned, where transmission is by means of an independent signal activated by accessing page 888 of teletext. Intralingual subtitling is intended to meet the needs of the deaf and hard of hearing, and

> involves turning the oral content of actors' dialogues into written speech, without loss of all the paratextual information which contributes to the development of the plot or setting the scene, which deaf people are unable to access from the soundtrack, for example telephones ringing, knocks on the door, and the like.

In this definition, Díaz-Cintas once again makes a distinction between interlingual and intralingual subtitles, emphasizing that the target audience of the latter are persons with hearing loss. Additionally, he includes a reference to the importance of any extralinguistic information that is intended to help viewers with hearing loss understand the plot. However, Díaz-Cintas (2003) also points out that this differentiation does not reflect reality. The author argues that this assumption would imply that persons with hearing loss only watch programs produced in their mother tongue, which is certainly not true. According to the author, viewers with hearing loss are therefore forced to consume interlingual subtitles that are not suitable for their needs.

The definition by Pereira (2005) is broader in that sense, and it considers the possibility of interlingual SDH when the author states that SDH “sometimes occurs between languages” (Pereira, 2005, as cited in Utray, Pereira & Orero 2009, p. 249):

> subtitling for the deaf and hard of hearing (SDHH) could be defined as a modality of mode transaction (from oral to writing), which sometimes occurs between languages. It consists of displaying a written text on screen which offers a semantic account of what is happening during a given programme: not only what is said but also how it is said (emphasis, tone of voice, accents, foreign languages, different noises, etc.) and who says it. It includes contextual sounds such as environmental music and noises, and also the numerous discursive elements which appear on the screen (letters, words on posters, etc.).

The topic of interlingual SDH has been further studied by Szarkowska (2013), contributing to the definition of SDH. The author defines this concept as follows:

> While intralingual SDH is sometimes thought to consist merely in transcribing a text and supplementing it with information on sounds and speakers, interlingual subtitling consists in actually translating a text. Interlingual SDH would thus combine translation skills with the knowledge of how to cater for the needs of hearing-impaired spectators. This points to the process of merging of the two well-

> established AVT modalities − regular interlingual subtitling and intralingual SDH − and the emergence of a new one: interlingual SDH. (Szarkowska, 2013, p. 69)

In this study, the author presents the differences between interlingual subtitles, intralingual SDH and interlingual SDH (in English and Polish) based on a limited descriptive study, and encourages the AVT scholar community to further research this topic. Differences are encountered not just in the form (for example, SDH includes sound representation or character identification) but also in the content (SDH includes more information than interlingual subtitles).

More recently, Neves (2018b, p. 82) defined SDH as "[s]uch subtitles are designed for people with hearing impairment because, in addition to rendering speech, they identify speakers and provide extra information about sound effects and music." The author introduces the concepts of "enriched content" and "responsive design," both of which are common in digital environments, in relation to subtitling. Neves (2018b, p. 83) states that the new terms:

> […] speak for a convergent and user-centred reality in which subtitles are part of a complex yet flexible multi-layered media landscape. 'Enriched' speaks for all the added elements that make subtitles relevant to specific users and 'responsive' for the standardized properties enabling subtitles to travel across platforms and media. The term also accounts for the growing interaction between the person and technology, at both ends of the process: production and reception.

Neves' user-centered approach to subtitling is in line with this Ph.D. dissertation. We consider the term "enriched responsive subtitling" to be adequate in the context of this study because it allows for new subtitling features that are present in VR environments. Interaction with VR content leads to new levels of enriched content, for example, the need to include directions in subtitles to guide the users to the speakers or sounds. This new element is not present in previous definitions. Moreover, the author delves deeper into the concept of user-centeredness arguing that:

> 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. (Neves, 2018b, p. 91)

Although the main criteria could be further discussed, we agree with Neves in the definition of subtitling as a 'user experience' that needs to be fulfilled. Neves (2018b) states that in order to achieve better-quality subtitles, it is necessary to better understand the profile and needs of the audience.

So far, there are two common denominators in all SDH definitions: the target audience and additional information. Regarding the target audience, in the abovementioned definitions, some authors have referred to SDH recipients as the deaf and hard of hearing (Gambier, 2003; Díaz-Cintas, 2003; Pereira, 2005), hearing-impaired spectators (Szarkowska, 2013), or people with hearing impairment (Neves, 2018b). Some AVT scholars have tackled the topic of describing and identifying the different groups of people with hearing loss (Neves, 2005; Báez Montero & Fernández Soneira, 2010). One of the issues highlighted is a lack of homogeneity among recipients (Neves, 2008). Neves (2008, p. 131) identified several groups among SDH users:

- deaf and hard of hearing viewers;
- pre-lingually and post-lingually deaf;
- oralising and signing deaf;
- deaf who feel they belong to hearing majority social group and (capital D) Deaf who assume themselves as a linguistic minority;
- deaf for whom the written text is a second language;
- deafened viewers who have residual hearing and/or hearing memory.

Regarding the additional information found in SDH, the cited authors identified: "paratextual information […] for example telephones ringing, knocks on the door, and the like" (Díaz-Cintas, 2003, pp. 199-200), "not only what is said but also how it is said (emphasis, tone of voice, accents, foreign languages, different noises, etc.) and who says it" (Pereira, 2005, as cited in Utray et al., 2009, p. 249), "contextual sounds such as environmental music and noises, and also the numerous discursive elements which appear on the screen (letters, words on posters, etc.)" (Pereira, 2005, as cited in Utray et al., 2009, p. 249), and "they [SDH] identify speakers and provide extra information about sound effects and music" (Neves, 2018b, p. 83).

For a more systematic analysis of SDH features, we adopted the taxonomy proposed by Arnáiz-Uzquiza (2012) for SDH, based on the subtitling taxonomy by Bartoll (2008). The author proposes six categories (Arnáiz-Uzquiza, 2012):

- Linguistic: related to language and density.
- Pragmatic: related to author, recipient, moment of creation, and intention.
- Aesthetic: related to location, position, font, color, and justification.
- Technical: related to method of creation, format, storage, distribution, and medium.
- Aesthetic-technical: related to speed, integration, and optionality.
- Extralinguistic sound parameters: related to all non-verbal sound information included in the audiovisual work, such as character identification, paralinguistic information, information on sound effects, and musical information.

Taxonomies are a useful tool for analysis, especially in descriptive studies (see Chapter 3). The distinction between subtitling for the hearing and subtitling for persons with hearing loss has been imposed by the industry and academia. However, from a theoretical point of view, we understand there should be no differences between subtitling and SDH, and the user-centered definition suggested in subsection 1.3.2 also proves valid when referring to SDH:

> 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 and extralinguistic aspects of the content for its understanding) that are legible, readable, comprehensible and accessible for the intended users.

This user-centered approach to subtitling can be enabled by new technologies and the customization possibilities they bring. The media industry could shift towards a highly customizable subtitling approach where users can choose between subtitles with script only, or script plus character identification, or script plus sound information, etc. The possibilities would be endless, and subtitling would become a unique product with different layers that could be activated or deactivated depending on the users' preferences and needs. Returning to the concept of "enriched responsive subtitling," Neves (2018b, p. 92) concludes that:

> As user-centred technological environments become ever more ubiquitous, viewers will be able to choose specific formats that suit their personal needs. The traditional disability-oriented 'subtitling for the deaf and hard of hearing' paradigm will thus shift towards a more encompassing framework characterized by the use of 'enriched (responsive) subtitles'. Adopting a non-discriminating

> terminology will signal a greater degree of respect for diversity, consistent with the spirit of the 2001 UNESCO Universal Declaration on Cultural Diversity and the 2006 UN Convention on the Rights of Persons with Disabilities. This should not mark 'the end' of SDH, but rather the incorporation of SDH standards as a subtitling variety to be made available to every viewer on demand.

According to the author, we should move away from the distinction between subtitles for the hearing and subtitles for persons with hearing loss towards a more user-centered approach to subtitling, which can only be enabled thanks to technological advances. In our study, we have taken a similar approach based on a user-centered idea of subtitling, which will be put into context in the following section discussing the UCT model.

1.3.4. User-Centered Translation (UCT) model

The UCT model by Suojanen et al. (2015) is an approach to translation theory and practice based on a user-centered methodology. The authors argue that the UCT model can be considered as both a translation theory and a practical model and it is aimed at encouraging interaction between scholars and professionals. According to the authors: "[i]n user-centered translation, information about users is gathered iteratively throughout the process and through different methods, and this information is used to create a usable translation" (Suojanen et al., 2015, p. 4). The motivation of this dissertation is to provide subtitles in immersive content. A user-centered approach was considered the best solution to overcome the obstacles caused by the lack of standard solutions and previous studies in this field. The users were at the center of the study and different research tools were employed to gather iterative feedback such as focus groups, usability testing, and reception studies, as will be explained in section 1.4. The authors of the theory illustrate the UCT process as shown in Figure 3:

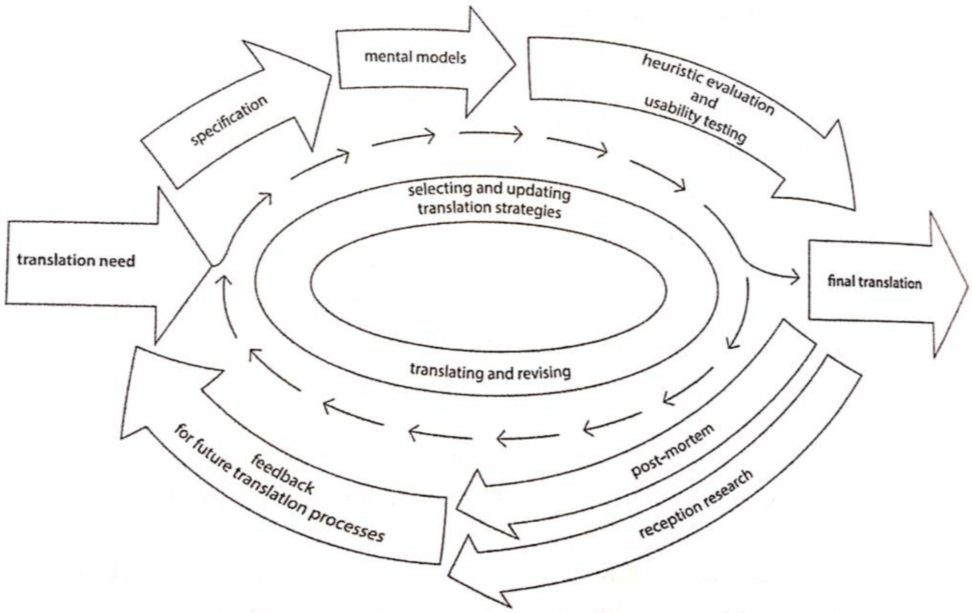


Figure 3. The UCT translation process (Suojanen et al., 2015, p. 4).

The above diagram shows the different elements that are part of the UCT model (Suojanen et al., 2015, pp. 4-6). In Figure 4, the UCT process diagram has been adapted for this dissertation, replacing translation with subtitling.

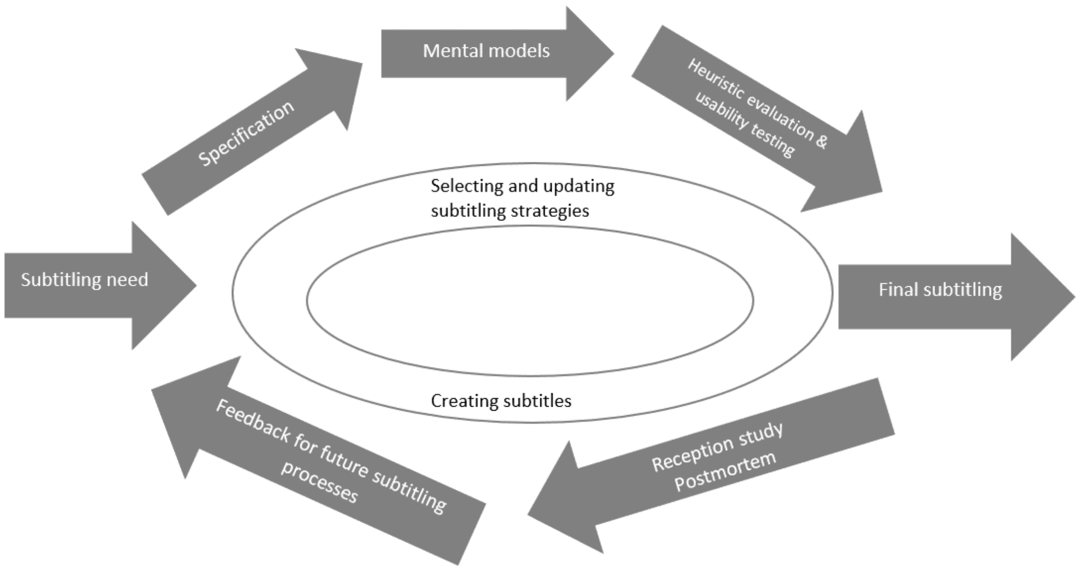


Figure 4. UCT process adapted to subtitling based on original UCT diagram (Suojanen et al., 2015, p. 4).

For this dissertation, we have adapted the UCT model to the specific process of subtitling, as explained in Table 1.

Element	UCT definition
Inner circle: subtitling strategies and creation	This refers to the subtitling process and the strategies that are applied and iteratively evaluated in order to deliver a usable solution for users.
Subtitling need	The communicative need behind the subtitling and the users' requirements and expectations.
Specification	The goals of the subtitled content for the different stakeholders.
Mental models	Models for profiling the users of the subtitled content.
Heuristic evaluation and usability testing	In UCT, it is encouraged to iteratively test the validity of the subtitled content. To that end, there are methods such as heuristic evaluation with a group of experts and usability testing with real end users.
Post-mortem	After the subtitling process is finished, the team analyzes the output and gathers feedback to fine-tune the next project cycles.
Reception research	After the subtitling process, it is advisable to test the end result to understand user preferences and reactions to the chosen solutions. The information is then used to improve the next project cycles.

Table 1. UCT subtitling process.

The UCT model defines usability as "the ease of use of a product in a specified context of use; users are able to use a product effectively, efficiently and to their satisfaction" (Suojanen et al., 2015, p. 13). The authors further define the concept of usability explaining the system acceptability model established by Nielsen (1993) as shown in Figure 5.

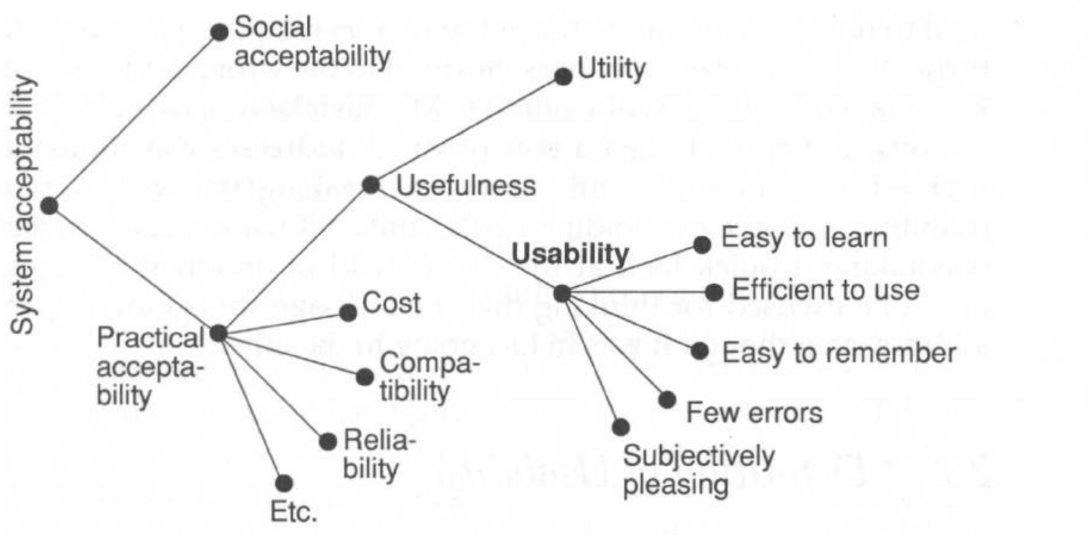


Figure 5. Model of attributes of system acceptability (Nielsen, 1993, p. 25).

According to Nielsen (1993, p. 26), usability has five attributes:

- *Learnability*: The system should be easy to learn so that the user can rapidly start getting some work done with the system.
- *Efficiency*: The system should be efficient to use, so that once the user has learned the system, a high level of productivity is possible.
- *Memorability*: The system should be easy to remember, so that the casual user is able to return to the system after some period of not having used it, without having to learn everything all over again.
- *Errors*: The system should have a low error rate, so that users make few errors during the use of the system, and so that if they do make errors they can easily recover from them. Further, catastrophic must not occur.
- *Satisfaction*: The system should be pleasant to use, so that users are subjectively satisfied when using it; they like it.

Another element that is especially relevant when designing usable subtitles is intuition because the users' preferences can be influenced by habit (Bartoll & Martínez-Tejerina, 2010). The authors of the UCT model define intuition as "our familiarity with something in light of our earlier world of experience with it. If a device that we encounter reminds us of devices we already know, it is intuitive to us and we are able to use it" (Suajonen et al., 2015, p. 16). This element came up when gathering feedback from users during focus groups when some of the users said that they would like to see subtitles in 360° content "like on TV."

These attributes should all be in mind when designing a usable product. However, when subtitles are discussed, we do not only consider the usability of the product (how it is presented in the medium) but also the textual elements of usability. This is also tackled in the UCT model, where the authors identify four elements: legibility, readability, comprehensibility, and accessibility. These concepts are defined as follows (Suajonen et al., 2015, p. 48):

- The legibility of a text means that the text is visible, and that the characters, words and lines are discernible and the text decipherable.
- The readability of a text is affected by stylistic factors such as word choice, argumentation structure, the rhythm of the text, smooth transitions between text parts, active versus passive voice and reader-orientation.

- Comprehensibility refers to how well a text fulfills its communicative function – that is, how well readers understand the text.
- Accessibility refers to efforts making products and services available to all people, regardless of their individual abilities or backgrounds.

In our reception study, the focus was placed on legibility (can subtitles be properly read?) and accessibility (is the content accessible for a diverse group of users?). The purpose was to provide a subtitling solution for immersive content that was both usable and accepted by users. We are currently at a very early stage of research on subtitling in immersive content, and readability and comprehensibility were not yet part of our study; after all, to measure such parameters, first it was necessary to provide subtitles that could be read and accessed by the target audience. According to the UCT model authors "[o]ne key consideration in enhancing accessibility services is usability, and providing accessibility is thus an area where user-centered thinking is needed." (Suajonen et al., 2015, p. 58) Therefore, even though accessibility and usability are different concepts, they are intrinsically related and the key to both is their user-centered approach. One last element was considered necessary for subtitles in 360° content to be successful, which is commonly known as immersion (but for this paper we are using the more accurate concept of presence). Subtitles that are created for this type of content should not disrupt the immersive experience of users. The concept of presence is further discussed in the next subsection.

1.3.5. Presence

The concept of presence has been used to describe the feeling experienced in a virtual environment (VE). Feeling present is not common in traditional media such as TV or cinema. However, VR or 3D games "seem to be a reliable source of this experience" (Schubert, Friedmann, & Regenbrecht, 2001, p. 266). The study of this presence has attracted the attention of many scholars in an effort to understand what happens in the human brain when faced with a VR experience (Heeter, 1992; Steuer, 1992; Lombard & Ditton, 1997; Slater & Wilbur, 1997; Ijsselsteijn, de Ridder, Freeman, & Avons, 2000; Lee, 2004). Reeves (1991) put it as the feeling of being there, an expression that has been repeated over the years to describe the sense of presence in a VE. Lombard and Ditton (1997, n. p.) define presence as "the perceptual illusion of nonmediation." The authors state that this illusion "occurs when a person fails to perceive or acknowledge the existence of a medium in his/her communication environment and responds as he/she would if the medium were not there." (Lombard & Ditton, 1997, n. p.)

Steuer (1992) criticized early conceptions of VR that defined it only from a technological point of view. According to Steuer (1992, p. 73):

> a device-driven definition of virtual reality is unacceptable: It fails to provide any insight into the processes or effects of using these systems, fails to provide a conceptual framework from which to make regulatory decisions, fails to provide an aesthetic from which to create media products and fails to provide a method for consumers to rely on their previous experiences with other media in understanding the nature of virtual reality.

The author then proposes adopting the concept of presence in theorizing VR. He defines presence as "the sense of being in an environment" (Steuer, 1992, p. 75) based on the theories by Gibson (1979). Steuer differentiates between presence (unmediated, natural, the immediate physical surroundings) and telepresence (mediated by a communication technology). In Steuer's words (1992, p. 76): "Telepresence is defined as the experience of presence in an environment by means of a communication medium. In other words, presence refers to the natural perception of an environment, and telepresence refers to the mediated perception of an environment."

This early definition by Steuer opts for the word "telepresence" to refer to the sense of being there in a VE. However, "presence" is a more widespread term throughout the literature, and this is the term that has been adopted for this dissertation (Ijsselsteijn et al., 2000; Lessiter, Freeman, Keogh, & Davidoff, 2001; Schubert, 2003; Lee, 2004). In fact, Lessiter et al. (2001) argue that the study of presence as a concept should be limited to the study of experiences in mediated environments and that real-world experiences can be measured using other constructs such as attention or arousal, for example.

Following the discussion on VR terminology, Slater and Wilbur (1997) introduced the distinction between immersion and presence. The authors define immersion as "a description of a technology, and describes the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant." (Slater & Wilbur, 1997, p. 3) Immersion is, therefore, delimited by four elements (Slater & Wilbur, 1997, p. 3): inclusive (referring to "the extent to which physical reality is shut out"), extensive (referring to the sensory feedback provided by the VR), surrounding (whether the VR is panoramic or narrowed to a limited field), and vivid (related to resolution, fidelity or quality of displays). Presence, on the other hand, is "a state of consciousness, the (psychological) sense of being in the virtual environment" (Slater & Wilbur, 1997, p. 5). Presence is related to the sense of physically being there, the impact of immersion on the individuals, and the plot that is supposed to deliver a greater sense of presence if it presents an "alternate self-contained world" (Slater & Wilbur, 1997, p. 5). According to the authors' theory, a one-to-one relationship occurs between immersion and presence. In their words: "presence is the potential psychological and behavioural response to immersion." (Slater & Wilbur, 1997, p. 9)

The quality of the experience delivered by the VR system may have a direct impact on the sense of presence experienced by users. However, Schubert et al. (2001) argue that assuming this one-to-one relationship can be misleading. The authors state that the cognitive processes leading from the perception of the VE to the feeling of presence must be considered. Cognitive processes mediate the impact of immersion on the construction of presence. According to the authors, "[s]timuli from a VE are only the raw material for the mind that constructs a mental picture of a surrounding world, instead of a mental picture of pixels on the display in front of the eyes […]" (Schubert et al., 2001, p. 4).

For the context of this study, we understand presence as defined by Schubert et al. (2001): A sum of cognitive processes that transform the sensorial input delivered by technology into the sense of presence. According to the authors, the main factors contributing to the sense of presence are: a) that users feel physically transported to the VE, as opposed to seeing it from outside; b) that users are focused on what is happening in the VE, instead of paying attention to the real environment or the hardware; and c) that users perceive the virtual world as the real world.

1.4. Methodology

The methodology of the different parts of this dissertation as a compendium of publications is detailed in each chapter corresponding to each article (as well as in the two additional articles included in Annex 1). However, it was considered beneficial to include an overview of the methodology in this introductory chapter for the sake of clarity and consistency. For each main goal and specific objective, a different study was carried out:

Goal 1: To identify challenges and subtitling strategies for immersive content.

Firstly, a focus group was conducted with end users in order to gather feedback about how they were expecting to see subtitles in immersive content, before any solutions were developed, as explained in Chapter 2. The feedback was used to create subtitling solutions, which were further tested and redefined until acceptable solutions were put in place (see Annex 1).

Secondly, a 360° video multimedia corpus was analyzed to explore commercially available subtitling solutions (if any), as explained in Chapter 3. 360° videos from the NYT and the BBC were carefully analyzed in search of potential subtitling solutions that could be implemented in designing subtitling strategies for the study.

Goal 2: To identify professional subtitlers' needs for immersive content.

Firstly, a focus group was carried out with professional subtitlers in order to identify their main concerns and needs in regard to creating subtitles for immersive content, as explained in Chapter 2.

Based on this feedback, the development team in the ImAc project developed a prototype. Then, the prototype of the subtitle editor was tested in a reception study to gather more in-depth feedback from professional subtitlers as part of this dissertation. The purpose was to find out which features were working better and which needed further development.

Goal 3: To evaluate different subtitling strategies in immersive content (360° videos).

Based on all the information and feedback gathered from the previous goals, we defined different subtitling strategies that were tested in a larger reception study involving 40 participants (20 hearing and 20 with hearing loss). We tested two main features of subtitles for immersive media: position and directions. Regarding position, we compared always-visible subtitles (that is, subtitles that are always located in front of the viewer) and fixed-position subtitles (that is, subtitles that are located in a fixed position every 120º within the virtual sphere). For directions, we compared arrows with a radar. The main purpose of the test was to gather feedback from users regarding their preferences and the level of presence experienced with each strategy.

The relationship between the different parts of this study can be better understood if taken as a whole, as seen in Table 2.

Date	Research tool	Approach	Nº of participants	Expected outcome	Article
November 2017	Focus group	Qualitative	14	Gathering user feedback about their expectations and preferences regarding subtitles in 360° content. Gathering professional subtitlers' feedback about their expectations and preferences regarding a hypothetical subtitle editor for 360° content.	Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group (Chapter 2)
April 2018	Multimedia corpus	Qualitative	N/A	Gathering information about the state of the art regarding subtitles in 360° content.	(Sub)titles in cinematic virtual reality: a descriptive study (Chapter 3)
May 2018	Usability testing (ad-hoc preference questionnaires)	Qualitative	6	Gathering iterative user feedback about the first prototype of subtitles in 360° content with a limited number of users. The prototype is designed based on the feedback gathered in the focus group.	From disabilities to capabilities: testing subtitles in immersive environments with end users (Annex 1)
July 2018	Usability test (SUS questionnaire plus ad-hoc preference questionnaire with open questions)	Quantitative and qualitative	27	Gathering iterative feedback about the usability of the first prototype of the subtitle editor for 360° content. The subtitle editor is designed based on the feedback gathered in the focus group.	Technology for subtitling: a 360-degree turn (Chapter 4)
November 2018	Usability testing (IPQ questionnaire plus ad-hoc preference questionnaire)	Qualitative	8	Gathering iterative feedback about subtitles in 360° content, designed based on previous user feedback and state of the art. In this case, not only preferences but also presence is measured.	Making interaction with virtual reality accessible: rendering and guiding methods for subtitles (Annex 1)
March-April, 2019	Reception study (IPQ questionnaire, ad-hoc preference questionnaire plus head movement tracking)	Quantitative and qualitative	40	Gathering feedback about user preferences and levels of presence comparing different subtitling modes in 360° content. The subtitles in this reception study were improved and refined following previous user feedback.	Subtitles in virtual reality: a reception study (Chapter 5)

Table 2. The iterative steps of the study presented chronologically.

1.4.1. Research tools

This study adopted a mixed-method design, combining qualitative and quantitative approaches and triangulating results. This helped to overcome the potential limitations of the applied methods. Also, the nature of the medium under research (360° videos) and the high impact of technology made it important to take a multidisciplinary approach. According to Orero et al. (2018), this is common practice in experimental research on AVT studies, "working beyond the boundaries of and integrating approaches from film studies, literary studies, psycholinguistics, cognitive science, and TS." (Orero et al., 2018, p. 106) In that sense, concepts and tools have been borrowed from usability studies and psychology (presence) and integrated within the study.

It is also common in AVT experimental research to employ questionnaires and self-rating scales, and it is advisable to work with previously validated and widely used questionnaires (Orero et al., 2018). For this study, two previously validated questionnaires were used: the System Usability Scale (SUS) questionnaire (in its original English form) and the IPQ questionnaire (in a non-validated Spanish version).

The SUS questionnaire was conducted to validate the usability of the prototype of the subtitle web editor that was developed in the framework of the ImAc project by the software development company Anglatècnic[6]. Software is an essential part of subtitling practice and it is crucial to involve professional subtitlers in the development of new tools that will impact their professional practice, following a user-centric approach. Even though the usability studies came from a different discipline, it was relevant to include this study in the framework of this research. The questionnaire and the results of the test will be discussed in Chapter 4.

The Igroup Presence Questionnaire (IPQ) was used to measure presence in the reception study because we understand presence as explained by the author of the questionnaire (Schubert, 2003). Scholars researching the concept have not reached a consensus regarding which factors contribute to the construction of presence (Slater & Wilbur, 1997). Questionnaires have been widely used in the research of presence, albeit with some criticism. For example, Slater (1999, p. 10) criticizes the validity of the Witmer and Singer questionnaire (1998) in particular, and of questionnaires in general:

---

[6] https://www.anglatecnic.com/

> I would rather not use questionnaires at all. In many ways very profound insights into the nature of presence can be gained from ethnographic studies (for example, McGreevy, 1993). However, at the end of the day I use questionnaires because, for the time being, I do not know what else to do, and in order to construct predictive equations, concerned with how presence varies with other factors in groups of people, some method of quantification is necessary.

Schubert (2003) developed the IPQ with the aim of measuring the sense of presence. The questionnaire is based on the theory developed by the authors explaining the cognitive processes behind the sense of presence in VEs, mainly the construction of the own body in the VE and the suppression of immediate sensory input (Schubert et al., 2001). According to the authors, "[e]mbodied presence develops from the mental representation of navigation (movement) of the own body (or body parts) as a possible action in the virtual world." (Schubert, Friedmann, & Regenbrecht, 1999, as cited in Schubert et al., 2001, p. 268) The authors explain suppression as follows: "[s]imilarly, when perceiving a VE, usually a number of conflicting sensory inputs must be suppressed, including distracting stimuli from the hardware or the real environment." (Schubert et al., 2001, p. 268)

The abovementioned cognitive processes lead to the construction of presence when experiencing VR. The authors state that the "sense of presence should involve at least two components: The sense that we are located in and act from within the VE, and the sense that we are concentrating on the VE and ignore the real environment." (Schubert et al., 2001, p. 269) When designing the IPQ questionnaire, the authors included questions related to these elements. The IPQ questionnaire considers one general presence item and three subscales: spatial presence, involvement, and realness. According to Schubert (2003, p. 71):

> These two processes are 'echoed' in the feelings of spatial presence and involvement: The users feel present when their mental representation includes plenty of bodily actions that are possible in the VE, and the users feel involved when they successfully suppress the real environment.

One of the main innovations in this study is the medium in which subtitles are being implemented and tested. Immersive media should, as their name suggests, offer a more immersive experience compared to traditional media such as TV or cinema. One of the premises of offering access services in this medium is that the experience of feeling present in the virtual world should not be disrupted by subtitles. Therefore, it was important to measure the impact of the different subtitling strategies on presence. The questionnaire and the results of the reception study will be further discussed in Chapter 5.

Apart from quantitative measures such as the abovementioned questionnaires, qualitative methods were implemented in the study. For example, ad-hoc questionnaires were used with participants to gather qualitative feedback because it was important to know how users felt about the developed subtitling solutions or the subtitle web editor. Focus groups, as well as usability tests with a limited number of groups, were put in place in order to gather iterative feedback before the final solutions were designed. Head movement patterns were tracked in the reception study using the CVR tool developed by Rothe, Höllerer & Hussmann (2018). These patterns were observed on a qualitative basis in order to triangulate results with other tools used in the experiment (IPQ and ad-hoc preference questionnaire), as will be further explained in Chapter 5.

1.4.2. The UCT model in practice

According to the authors of the UCT model, "[a] significant difference between these reception studies and UCT is that UCT emphasizes the eliciting of user feedback iteratively and during the translation process, whereas reception studies typically test completed translations." (Suojanen et al., 2015, p. 112) This model has been followed in this study, applied to subtitling strategies for immersive content. We have iteratively gathered feedback from subtitling users (both hearing and with hearing loss) via focus groups and usability tests until acceptable subtitling solutions were found. Then, the refined solutions were tested in a larger reception study in order to conclude which subtitling strategies proved most suitable for viewers. The steps taken in this study related to the UCT process are specified in Table 3.

Element	Implementation in this study
Inner circle: subtitling strategies and creation	Different subtitling strategies for 360° videos were designed and iteratively tested in order to fine-tune the final solution and come up with usable subtitles.
Subtitling need	Immersive content is an emerging new medium and needs to be made accessible for all users.
Specification	An agreement was reached between the parties involved in the study (engineers, content creators, and AVT experts) and end users (feedback gathered via focus group) in order to define expectations for subtitles in 360° content.
Mental models	The target audience was profiled as users with hearing loss whose main language was oral language, as well as hearing users.
Heuristic evaluation and usability testing	The different subtitling solutions for 360° content were first tested by a group of experts to evaluate usability. Then, the refined subtitles were tested with a limited group of users to gather feedback and fine-tune the subtitling solutions that were tested in a reception study.
Post-mortem	After the first rounds of usability testing, feedback was gathered to improve the implementation of subtitles in 360° videos for the ensuing steps of the project.
Reception research	A reception study was carried out to evaluate the fine-tuned subtitling solutions in order to gather feedback regarding user preferences and the usability of the subtitles.

Table 3. Implementation of the UCT process in this study.

1.5. Structure of the dissertation

This dissertation is presented as a compendium of the following four articles:

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 https://www.jostrans.org/issue32/art_agullo.pdf**

This article reviews the current situation of immersive technologies and their applications in media. Also, research studies carried out so far concerning subtitling in immersive media are discussed, as well as the current implementation of subtitles in immersive content, such as VR video games or 360° videos. Finally, the results from a focus group carried out in Spain with persons with hearing loss, as well as professional subtitlers, on how to subtitle 360° video content are presented. The results from the focus group shed light on how to address the new implications brought by immersive media in regard to subtitling. Some of the main areas discussed in the results are subtitle position, contrast considerations for correct reading, and how to indicate the location of the speakers, among others. Also, results show that users are willing to accept the implementation of new features in subtitles for immersive content, such as icons for non-speech information or improvements to current subtitling standards.

This article is presented in Chapter 2 and is linked to objectives *1.1) Gathering feedback from subtitle users about their expectations and recommendations regarding the consumption of subtitles in immersive content (360° content)*; and *2.1) Gathering feedback from professional subtitlers about their expectations and recommendations regarding the creation of subtitles for immersive content (360° content)*. This chapter is also linked to the first and second main goals of the present study*: 1) to identify challenges and subtitling strategies for immersive content*; and *2) to identify professional subtitlers' needs for immersive content*.

Agulló, B. & Matamala, A. (in press). (Sub)titles in cinematic virtual reality: a descriptive study. *Onomazéin*.

This article presents the state of the art of cinematic virtual reality content and discusses the current challenges faced by filmmakers when dealing with this medium and the impact of immersive content on subtitling practices. Moreover, it also reviews the different studies on subtitles in 360° videos carried out so far and the results obtained. Finally, the results of a corpus analysis are presented in order to illustrate current 360° video subtitling practices by The New York Times and the BBC. The results have shed some light on issues such as position, innovative graphic strategies or the different functions that are challenging current standard subtitling practices in 2D content.

This article is presented in Chapter 3 and is linked to objective *1.2) Describing the state of the art of immersive technology in regard to media applications*. This chapter is also linked to the first main goal of the present study: *to identify challenges and subtitling strategies for immersive content*.

Agulló, B. (in press). Technology for subtitling: a 360-degree turn. *Hermeneus*, *22*.

This article presents an updated review of current subtitling technology to contextualize the study. It then goes on to review the main immersive environments (3D, AR and VR) and their implications for subtitling. The focus of the study is on VR and, therefore, the main challenges of subtitling 360° content are presented. To respond to the needs of subtitling this type of content, a prototype version of a subtitle editor was developed and presented to twenty-seven professional subtitlers who tested the tool and reported the corresponding feedback on usability and preferences. The study proves the importance of carrying out usability tests with end users when developing specific software. Finally, the challenges faced by subtitlers in new audiovisual media such as 360° content are presented.

This article is presented in Chapter 4 and is linked to objective *2.2) Testing a 360° content subtitle editor prototype with professional subtitlers in order to gather their feedback in regard to the needs for subtitling 360° videos*. This chapter is also linked to the second main goal of the present study: *to identify professional subtitlers' needs for immersive content.*

Agulló, B. & Matamala, A. (in press). Subtitles in virtual reality: a reception study. *Íkala*.

In this article, we review the topic of subtitling immersive content, the previous studies, and the main challenges faced when designing subtitles for 360° content: position and guiding methods. The criteria for subtitle position is still not defined for immersive content as it is for content viewed in a traditional context and solutions are required. Guiding mechanisms are necessary for circumstances in which the speakers are not visible and viewers lacking an audio cue need visual information to guide them through the virtual scene. For each area under investigation in this study, two methods are compared. For position, always-visible position (when subtitles are consistently in front of the viewer) is compared with fixed-position (when subtitles are fixed and evenly spaced every 120º in the 360° scene). For guiding methods, arrows (that are integrated into the subtitles and indicate where the viewers need to direct their gaze) are compared to a radar (located in a fixed position in the field of view, indicating where the speaker is located in relation to the position of the viewer). Feedback on preferences, immersion (using the IPQ questionnaire) and head movements is gathered from 40 participants (20 hearing and 20 hard of hearing).

This article is presented in Chapter 5 and is linked to the third main goal: *to evaluate different subtitling strategies in immersive content (360° videos)*.

All articles are presented in this dissertation as accepted for publication; no changes were introduced (only the font and table and figure numbers were unified within the document). A copy of the published articles is presented in Annex 2. The rest of the articles are in press and will be published in 2020 or 2021, after this Ph.D. is completed.

Chapter 6 presents the summary of the results, as requested by the regulations on theses by compendium.

Chapter 7 discusses the results and conclusions of the present dissertation, as well as its theoretical and practical implications and how they can be further developed in future studies.

Finally, the unified bibliography for the dissertation is presented.

The following annexes can be found at the end of the document:

Annex 1: Additional articles

The two articles included in Annex 1 are not part of the main body of this Ph.D. dissertation. However, they are important to understand the user-centered approach and the iterative gathering of user feedback that has been applied in this study.

1.1. Article 1: Agulló, B., Matamala, A., & 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

User testing in MA has often profiled users based on their disabilities. SDH, for instance, have been generally tested with their expected target audience, which is persons with hearing loss. This article argues that selecting users based on sensory disabilities may not be the best strategy to obtain relevant results, as other capabilities—for instance, technological capabilities—may have a greater impact on the results. Moreover, the article argues that access services should not be exclusively for persons with disabilities but also for other audiences. If accessibility is mainstreamed, and ideally integrated in the creation and production processes, testing should expand beyond an exclusive approach based on accessibility to a more general approach based on usability where users with diverse capabilities are considered.

To illustrate this point and propose a new approach to user testing in MA, moving from a disability to a capability model, specific examples from ImAc are shown in a chronological order. Then, the article presents the initial testing, targeting persons with disabilities, and describes the poor data results leading to a new approach. A new testing focus is proposed, and the methodological shift is justified. After that, the second test in which the new approach is implemented is described, using the same stimuli but users with different levels of knowledge regarding new technologies. The article finishes with conclusions in which the door is opened to move from an accessibility approach to testing to a usability approach.

1.2. Article 2: 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

This article explores the appropriateness of two rendering modes (fixed-position and always-visible) and two guiding methods (arrows and auto-positioning) for subtitles in 360° video. All conditions considered have been implemented and integrated in an end-to-end platform (from production to consumption) for their validation and evaluation. A pilot study on eight end users has been conducted with the goals of determining the preferred options for users, identifying the options that result in higher presence, and gathering any additional valuable feedback from end users.

Annex 2: Published article within this dissertation

Apart from the articles in Annex 1, one more article within this dissertation has been already published:

2.1. Article 1: 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

Annex 3: Ethical Committee Documentation

3.1. ImAc Project's ethical procedure approved and signed

3.2. Written informed consent forms and information sheets for participating in the experiments (Spanish)

3.3. Written informed consent forms for video, photographs or audio recording (Spanish)

3.4. Written consent form online (for subtitle editor online test) (English)

Annex 4: Questionnaires for all tests

4.1. Demographic questionnaire – Focus group (Catalan)

4.2. Demographic questionnaire – For article "From disabilities to capabilities: testing subtitles in immersive environments with end users"

4.3. Demographic questionnaire – Web editor test

4.4. Postquestionnaire – Web editor test

4.5. Demographic questionnaire – For article "Making interaction with virtual reality accessible: rendering and guiding methods for subtitles"

4.6. Postquestionnaire (behavior) – For article "Making interaction with virtual reality accessible: rendering and guiding methods for subtitles"

4.7. Postquestionnaire (guiding) – For article "Making interaction with virtual reality accessible: rendering and guiding methods for subtitles"

4.8. IPQ Questionnaire (Spanish)

4.9. Demographic questionnaire – For article "Subtitles in virtual reality: a reception study"

4.10. Postquestionnaire (behavior) – For article "Subtitles in virtual reality: a reception study"

4.11. Postquestionnaire (guiding) – For article "Subtitles in virtual reality: a reception study"

Annex 5: Digital and multimedia resources

5.1. Multimedia corpus:

a) Excel sheet – Pilot version

b) Excel sheet – Final version with all data

5.2. Videos for the reception study.

Chapter 2. Article 1. Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group

Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group

Belén Agulló, Universitat Autònoma de Barcelona

Anna Matamala, Universitat Autònoma de Barcelona

Immersive media such as virtual reality or 360º content is increasingly present in our society. However, immersive content is not always accessible to all, and research is needed on how to cater for the needs of different kinds of users. This article will review the current situation of immersive technologies and their applications in media. Also, research studies carried out so far concerning subtitling and SDH in immersive media will be discussed, as well as current implementation of subtitles in immersive content, such as VR video games or 360º videos. Finally, the results from a focus group carried out in Spain with deaf and hard-of-hearing users, as well as professional subtitlers, on how to subtitle 360º video content will be presented. The results from the focus group shed light on how to address the new implications brought by immersive media in regard to subtitling. Some of the main areas discussed in the results are: subtitle position, contrast considerations for a correct reading, and how to indicate the location of the speakers, among others. Also, results show that users are willing to accept the implementation of new features in SDH in immersive content, such as icons for non-speech information or improvements to current standards.

Subtitles, subtitling for the deaf and hard-of-hearing, SDH, immersive content, virtual reality, 360º content, focus group.

2.1. Introduction

Immersive media are increasingly present in our society and considerable research efforts are put in developing better immersive technologies (Jayaraj *et al.* 2017). According to a report by VR Intelligence (2017), nearly half of the surveyed VR companies (46%) reported a strong or very strong growth in sales. However, some reports agree that the main reasons preventing users to buy VR headsets are economic, because prices are too high, and the lack of available content (VR Intelligence 2017; Sketchfab 2017).

Even if data point to expected growth, immersive content is not always accessible to all, and research is needed on how to cater for the needs of diverse users. Audiovisual Translation (AVT) and more specifically Media Accessibility (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 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 other. Still, most research has dealt with traditional media such as TV or cinema (Perego *et al.* 2015; Romero-Fresco (ed.) 2015), museums (Jiménez Hurtado *et al.* 2012; Szarkowska *et al*. 2016; Neves 2018) or live events (Orero and Matamala 2007; Udo and Fels 2011). In these environments, as in many others such as the localisation and game industry, accessibility has generally been considered an afterthought, despite many voices asking for the inclusion of accessibility in the creation process (Romero-Fresco 2013). To date, little research on accessibility in immersive media has been carried out and immersive technologies are on the rise, but they are still not fully implemented. This scenario was seen as an opportunity to start researching access services while immersive media were being developed, and the ImAc project was set up.

ImAc is a European project funded by the European Commission that aims to research how access services (subtitling, AD, audio subtitles, SL) can be integrated with immersive media. The project aims to move away from the constraints of existing technologies into an environment where consumers can fully customise their experience. The key action in ImAc is to ensure that immersive experiences address the needs of different kinds of users. ImAc follows a user-centred methodology (Matamala *et al*. 2018), so the first step in the project has been to ask users about their expectations. Two focus groups have been carried out in two countries (Germany and Spain) with different user profiles to gather feedback, define user scenarios and establish user requirements regarding SDH in 360º video content. This article aims to discuss the results of the focus group carried out at the Catalan Media Corporation (CCMA) in Spain.

The article begins with an overview of immersive technologies and immersive content in media, in order to contextualise our research. It then explains the limited research that has been carried out so far concerning subtitling in immersive media. Section 5 describes the methodology for the focus group and its implementation, and Section 6 reviews the results.

2.2. Immersive content: an overview

Immersive content allows users to feel as if they were physically transported into a different location. There are different solutions that can provide such experience. Fulldomes are one of those solutions. These are based on panoramic 360º videos projected on a dome structure, such as those that can be seen in planetariums, museums or flight simulators, for example.

Stereoscopic 3D technology represents another type of immersive technology that has had a relative presence in cinemas and homes during the past decade. Specifically, stereoscopy or 3D imaging:

> refers to a technique to create or enhance the illusion of depth in an image by presenting two offset images separately to the left and right eye of the viewer. The two 2D images are then combined in the brain to give the perception of 3D depth. The visual cortex of the brain fuses the two images into the perception of a three-dimensional scene or composition. (Agulló and Orero 2017: 92)

However, 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). Depending on the display design, the way of accessing stereoscopic 3D images differs (anaglyph glasses, head-mounted displays, active/passive glasses, etc.). This lack of standardisation, the intrusive nature of 3D and some uncomfortable side effects (headache or eyestrain) might prevent stereoscopy to become the main display for audiovisual (AV) products (Belton 2012).

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. Sherman and Craig (2003) define four key factors in VR (virtual world, immersion, sensory feedback and interactivity) which result in the following definition:

> **virtual reality** [is] a medium composed of interactive computer simulations that sense the participant's position and actions and replace or augment the feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world). (Sherman and Craig 2003: 13)

VR has also been referred to as a new medium, like the telephone or television, by authors such as Steuer (1992), who considers VR as a set of technical components, for example computers, head-mounted displays, sensors, among others. Most recently, VR has also been referred to as a set of computer-generated images that reproduce a reality and allow users to interact with their surroundings with the appropriate equipment (BBC 2014). However, some authors (Biocca 1992; Steuer 1992) consider technical definitions of VR to be limited. They suggest defining VR in terms of human experience and introducing the concept of *presence*, which can be defined as “the experience of one’s physical environment; it refers not to one’s surroundings as they exist in the physical world, but to the perception of those surroundings as mediated by both automatic and controlled mental processes” (Steuer 1992: 75).

This definition can be applied to different types of immersive content which, contrary to VR, does not need to involve an interactive response and an advanced equipment. Therefore, 360º or omnidirectional videos can also be defined as immersive content: they reproduce highly-realistic images recorded with special camera sets that represent a reality in which the users are observants and cannot interact (BBC 2014). Moreover, images can be combined with audio technologies such as immersive audio aimed to increase immersion. The spatial sounds are created with specially designed microphones that simulate the way in which human beings receive aural stimuli through the auditory system (BBC 2012).

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

> 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.

In this definition, the authors explain their concept of “virtuality continuum” in which MR is the hypernym that includes the more specific term AR. Azuma (1997) further defines AR as a medium that combines the real world with virtual objects that appear superimposed in the real world. 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).

VR and 360º content is accessed by different types of equipment. It can be directly viewed on a flat screen (for example, a computer, smartphone or a TV set) with a remote, touch pad or mouse to change the direction of the field of view (FoV), or it can be accessed with a head-mounted display (HMD), which can be either a tethered or a mobile device. Tethered HMD, such as PlayStation VR, Oculus Rift or Vive, incorporate high definition screens and are connected to high-performance computers or last generation consoles (Deloitte 2016). They are generally used for gaming purposes and the quality of the VR experience is higher. On the other hand, mobile HMD, such as Samsung Gear VR or Google Cardboard, are dependent on smartphone technology (such as accelerometer or gyroscope).

2.3. Immersive content in media

Broadcasters and video content developers are starting to experiment with immersive content. According to a report on VR by the European Broadcasting Union (EBU 2017a), 49% of its members are starting to explore or further develop immersive content. Most EBU members think that the potential of immersive content is clear, because it offers new opportunities to tell stories from a different perspective. However, factors such as technical limitations, lack of knowledge and uncertainty about the return on investment are holding some of them back (EBU 2017a). The current preferred format is 360º video over VR or AR/MR and the trend in terms of duration is 5-10 minutes. Most stories told in 360º/VR are "history or news and current affairs products, as 360º/VR allows the user to gain better understanding of the story being told" (EBU 2017a: 9). Sports and music events are clear candidates for 360º/VR. In the report, attention is also directed at the challenges posed in terms of storytelling, since the plots are non-linear and the level of user interaction is not determined.

Immersive journalism is featured as a key concept in the EBU report. This relatively new concept entails "the production of news in a form in which people can gain first-person experiences of events or situation described in news stories" (De la Peña *et al.* 2010: 291). According to De la Peña *et al.*, VR is a perfect media that could help journalists in eliciting deeper emotions in the audience. Major broadcasters and newspapers such as the BBC, *The New York Times*, *The Washington Post* and *ABC News* have already started working on this. To name just two examples, in 2015 *The New York Times* decided to reward their subscribers with a pair of Google Cardboard glasses, as a strategy to promote their own immersive content (Wohlsen 2015). The company even launched its own smartphone application (NYT VR). More recently, as reported by Jones (2017), *ABC News* recorded a 360º short film about North Korea with a relatively positive outcome. Immersive journalistic experiences have been proven to elicit emotions and increase audience engagement (Jones 2017), showing greater potential for journalism and broadcasting.

As a genre, fiction appears to be a rather underexplored area in relation to immersive content. However, there are some indicators which show that VR has had a moderate penetration in the entertainment industry. For example, the appearance of immersive content in major film and TV industry events is one of them. In 2016, the VR film *Henry* was awarded an Emmy for Outstanding Original Interactive Program (Oculus Team 2016). In 2017, the immersive short film *Pearl* was nominated for the Oscars, under the category of Best Animated Short Film (Hall 2017). Pixar has also developed a VR and interactive experience to market their animated film *Coco* (Lee 2017). However, most fictional immersive experiences are computer-generated products and not images recorded with 360º cameras. According to the BBC, "truly interactive VR video is in its infancy and can be expensive to create, but total or partial animation or CGI can be used very effectively and efficiently, while other production techniques may yet emerge or become more accessible over time" (Conroy 2017). Therefore, it could be inferred that immersive technologies are still not sufficiently developed to be implemented in creating successful fictional films or movies. Some of the reasons could be the hindrances posed by technology, which is delivering a quality that is still not considered suitable for the audience (EBU 2017a); also, the lack of knowledge in delivering well-written immersive stories. It is in this context of development that the integration of access services in the production line should be researched, adopting a user-centred approach. The users' voice needs to be heard before the technology is fully implemented.

2.4. Subtitling and SDH in immersive media

AVT and MA research on immersive content is at an early stage. Nevertheless, we can find studies that have addressed the challenges of creating and consuming subtitles in stereoscopic 3D content (Vilaró 2011; Lambooij *et al.* 2013). Some of the main issues when integrating subtitles in 3D imaging is that superimposing 2D subtitles on a 3D image can generate effects such as ghosting, which hinders the readability of the subtitles and can cause headaches and eyestrain (Agulló and Orero 2017). However, the implementation of some techniques such as the positioning of the subtitle close to the screen plane or the use of lighting, shades and colours to reduce contrast between the screen and the subtitle (González-Zúñiga *et al.* 2013) could contribute to minimising the impact of such issues.

Reception studies on access services in VR and 360º content are almost non-existent, with research on subtitling by the BBC being an exception. Although not focused specifically on SDH, audiences with hearing loss might be potential users of intralinguistic subtitles. The BBC research designed four subtitle scenarios for 360º content (Brown *et al.* 2018: 3-6): (a) subtitles equally spaced by 120º in a fixed position below the eye line; (b) subtitles following head immediately always in front of the user; (c) subtitles following head with lag in front of the user; and (d) subtitles appearing in front of users, and then fixed until they disappear. 24 participants, frequent users of TV subtitles, took part in the study. They randomly viewed the four variables on an Oculus Rift HMD, and replied to a questionnaire (Brown *et al.* 2018: 7-9). Results show that the preferred solution was (b), in line with a similar behaviour to subtitles in 2D. Their conclusion was that sometimes the simplest solution is the best (Brown *et al.* 2018: 29-33), but it remains to be seen whether results would be the same with longer content.

2.4.1. Revisiting subtitling parameters in immersive content

User reception studies in subtitling and in SDH have allowed the definition of a set of preferred parameters for users (Jensema *et al.* 1996; Romero-Fresco 2009; Matamala and Orero (eds) 2010) in a myriad of aspects such as: the number of characters and lines per subtitle; subtitle editing; font type and size; boxes, borders and shadows; justification and spacing; paralinguistic information, and subtitle speed (Neves 2005; Arnáiz-Uzquiza 2012; Romero-Fresco (ed.) 2015). Immersive environments, however, pose specific challenges that need to be considered.

Subtitle positioning, which is a widely researched (Bartoll and Martínez-Tejerina 2010) and standardised parameter for 2D products, is one of the main issues when designing subtitles for immersive content, since user behaviour in the immersive environment is unpredictable (Arrés *forthcoming*). While safe areas for subtitling are already defined and guidelines are provided for content consumed in TV or flat screens (EBU 2017b), recommendations for safe areas in immersive devices such as HMD are still lacking. The FoV for users in VR environments is wider than the FoV in 2D products on a flat screen. But to the best of our knowledge, eye-tracking studies, showing where the users direct their attention when reading subtitles in VR or 360º content, are lacking.

Therefore, the safe area for subtitles in immersive environments needs to be defined very carefully and tested later on.

The interactivity and the freedom of movement that are inherent to immersive products also impact other subtitling parameters such as character identification, because there is the possibility that a character speaks but is located outside the FoV of the user. In 2D SDH, character identification is usually solved by using different colours, name tags or speaker-dependent placement of subtitles or a combination of these (Neves 2005; Arnáiz-Uzquiza 2012). However, immersive content introduces a new dimension: direction, which is particularly relevant in immersive environments.

Another SDH parameter that may change in subtitles for immersive content is the display of non-speech information, such as music or sounds. Immersive technologies offer new opportunities for implementing new features in SDH. In previous studies, the implementation of graphic elements such as icons to display non-speech information has been suggested as a way to optimise reception (Civera and Orero 2010). Other authors have also tested the reception of emotions via emoticons (Arnáiz-Uzquiza 2015) and other creative approaches (Sala Robert 2016). Although these are not extended practices for SDH, the technical advances provided by immersive technologies could open the possibility of introducing new elements that might counterbalance other VR limitations. However, it remains to be seen whether alternative approaches such as using icons may help reduce that kind of discomfort.

2.4.2. Some examples from current practices

Some randomly selected current experiences in subtitling immersive environments can give us food for thought as to the opportunities and challenges subtitles in immersive AV products may pose. The Spanish television series *El Ministerio del Tiempo* (The Ministry of Time) launched an immersive experience in the form of an interactive episode. The episode “El tiempo en tus manos” (The time is in your hands) is one of the first fictional TV episodes launched in an immersive format. The interlingual subtitles (Spanish into English) in this short episode are positioned slightly below the centre of the screen, following the movement of the head and floating through the screen as the user’s head moves. The transition of the subtitles, when the users move their heads, presents a slight delay in reaction time. Therefore, when the movement is abrupt, subtitles are not positioned in the centre of the screen, but float in the direction of the movement of the head. They only settle into a fixed position in the centre of the image when the user's head is still. The font type is a white sans serif font without a background box. The justification is centred. The segmentation rules are not followed, and some linguistic issues are found, such as missing information.

Another example is the clip *The Displaced* created by *The New York Times*, in which children who have been driven away from their homes explain their current situation as refugees. In this video, subtitles are burnt in in three different fixed positions in the 360º video, so when the users move their head to explore the scenario, they will always find the subtitles somewhere in their FoV (Brown *et al.* 2017). The font is white sans serif, smaller than the previous example, which hinders readability. Moreover, these subtitles do not include a background box and the contrast is very low, meaning that sometimes the text is very difficult to read.

Video games provide other examples of subtitles in immersive environments, although subtitling practices in video games do not always follow the same rules as in other AV content (Mangiron 2013). In the game *Eve Valkyrie*, for PlayStation VR, intra- and interlingual subtitles are located in a fixed position in the centre of the screen. Therefore, if the users turn their head towards a different part of the scene, they will not be able to read the subtitle. This strategy might result into less freedom of movement, but it might also help avoiding distractions from the main action when the narrative requires the user's attention. The font of the subtitles is sans serif and yellow. Subtitles contain more than 2 lines in many cases, and do not follow segmentation rules.

Summer Lesson and *London Heist*, both games for PlayStation VR, use a similar strategy to implement subtitles. In this case, intra- and interlingual subtitles are always displayed in front of the user, at the bottom of the FoV and centred, which is less intrusive for the scene. They both use sans serif fonts in white. The subtitles in *London Heist* include a black background box to facilitate reading. This strategy could be appropriate for immersive environments, because the user has freedom of movement and it would be very complex for the professional subtitler to foresee where the video background could interfere with the reading and, therefore, change the position of the subtitle, as is the case in current subtitling practices for 2D content. Finally, another strategy for implementing subtitles in VR games appears in *Battle Zone*, for PlayStation VR. In this game, subtitles are integrated in the scene, as they would appear in a head-up display. In this example, the subtitles are not obtrusive in the scene because they appear as if they were part of the environment, integrated in a futuristic spaceship.

The previous examples have shown how some critical issues such as subtitling positioning have been implemented in a selection of subtitled immersive content. Others, such as character identification or non-speech information display, have not been addressed at all. However, it is paramount to gather user feedback in order to generate subtitles that can be easily implemented and accepted by end users. Focus groups such as the one presented in the next section can contribute to such an end.

2.5. Gathering user feedback: focus group methodology

Focus groups were considered appropriate for identifying user needs at the beginning of the ImAc project, before access services in immersive media were actually implemented. A shared methodology was developed for the five focus groups (three on AD and two on SDH) which took place in four different countries, and ethical clearance was obtained. The preparation stage for the focus groups involved two main steps.

To the best of our knowledge, access services such as subtitles or audio description in immersive media were not fully implemented at the time of conducting the focus groups. Therefore, in order to identify the most relevant questions to be posed to participants, it was necessary to define user types and scenarios. Two main user profiles were defined: those creating the services, i.e. professional users (for instance, IT, graphic designers, subtitlers, audio describers, and SL interpreters), and those consuming the services, i.e. home users (for instance, deaf, hard-of-hearing, blind, low vision users, the elderly). At this stage it was decided that the focus would be mainly on those consuming the services, gathering additional data from a few professionals and opening the door to future research with other professional profiles such as content creators. It was also decided that home users would be advanced, meaning they would have some knowledge or special interest on the technologies being developed.

The focus group presented in this paper was organised by the Catalan Media Corporation (CCMA) in collaboration with Universitat Autònoma de Barcelona and was held on 28 November 2017 at the CCMA premises. The main aim of the focus group was to obtain feedback regarding expectations, recommendations and desires from professional and home users when consuming and editing SDH in 360º videos, as well as SL access services. The focus of this article is on subtitling, so only the results related to SDH will be reported. The results of this focus group reflect the needs of the Spanish SDH audience. Another focus group regarding SDH was carried out in Berlin (Germany) by another project partner for which the results have not been published. The contents of the focus groups were different and so was the target audience (different language and subtitling habits). Therefore, the results were not fully comparable, and it was decided not to include them in the present study.

2.5.1. Participants

There were 14 participants (6 males, 8 females): 10 advanced home users (6 signers, 4 oralists) and 4 professional users (2 subtitlers, 1 technical expert, 1 representative from a user association). Age range was 21-40 (3), 41-60 (7), and +60 (4). Three participants had secondary education studies, four participants had further education studies, six had university studies and one person did not reply to this question. Three of them reported having a device to access VR content (VCR, glasses and PC, respectively). Mobile phones were the technology most frequently used by the participants on a daily basis (14), followed by TV (14), laptop (12), PC (10) and tablet (8). The advanced home users were deaf (8) and hearing-impaired people (2), most having the disability from birth (4) or when they were between 0-4 years (5) or 41-60 years (1). The preferred devices for watching online video content was PC (7) and laptop (7), followed by smartphone (5), tablet (3) and TV (3).

Even though the recommended group size is up to 10 (Bryman 2004: 507), the profiles of deaf and hard-of-hearing users are diverse (Báez Montero and Fernández Soneira 2010: 26) and it was therefore considered that including the maximum number of home users would provide a more accurate demographic sample. The diversity in the results of the focus group confirmed that this approach was appropriate. Moreover, following standard recommendations on focus groups, it was deliberately decided to over-recruit in order to allow for no-shows (Wilkinson 1999: 188) and a higher number of short suggestions (Morgan 1998: 75).

2.5.2. Procedure

The focus group included five stages. First, participants were welcomed by the facilitator who briefly explained the aim of the ImAc project. The focus group took place in a meeting room equipped with a table and chairs for the participants and a computer and a large TV screen to show the examples to be discussed. A SL interpreter was present, as well as two researchers who took notes and summarised the conclusions in real time. Secondly, the aim of the focus group was explained to the participants, and they were asked to sign informed consent sheets. The third step was filling in a short questionnaire on demographic information. Finally, the group discussion began. To trigger the discussion, the facilitator gave a short introduction to VR and 360º content and explained how 360º content can be accessed, showing VR glasses to the participants. He explained that 360º content can also be accessed on a flat TV screen using a mouse to move around the 360º scene. As a specific example, an excerpt of the TV show *Polònia* was shown to participants on a flat TV screen. Different types of subtitles were presented to give users some ideas about how SDH could be implemented in immersive content and to stimulate their imagination: subtitles located in a fixed position, subtitles located close to the speaking character, and subtitles located each 120º in the 360º view. The facilitator also posed questions about how users would like to interact with this type of access services and what features a future platform giving access to these services should have. Together with these stimuli, the facilitator also used a list of guiding questions grouped under major topics to generate participants' reactions, taking special care to allow participants to raise aspects that they considered relevant even if not included in the list. A balance between an open-ended and a structured approach was sought, and the result was a lively discussion in which interesting suggestions were made.

As the focus group took place, one researcher was drafting a list of conclusions. Reading these conclusions and agreeing on them was the last step of the focus group, which lasted 90 minutes. At the end of the session, participants were thanked for their participation and they were told about the next steps in the project.

2.6. Focus group results

Data analysis followed a qualitative approach, due to the number of participants and the methodological tool chosen. As explained above, two researchers took notes on a shared document and summarised the conclusions in real time. After the focus group, the notes were thoroughly revised and tagged. This procedure allowed to identify three main areas in which users voiced their views: (1) feedback from advanced home users concerning the services; (2) feedback from advanced home users concerning the interaction with a future platform giving access to the services; and (3) feedback from professional users concerning content creation. The analysis also allowed to define aspects in which there was consensus among users and aspects in which opinions diverged, as described next.

2.6.1. Advanced home users: services

In general, home users considered that subtitles in immersive media should be based on approved subtitling rules (AENOR 2003) and, if necessary, improvements might be implemented to adapt existing rules to the new needs posed by immersive environments.

Regarding the position of the subtitles, users suggested that subtitles should always appear in a fixed position in relation to the users' FoV. They also agreed that subtitles should always appear at the bottom — except in specific cases, such as football matches. There was a brief discussion about the possibility of customising the position of the subtitles. It was even suggested that the placement of subtitles in real time should be changed while watching the 360º video. However, users finally disregarded this option, since they all agreed that subtitles at the bottom of the FoV was the most comfortable solution. It remains to be seen whether research will actually confirm this is the best solution.

Most participants were concerned about the fact that sometimes the subtitles could not be read because of the background image. They stated that it is important, therefore, to have the possibility to choose subtitles with a black background box to facilitate the reading. Also, some participants expressed their worry about the fact that subtitles in immersive media could be disruptive if they appear in a close-up or some other scenarios where the subtitle is obstructing the image. They said that subtitle editors should pay special attention to avoid disrupting the immersive experience.

For character identification, users stated that it is necessary to maintain colour coding to identify characters, as this is already done in SDH for 2D content.

Concerning the display of non-speech information (sounds, music, paralinguistic information, etc.), different options were proposed. In general, users requested that basic subtitling elements that have been previously approved in the regulations (for example, how to indicate music) should be retained. However, they accepted that new technologies may bring new possibilities. Some users preferred to receive non-speech information in the form of text in brackets, as is now the case in most subtitled TV programmes. Other users, considering the new technology in use, preferred to receive this information as icons. In that sense, users suggested the possibility of using a closed list of icons. For example, a lightning icon to indicate the sound of a storm. Regarding the position of non-speech information, users did not reach a consensus. Some stated that they preferred them to be located at the top, others at the bottom close to the subtitle (dialogue), and others would like to move them to a different location. In general, participants did not like the idea of having non-speech information at the top, as it is stated in the current Spanish UNE rule for SDH, because they do not have time to read both the subtitle (at the bottom) and the non-speech information (at the top). They suggested to change this in immersive environments and place the non-speech information in form of icons or text between brackets close to the subtitle area, stating that this would be easier to process. Also, some hard-of-hearing participants stated that they do not need non-speech information in the subtitles, and they would prefer to deactivate them if possible. In this sense, users would like to be able to customise the position of non-speech information.

Users also considered the challenges that the new dimension brought by immersive media (i.e. space and the need to indicate directions) would entail when it comes to SDH in immersive content. In that sense, users stated that it was difficult to know where to look to see the character speaking. They considered that the subtitle should indicate how you need to move your head (four directions), with icons (arrows), indicators between brackets (to the left, to the right) or some sort of mechanism. It was suggested that a compass or radar could be used to that end and that it should be always visible on the screen. Participants also agreed that the radar or compass should be close to the subtitle, otherwise it could be distracting.

As for the subtitle content, users insisted that it should include all the information, both on screen and off screen; in other words, dialogues taking place both within and outside the user's FoV. They suggested that this could be indicated with ON and OFF tags. They also stated that there are different needs among users and, consequently, subtitles must be adapted to different profiles. For example, there could be different levels of speed (faster/slower). However, users considered that summarised or simplified subtitles do not generally help deaf people, because this type of subtitle make it more difficult to follow the AV content. Nevertheless, they conceded that simplified subtitles may be useful for users with other types of needs and could be considered an alternative. It was clear that user profiles are diverse, and customisation should be a priority.

2.6.2. Advanced home users: platform interaction

Users were asked about the options and features that they would like to have in an immersive platform which would give access to virtual content with access services implemented. At the time of the focus group, no prototype was available. Therefore, thought-provoking questions were presented to participants based on hypothetical user scenarios.

Regarding interaction with an immersive interface, users positively valued the possibility of personalisation, i.e. having different layers that could be activated or not. For example, some participants preferred subtitles only for dialogues, others needed non-speech information indications and others wanted to have as much indications as possible, including directions. However, some elements were not considered in need of customisation, such as the position of the subtitles, which it was reported should always be at the bottom of the field of view, because they considered it would be easier to read. Moreover, both professional and home users considered that the user should customise this future platform the first time but then those parameters should be recorded by the interface for future use. Users also suggested that this customisation should be transferrable from one device to the other (importing profile, that is, the user profile could be imported) and they requested the possibility of creating more than one profile. They also considered the possibility of transferring a profile from the user’s device to another external device (for example, at a friend’s home).

Regarding interaction with access services, users positively valued the possibility of alternative interactions (for example, voice commands), although they did not find it necessary for their specific needs and indicated that implementation costs should be taken into account. However, they added that if this platform were to be developed for other profile types (for example, blind users), it could be an additional resource.

Regarding companion screens, participants liked the possibility of using the smartphone to interact with the platform as a touch-screen (like a “mouse”) and to customise their preferences. One user even suggested the possibility of including a finger-sensor that would allow users to see their own fingers on the virtual image. There were different opinions regarding the need of reproducing the same content on the smartphone screen, since the smartphone is often used as an element to access additional content. When accessing AV content together with other people, users did not want to consume subtitles on a different screen because this made them feel excluded.

2.6.3. Professional users: SDH creation

Professional users expressed their proposals regarding the production of SDH in immersive environments. They agreed that vertical positioning of the subtitles could be an interesting option for separating dialogue subtitles from non-speech information, although they considered that home users must be able to decide or set up where they prefer to locate the subtitle.

Regarding the production of subtitles, they stated that they preferred an on-screen display (player) showing one dynamic angle of the 360º view, so that they could choose which angle to see using cursors or mouse movements. Professional users considered that they should be able to test the results with both HMD and flat screen (for instance, a PC screen).

Regarding the subtitling tool, users indicated that they would need a subtitling editor similar to the existing ones for SDH, but it should add the 360º displaying and the possibility of adding emoticons and text messages to show sound actions that take place parallel to the dialogue subtitles. They also highlighted the need for the editor tool to offer original 360º immersive audio because it is important to identify where the sound/dialogues come from, as this information is requested by end users.

2.7. Conclusions

This article has described the role of immersive media in our society and has put forward the need to make them accessible to all users. The emphasis has been put on how subtitles can be integrated in immersive environments, with reference to the limited existing research and practice. Adopting a user-centred approach, the results of a focus group on SDH developed as part of the ImAc project have been presented. According to participants' feedback, SDH for 360º videos should: (1) be located in a fixed position and always visible in relation to the FoV and preferably at the bottom; (2) have a background box to avoid contrast issues with an unpredictable background; and (3) include a system to indicate directions when the speaker is outside the FoV, such as arrows, a compass or text between brackets. Also, results show that home users are willing to accept the implementation of new features in SDH in immersive content, such as icons for non-speech information, because of the new possibilities and dimensions brought by this medium. Moreover, customisation options appear to be a desirable feature among participants. Users also show their agreement and interest in continuing established practices and regulations for SDH, such as the Spanish subtitling standard UNE 153010

(AENOR 2003). However, they agree in introducing some changes to improve the current standards. For example, they would like to have the non-speech information or direction information closed to the subtitle area and not at the top as it currently is, in order to avoid distractions.

One of the limitations of the present study is that it only applies to the Spanish audience and should be replicated in other countries to confirm the validity and generalisation of the conclusions. Another focus group was carried out in Germany for the ImAc project,[7] with similar results (German participants would also like the subtitles always visible in the FoV and also suggested using an arrow to indicate the location of the speaker). However, the contents and examples were not the same, because the audiences spoke different languages and, therefore, the results are not comparable. It was not the intention of the project to compare the two focus groups, but rather gather a general feedback from end users that would set the basis to start developing a prototype for access services in 360º videos. The prototype will be later tested with a larger number of participants in the next stages of the project.

Therefore, the next step will be to transfer user feedback into user requirements and implement the features in immersive content in order to verify whether it is technically possible or whether there are limitations.

Once implemented, user testing will be necessary to verify or reject the validity of the proposed SDH models.

In conclusion, it is clear that research is needed in the field of MA for immersive media. For this purpose, the ImAc project will be a perfect laboratory environment for the development of a successful SDH model for immersive content. This is a significant step in the field of AVT and MA, since the design of accessibility will be taken into account before the technology is fully implemented in society.

2.8. Acknowledgements

ImAc has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No.761974. The authors are members of TransMedia Catalonia, a research group funded by the Secretaria d’Universitats i Recerca del Departament d’Empresa i Coneixement de la Generalitat de Catalunya, under

[7] Results for the German focus group can be found in the public report: http://www.imac-project.eu/documentation/deliverables/.

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 Belén Agulló'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 the Universitat Autònoma de Barcelona.

We would like to thank CCMA, and especially Francesc Mas and Jordi Mata, who were involved in the focus group management at the CCMA premises. And, of course, to all participants who took part in the focus group.

Bibliography

Agulló, Belén and Pilar Orero (2017). "3D Movie Subtitling: Searching for the best viewing experience." *CoMe - Studi di Comunicazione e Mediazione linguistica e culturale* 2, 91-101.

Arnáiz-Uzquiza, Verónica (2012). *Subtitling for the deaf and the hard-of-hearing: Some parameters and their evaluation*. PhD thesis. Universitat Autònoma de Barcelona. https://www.tdx.cat/handle/10803/117528 (consulted on 08.01.2019).

Arnáiz-Uzquiza, Verónica (2015). "Eye tracking in Spain." Romero-Fresco, Pablo (ed.) (2015). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe*. Bern: Peter Lang, 263-287.

Arrés, Eugenia (forthcoming). "Traducir en entornos de realidad virtual: cómo garantizar la inmersión del usuario." To appear in *New Technologies Applied to Translation Studies: strategies, tools and resources*.

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*. Madrid: Asociación Española de Normalización y Certificación.

Azuma, Ronald (1997). "A Survey of Augmented Reality." *Presence: Teleoperators and Virtual Environments* 6(4), 355-385.

Azuma, Ronald *et al.* (2001). "Recent Advances in Augmented Reality." *IEEE Computer Graphics and Applications* 21(6), 34-47.

Báez Montero, Inmaculada Concepción and Ana María Fernández Soneira (2010). “Spanish deaf people as recipients of closed captioning.” Anna Matamala and Pilar Orero (eds) (2010). *Listening to subtitles: Subtitles for the deaf and hard-of-hearing*. Frankfurt: Peter Lang, 25-44.

Bartoll, Eduard and Anjana Martínez-Tejerina (2010). “The positioning of subtitles for the deaf and hard of hearing.” Anna Matamala and Pilar Orero (eds) (2010). *Listening to subtitles: Subtitles for the deaf and hard-of-hearing*. Frankfurt: Peter Lang, 69-86.

British Broadcasting Corporation (BBC) (2012). “Binaural Sound: Immersive spatial audio for headphones.” http://www.bbc.co.uk/rd/projects/binaural-broadcasting (consulted on 19.12.2017).

British Broadcasting Corporation (BBC) (2014). “360 Video and Virtual Reality: Investigating and developing 360-degree video and VR for broadcast-related applications.” http://www.bbc.co.uk/rd/projects/360-video-virtual-reality (consulted on 19.12.2017).

Belton, John (2012) “Digital 3d cinema: Digital cinema’s missing novelty phase.” *Film History: An International Journal* 24(2), 187–195.

Biocca, Frank (1992). “Virtual Reality Technology: A Tutorial.” *Journal of Communication* 42(4), 23–72.

Brown, Andy *et al*. (2017). “Subtitles in 360-degree Video.” Judith Redi *et al.* (eds) (2017). *TVX '17 ACM International Conference on Interactive Experiences for TV and Online Video* (Hilversum, The Netherlands, 14-16 June 2017). New York: ACM, 3-8.

Brown, Andy *et al.* (2018). “Exploring Subtitle Behaviour for 360° Video. White Paper WHP 330.” https://www.bbc.co.uk/rd/publications/whitepaper330 (consulted on 13.02.2019).

Bryman, Alan (2004). *Social Research Methods*. New York: Oxford University Press.

Civera, Clara and Pilar Orero (2010). “Introducing icons in subtitles for deaf and hard of hearing: Optimising reception?” Anna Matamala and Pilar Orero (eds) (2010). *Listening to subtitles: Subtitles for the deaf and hard-of-hearing*. Frankfurt: Peter Lang, 149-162.

Conroy, Andy (2017). “The BBC and Virtual Reality.” *BBC Research & Development*. http://www.bbc.co.uk/rd/blog/2016-06-the-bbc-and-virtual-reality (consulted on 21.12.2017).

De la Peña, Nonny *et al*. (2010). “Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News.” *Presence: Teleoperators and Virtual Environments* 19(4), 291-301.

Deloitte (2016). *Technology, Media & Communications Predictions*. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Technology-Media-Telecommunications/gx-tmt-prediction-2016-full-report.pdf (consulted on 20.12.2017).

European Broadcasting Union (EBU) (2017a). *Virtual Reality: How are public broadcasters using it?* https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it (consulted on 18.12.2017).

European Broadcasting Union (EBU) (2017b). *Safe Areas for 16:9 Television Production*. https://tech.ebu.ch/docs/r/r095.pdf (consulted on 31.01.2018).

González-Zúñiga, Diekus, Carrabina, Jordi and Pilar Orero (2013). “Evaluation of Depth Cues in 3D Subtitling.” *Online Journal of Art and Design* 1(3), 16–29.

Greco, Gian Maria (2016). “On Accessibility as a Human Right, with an Application to Media Accessibility.” Anna Matamala and Pilar Orero (eds) (2016). *Researching Audio Description.* London: Palgrave Macmillan, 11-33.

Hall, Charli (2017). *Watch VR’s first Oscar-nominated short film in Polygon*. https://www.polygon.com/2017/1/24/14370892/virtual-reality-first-oscar-nominated-short-film-pearl (consulted 24.10.2017).

Holliman, Nicolas *et al*. (2011). “Three-dimensional displays: a review and applications analysis.” *IEEE Trans. Broadcast* 57(2), 362–371.

Jayaraj, Lionel, Wood, James and Gibson, Marcia (2017). “Improving the Immersion in Virtual Reality with Real-Time Avatar and Haptic Feedback in a Cricket Simulation.” Wolfgang Broll *et al.* (eds) (2017). *Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality*. Los Alamitos, California: Conference Publishing Services, 310-314.

Jensema, Carl, McCann, Ralph and Scott Ramsey (1996). "Closed-captioned television presentation speed and vocabulary." *American Annals of the Deaf* 141(4), 284-292.

Jiménez Hurtado, Catalina, Seibel, Claudia and Silvia Soler Gallego (2012). "Museos para todos. La traducción e interpretación para entornos multimodales como herramientas de accesibilidad universal." *MonTI. Monografies de Traducció i Interpretació* 4, 349-383.

Jones, Sarah (2017). "Disrupting the narrative: immersive journalism in virtual reality." *Journal of Media Practice* 18(2-3), 171-185.

Lambooij, Marc, Murdoch, Michael, Ijsselsteijn, Wijnand and Ingrid Heynderickx (2013). "The impact of video characteristics and subtitles on visual comfort of 3D TV." *Displays* 34(1), 8-16.

Lee, Nicole (2017). *Pixar's VR debut takes you inside the entrancing world of 'Coco'*. Engadget. https://www.engadget.com/2017/11/15/pixar-coco-vr/ (consulted on 21.12.2017).

Mangiron, Carme (2013). "Subtitling in game localisation: a descriptive study." *Perspectives: Studies in Translation Theory and Practice* 21(1), 42-56.

Matamala, Anna and Pilar Orero (eds) (2010). *Listening to subtitles: Subtitles for the deaf and hard of hearing*. Bern: Peter Lang.

Matamala, Anna *et al.* (2018). "User-centric approaches in Access Services Evaluation: Profiling the End User." *Proceedings of LREC Workshop "Improving Social Inclusion using NLP: Tools, Methods and Resources", ISINLP2 2018*. https://ddd.uab.cat/record/189707 (consulted on 13.02.2019).

Milgram, Paul and Fumio Kishino (1994). "A taxonomy of mixed reality visual displays." *IEICE Transactions on Information Systems* 77, 1321-1329.

Morgan, David L. (1998). *Planning Focus Groups*. Thousand Oaks: Sage.

Neves, Joselia (2005). *Audiovisual translation: Subtitling for the deaf and hard of hearing*. PhD Thesis. Roehampton University. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.1405&rep=rep1&type=pdf (consulted on 08.01.2019).

Neves, Joselia (2018). “Cultures of Accessibility: Translation making cultural heritage in museums accessible to people of all abilities”. Sue-Ann Harding and Ovidi Carbonell Cortes (eds). *The Routledge Handbook of Translation and Culture*. London/New York: Routledge, 415-430.

Oculus Team (2016). *Oculus Film Short “Henry” Wins an Emmy!* Oculus Blog. https://www.oculus.com/blog/oculus-film-short-henry-wins-an-emmy/ (consulted on 24.10.2017).

Orero, Pilar and Anna Matamala (2007). “Accessible opera: Overcoming linguistic and sensorial barriers.” *Perspectives: Studies in Translation Theory and Practice* 15(4), 262-277.

Perego, Elisa, Del Missier, Fabio and Sara Bottiroli (2015). “Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects.” *Perspectives: Studies in Translation Theory and Practice* 23(1), 1-21.

Remael, Aline, Orero, Pilar and Mary Carroll (eds) (2014). *Audiovisual translation and media accessibility at the crossroads*. Amsterdam/New York: Rodopi.

Romero-Fresco, Pablo (2009). “More haste less seed: Edited vs. verbatim respoken subtitles.” *VIAL (Vigo International Journal of Applied Linguistics)* 6, 109-133.

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

Romero-Fresco, Pablo (ed.) (2015). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europ*e. Bern: Peter Lang.

Sala Robert, Elia (2016). *Creactive subtitles. Subtitling for all*. PhD Thesis. Universitat Pompeu Fabra. https://repositori.upf.edu/handle/10230/27737 (consulted on 08.01.2019).

Sherman, William and Alan Craig (2003). *Understanding Virtual Reality: Interface, Application, and Design*. San Francisco: Morgan Kaufmann Publishers (An imprint of Elsevier Science).

Sketchfab (2017). *VR Industry Trends: A comprehensive overview on the state of the VR Industry in Q2-2017*. https://sketchfab.com/trends/q2-2017 (consulted on 18.12.2017).

Steuer, Jonathan (1992). "Defining Virtual Reality: Dimensions Determining Telepresence." *Journal of Communication* 42(4), 73-93.

Szarkowska, Agnieszka *et al*. (2016). "Open Art: Designing Accessible Content in a Multimedia Guide App for Visitors with and without Sensory Impairments". Anna Matamala and Pilar Orero (eds). *Researching Audio Description. New Approaches*. London: Palgrave Macmillan, 301-320.

Udo, John Patrick and Deborah Fels (2011). "From the describer's mouth: reflections on creating unconventional audio description for live theatre." Adriana Serban, Anna Matamala and Jean-Marc Lavaur (eds). *Audiovisual Translation in Closeup: Practical and Theoretical Approaches*. Bern: Peter Lang, 257-278.

Vilaró, Anna (2011) "Does 3D stereopsis change the way we watch movies? An eye-tracking study". Paper presented at *4th International Conference Media for All* (London, 28 June-1 July 2011).

VR Intelligence (2017). *Virtual Reality Industry Survey*. http://vr-intelligence.com/vrx/docs/VRX-2017-Survey.pdf (consulted 19.12.2017).

Wilkinson, Sue (1999). "Focus Groups Methodology: A Review." *International Journal of Social Research Methodology* 21, 111-25.

Wohlsen, Marcus (2015). "Google Cardboard's New York Times Experiment Just Hooked a Generation on VR." *Wired Magazine*. https://www.wired.com/2015/11/google-cardboards-new-york-times-experiment-just-hooked-a-generation-on-vr/ (consulted on 21.12.2017).

Chapter 3. Article 2. (Sub)titles in cinematic virtual reality: a descriptive study

(Sub)titles in cinematic virtual reality: a descriptive study

Belén Agulló, Universitat Autònoma de Barcelona

Anna Matamala, Universitat Autònoma de Barcelona

Virtual reality has attracted the attention of industry and researchers. Its applications for entertainment and audiovisual content creation are endless. Filmmakers are experimenting with different techniques to create immersive stories. Also, subtitle creators and researchers are finding new ways to implement (sub)titles in this new medium. In this article, the state-of-the-art of cinematic virtual reality content is presented and the current challenges faced by filmmakers when dealing with this medium and the impact of immersive content on subtitling practices are discussed. Moreover, the different studies on subtitles in 360º videos carried out so far and the obtained results are reviewed. Finally, the results of a corpus analysis are presented in order to illustrate the current subtitle practices by The New York Times and the BBC. The results have shed some light on issues such as position, innovative graphic strategies or the different functions, challenging current subtitling standard practices in 2D content.

subtitles, subtitles for the deaf and hard of hearing, multimedia corpus, 360º videos, immersive media

3.1. Introduction[8]

Virtual reality is a dangerous medium. Those are the words of the Academy Award-winning filmmaker Steven Spielberg (Spielberg, quoted in Ferrari, 2016). It is dangerous for storytellers, because in this medium the viewers have the freedom to decide where to look at. This can be challenging for film directors who, until now, knew how to direct the viewers' attention (i.e., using close-ups or framing specific parts of the scene). The cinematic language in classical narrative media (such as TV or film) is established

[8] This article is related to the research carried out in the European funded project ImAC (GA: 761974). The authors are members of TransMedia Catalonia, an SGR research group funded by "Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya" (2017SGR113). This article is part of Belén Agulló'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 the Universitat Autònoma de Barcelona.

and audiences know what to expect from a film or an episode. Certain established rules are generally followed by directors, unless they want to surprise the viewer.

In subtitling, there are also standardised practices regarding many aspects: position, character identification, speed, number of lines, number of characters, etc. Filmmakers accept that subtitles will be integrated, or most frequently added, to their audiovisual work in order to reach a wider audience (including non-native speakers or persons with hearing loss). As in filmic creation, there are also Audiovisual Translation studies that challenge traditional subtitling practices, encouraging more creative and integrated subtitles (Lee *et al.*, 2007; Foerster, 2010; McClarty, 2012 and 2014; Fox, 2016a and 2016b).

The situation in immersive media is different. Cinematic virtual reality is still in its infancy and research on the creation of this type of content is ongoing (Sheikh *et al.*, 2017; Dooley, 2017; Mateer, 2017; Gödde *et al.*, 2018). Content creators and broadcasters are experimenting with this new medium (EBU, 2017), and storytelling and production strategies have not been established yet. Similarly, the integration of subtitles in immersive environments is yet to be defined and the challenges are multiple. Subtitles should be generated "in an immersive, engaging, emotive and aesthetically pleasing way" and "disrupt [the immersive experience] as little as possible" (Brown *et al.*, 2018: 1), guaranteeing both accessibility and usability.

A necessary first step before defining guidelines for the creation of subtitles in immersive environments is to describe the limited existing practices and gather viewers' feedback. This article will focus on the descriptive aspect and will analyse the (sub)titles found in a multimedia corpus of cinematic virtual reality content generated by The New York Times and the BBC[9].

The article begins with an overview of immersive content and it then explains the research that has been carried out so far concerning subtitling in 360º videos. Section 4 describes the methodology for the multimedia corpus creation and analysis, and Section 5 summarises the results. Section 6 introduces the discussion generated by the results and Section 7 presents the conclusions of the study.

[9] The *sub* of subtitles is written between brackets, because, as it will be explained below, in the case of cinematic virtual reality, position is relative and subtitles do not always appear at the bottom center of the screen or field of view.

3.2. Immersive content

Immersive technologies are mainly designed to elicit the feeling of being there (Heeter, 1992). This concept of presence has a clear potential for entertainment and audiovisual industries. According to a report on virtual reality (VR) issued by the European Broadcasting Union (EBU, 2017), 49% of its members are developing or planning to develop immersive content. They believe that this new medium offers new opportunities to tell stories from a different perspective and make them more engaging. Also, the video games industry is believed to be the most impacted by VR technology (VR Intelligence, 2017).

VR is a wide term that encompasses different types of devices, products and contents, from 360º videos that can be watched on YouTube on a smartphone to interactive video games to be played with a head-mounted display (HMD) such as Oculus Rift connected to a high-performance computer. In this article, when the terms 360º videos or immersive content are used, it refers to the concept of cinematic virtual reality (CVR) defined by Mateer (2017: 15):

> While a formal definition of CVR is still being developed, the emerging consensus is that the term refers to a type of immersive VR experience where individual users can look around synthetic worlds in 360°, often with stereoscopic views, and hear spatialised audio specifically designed to reinforce the veracity of the virtual environment (as a note, there are presently no initiating studies or foundational articles that can be seen as seminal at this point). Unlike traditional VR in which the virtual world is typically generated through graphics processing and audio triggers in real-time, CVR uses pre-rendered picture and sound elements exclusively. This means that the quality of these assets can approach that found in high-end television or feature film.

Other authors, such as MacQuarrie and Steed (2017: 45), also point out that the majority of CVR content are "monoscopic, passive, fixed-viewpoint 360° videos." They also believe that "real-time rendered, story-led experiences also straddle the boundary between film and virtual reality."

3.2.1. Features of cinematic virtual reality

From a technical point of view, 360º videos are mostly filmed using specially designed cameras that capture overlapping views that are then stitched together with video editing software. The result of that is a full sphere referred as viewing sphere (MacQuarrie and Steed, 2017).

The duration of cinematic virtual reality content tends to be short compared to traditional cinematic narrative content such as films or series. The average duration is from five to ten minutes (EBU, 2017; Dooley, 2017; MacQuarrie and Steed, 2017). According to the present corpus study, the duration of the CVR videos from The New York Times and BBC varies from two to four minutes. The reasons for this could be “the difficulties in storytelling and expense of production” (MacQuarrie and Steed, 2017: 46). Also, current HMD are not especially comfortable to wear (mainly because they are heavy), and there might be social isolation implications. Therefore, viewers might not be ready yet to spend longer times watching CVR content.

A well-written story is crucial to achieve an immersive experience in CVR content. As stated before, one of the main challenges for CVR content directors is the lack of control over viewers’ gaze directional behaviour, because they are free to look at any point in the viewing sphere. Other challenges that have been reported are the difficulties finding the right shots, hiding the crew kits and engaging the viewers (EBU, 2017). Due to these and other hindrances, a new grammar of filmmaking needs to be developed. In Dooley’s words (2017: 165): “Just as the filmmakers of the late nineteenth century took some time to experiment with screen grammar and establish the rules of narrative storytelling on the two-dimensional screen, so too are VR developers now exploring a new screen grammar for the 360-degree, interactive space.”

Different methods for directing attention are being explored and tested in CVR (Rothe *et al.*, 2017; Sheikh *et al.*, 2017; Mateer, 2017). Some of these techniques are based on movement, sound and lightning cues (Sheikh *et al.*, 2017). Also, moving objects could be used as a guiding strategy (Rothe *et al.*, 2017). Some techniques to direct viewers’ attention found in traditional filmic narrative content could also be used as suggested by Mateer (2017). He poses the example of *Schindler’s List* (1993) by Steven Spielberg, where the director uses the girl in the red coat to attract viewers’ attention. In this sense, “the VR writer not only guides the viewer through the story, but also through space” (Dooley, 2018: 102).

These and other CVR features such as the absence of a defined frame or different shots controlled by the director have a direct impact on subtitles. For example, the position needs to be defined without knowing how the background is going to look like, because it depends on where the viewer decides to look in the video. This could cause contrast issues or important parts of the video could be blocked by the subtitle. Moreover, if audio cues are an important technique for directors to attract attention, the location of the sounds should be made accessible to those viewers who cannot make use of the soundtrack (Agulló and Matamala, forthcoming). All in all, the grammar of CVR storytelling is being developed and, similarly, subtitling in this new environment is still to be defined.

3.2.2. Genres

According to the EBU report on VR (2017: 9), "a majority of the stories are either history or news and current affairs products." They also point out the potential of VR content for music (with binaural audio) and sport events, and to a lesser extent for fiction products and promotional material. Mateer (2017) also agrees that most CVR content features non-fiction stories, and he refers to promotional material, travel and sport events. He also provides some examples of action-based contents, such as "Lewis' Escape The Living Dead (2016) and Lin's HELP (2015)" (Mateer, 2017: 15). According to Grambart (2015), the current state of CVR can be compared to the early filmmaking, because both have started recording documentary or journalistic work. When cuts were introduced, and cinematography was born, filmmakers had the filmic language and strategies to tell more complex stories. CVR is now in that early period focusing on non-fictional content and a new cinematic grammar needs to be created in order to turn CVR into real storytelling (Grambart, 2015).

The current situation might be the reason why mostly broadcasters and journals such as The New York Times and the BBC are leading the creation of CVR content, telling real stories through what has been coined as immersive journalism (De la Peña *et al.*, 2010). According to the authors, immersive journalism "is the production of news in a form in which people can gain first-person experiences of the events or situation described in news stories" (De la Peña *et al.*, 2010: 291). The results of the corpus analysis, as will be seen later in the article, confirmed that most stories told are life stories or news, as well as documentaries tackling topics such as science and nature, arts and culture, travel or history. Fiction content is limited so far.

3.3. Subtitles in 360º content

Research in subtitles for immersive content is work in progress and some results have been published already (Agulló, 2018; Agulló and Matamala, forthcoming; Agulló *et al.*, forthcoming; Brown *et al.*, 2018; Fraile *et al.*, 2018; Montagud *et al.*, 2018; Rothe *et al.*, 2018). One of the main challenges to be solved in CVR narrative is how to attract viewers' attention, as explained before. In the case of subtitles, different challenges are being discussed, such as the position of the subtitles or speaker location methods.

The BBC Research & Development team has recently published a White Paper (Brown *et al.*, 2018) on subtitles in 360º videos. According to them, the main challenges presented by this new medium are caused by the freedom of movement that viewers have in the 360º sphere. This causes that "nowhere in the scene is guaranteed to be visible to the viewer" and "there will always be something behind the subtitle, and we may not know what this will be" (Brown *et al.*, 2018: 1). Other challenges are related to immersion, because preserving the immersive experience is vital for 360º content, and "it will be important that the subtitles disrupt this as little as possible" (*ibid.*, 1). Also, as there is the possibility that subtitles appear outside the Field of View (FoV) of the viewers, the effort to find the subtitles should be minimum. Finally, they also highlight the challenge of minimising VR sickness, because some viewers may experience dizziness when using VR and the implementation of the subtitles should not contribute to that (Brown *et al.*, 2018). For Rothe *et al.* (2018) the challenges are similar: position (where to place the subtitles), speaker location (and they insist that this is difficult for CVR content, because speakers can appear outside the FoV) and VR experience (related to the feeling of being immersed).

Taking into account these challenges and precedents, the BBC team designed and tested four different possible subtitle modes: (1) Evenly spaced: subtitles equally spaced by 120º in a fixed position below the eye line; (2) Follow head immediately: subtitles follow the viewer as they look around, displayed always in front of the them; (3) Follow with lag: the subtitle appears directly in front of the viewer and it remains there until the viewers look somewhere else; then, the subtitle rotates smoothly to the new position in front of the viewer; and (4) Appear in front, then fixed: subtitles appearing in front of viewers, and then fixed until they disappear (in this case, the subtitles do not follow the viewer if they look around). After this study in which 24 participants were involved, the BBC team concluded that the 'Follow head immediately' was the most suitable mode, because it gave the best overall user experience. According to the authors, this mode was easy to understand and gave the viewers the freedom to explore the video without missing any content (Brown *et al.*, 2018).

Rothe *et al.* (2018) also carried out an experiment following suggestions by the BBC. They focused on comparing two subtitle modes: static subtitles (subtitles that are anchored to the viewer's FoV, following their movements) and dynamic subtitles (subtitles that are dynamically placed within the 360º sphere). According to their study in which 34 participants were involved, the participants did not state a clear preference for any of the methods in the comparison part of the questionnaire. However, the questions about presence, sickness and workload favoured the dynamic subtitles.

Even if there is no clear solution, the challenges and different possibilities explored by previous studies open the path to keep improving subtitle integration in 360º content. Subtitles for the deaf and hard-of-hearing (SDH) features have not been researched in depth in previous studies, though. For example, a method needs to be designed to indicate the viewers where the speakers are located in the 360º sphere, so that they do not miss out the action. Some suggestions such as the usage of arrows or a compass have been made and tested (Agulló, 2018; Agulló *et al.*, 2018; Agulló and Matamala, forthcoming).

All in all, there are still open questions regarding subtitling features in immersive media, but a necessary first step is to analyse existing content, and this is the ultimate goal of this article.

3.4. Methodology: multimedia corpus

A multimedia corpus of CVR content was analysed, aiming to identify how titles, including both (sub)titles and other types of text on screen, were shown. It was not possible to foresee the type of elements that were to be found in the analysis, because subtitling practices are not standardised in CVR yet, and a closed sample of videos with (sub)titles was not available. Therefore, any type of textual element on screen was considered important for the analysis, except for video credits, company logos and title of the video, because these elements are generally not part of the narrative.

The approach to the analysis is qualitative, using the video as a unit and not each specific subtitle. For instance, when analysing the function of titles, the study did not extract the function of each individual title but analysed whether the pre-established functions were found or not in each video.

Considering that the medium is at a very early stage and there is a lack of standardisation, the interest of this study lies in identifying current practices and not quantifying their usage in videos. This analysis can be a departure point to start understanding the nature of subtitling in immersive content, the technical implications and shortcomings, and the possibilities offered.

3.4.1. Selection criteria

Audiovisual content developed in 360 degrees is not as available as 2D content, especially when it comes to professional quality. Also, accessing 360º videos and being able to process that information for analysis purposes is again not easy compared to 2D content. For example, easily play and pause the videos, extract subtitle tracks (most CVR videos present burnt-in subtitles) or take screenshots. A considerable amount of 360º videos can be found on YouTube platform, but not all of them are valid for research and even less include subtitles. Therefore, the first step in this analysis was to research different sources of 360º videos and define the selection criteria. The following criteria were considered: (1) videos should be created by professional, recognised broadcasters and/or producers; (2) videos should be CVR, that is, real images recorded with 360º camera sets, and not computer-generated image (CGI) content; (3) videos should be accessible for a wide audience (that is, they should be accessible from a PC or a smartphone, without the need of high-performance equipment); (4) videos should be non-interactive (which would exclude video games).

Following these criteria, it was decided to extract all the videos that were found in the NYT VR app and the BBC YouTube channels in April 2018. From the NYT VR App, 472 videos were selected on the 9th April 2018. From the 472 videos analysed in the NYT VR platform, two were discarded. One for not fulfilling the selection criteria (it was CGI created) and another for having been removed from the platform. The entire duration of the sample was 18 hours, 42 minutes and 42 seconds and the average duration per video was 2 minutes and 23 seconds. Those videos could be accessed via app in a smartphone or via browser (in YouTube or NYT VR Player) in any computer. From the BBC YouTube channels, 99 videos were selected on the 10th April 2018. From the 99 videos analysed in the BBC YouTube channels, six were discarded for not fulfilling the selection criteria (were CGI created). The entire duration of the sample was 6 hours, 23 minutes and 22 seconds, and the average duration per video was 3 minutes and 52 seconds (see Table 4). The videos could be accessed via YouTube in a smartphone or in a computer.

Broadcaster	No. of videos	Total duration of the sample	Average duration per video
The New York Times	472	18:42:42	00:02:23
BBC	99	06:23:22	00:03:52

Table 4. Summary of the analysed sample.

3.4.2. Data extraction process

Spreadsheets (Microsoft Excel) were used to collect, organise and analyse data. The identification of elements was based on Arnáiz-Uzquiza's taxonomy (2012), which proposes six different types of parameters: linguistic, sound extralinguistic, pragmatic, aesthetic, technical and aesthetic-technical. From the linguistic parameters, language features were considered (transcriptions –intralinguistic– or translations –interlinguistic–) and density (which includes character limitation, reading speed, etc.) was excluded, because this was not relevant for the study. Sound extralinguistic parameters were also included. Pragmatic parameters were not included, because they were not relevant in this case. Regarding the aesthetical parameters, font colour, font type, background box and position were considered. Finally, regarding technical and aesthetic-technical parameters, only the parameter implementation was included. Other general subtitling features such as character limitation, number of lines or segmentation rules were excluded, because subtitling in 360º content is at a very early stage, and the lack of standardisation would make the results irrelevant.

An initial data extraction process was tested with a limited number of videos and improved. The final structure gathered the following information for each video, having one tab for NYT VR content and one tab for BBC content.

1. A column for the title of the video, in an open field.
2. A column for the description of the video (provided by the broadcasters), in an open field.
3. A column for the duration of the video, in time format.
4. A column labelled "Text-on-screen?", to indicate if the video included any textual element. A picklist was provided (Yes, No).
5. A column labelled "Function". A preliminary analysis allowed to identify the following categories (included as a picklist):

a. Transcription: titles that included a written version of the voice-over.
b. Translation: titles that provided a written translation for the voice-over.
c. Speaker identification: titles that contain the name of the speaker.
d. Location identification: titles that contain the name of the location.
e. Directions: titles that contain directions to indicate the viewers to direct their gaze to a specific location (for example, ‘look up’).
f. Additional explanatory information: long titles including several sentences and/or paragraphs that add extra information about the main story of the video to inform the viewers.
g. Explanatory titles: short titles that include information other than speaker, location or directions.

6. A column labelled “Speaker identification”, to indicate the different strategies for character identification. A picklist was provided (Colours, Tags, Placement, Other (to be defined)).
7. A column labelled “Non-speech information”, to indicate the different types of non-verbal information. A picklist was provided (Sound effects, Music, Paralinguistic information).
8. A column labelled “Placement”. A picklist was provided (Evenly Spaced, Follow Head Immediately, Fixed position, Other (to be defined)). For the position of the titles, adapted version of the classification by Brown *et al.* (2018) is used:
    a. Evenly Spaced: subtitles are placed into the scene in different fixed positions, equally spaced around the video.
    b. Follow Head Immediately: subtitles are always visible in front of the viewer. If the viewers move their head, the subtitles move with them, always visible at the same location.
    c. Fixed position: subtitles are place into the scene in a single fixed position.
    d. Other (to be defined): this last category was added in case an unclassified type of implementation appeared.
9. A column labelled “Position”. A picklist was provided (Middle of the image, At the bottom, At the top, Next to the speaker, Next to an object or a person, On top of character(s), Integrated in an area of the image).
10. A column labelled “Font colour”, in an open field.
11. A column labelled “Background box”. A picklist was provided (Yes, No).

12. A column labelled "Graphic strategies", in an open field.
13. A column labelled "Font type". A picklist was provided (Sans-serif, Serif, Mixed).
14. A column labelled "Series", which identify if the video belonged to a specific series developed by the broadcasters, in an open field.
15. A column labelled "Category". This metadata would serve to understand which type of genres are the most popular among 360-degree content. The following picklist was created: News, Life Stories, Learning, Music, Sports, Travel, Science & Nature, History, Art, Culture & the Media, Comedy, Politics, Horror & Supernatural, Action & Adventure, Thriller, Crime, Drama, Promotional. This picklist is based on BBC genre classification[10]. The list was shortened for the sake of simplification, according to the most popular genres for 360-degree videos.
16. A column labelled "Date", which indicated the date of selection, in date format.
17. A column labelled "Timecode", which indicated the exact time in the video were the title appears, in time format.
18. A column labelled "Screenshot ID", in an open field. When analysing the videos, screenshots with their corresponding timecodes of all examples included in the corpus were created and stored.
19. A column labelled "Comments", in an open field.
20. A column labelled "URL", in an open field.

3.5. Results from a qualitative analysis

In this section, the obtained results are analysed separately: The New York Times and BBC content. This way, the different approaches taken by each broadcaster when implementing titles in 360º content can be seen, which present major differences as explained below.

From the 470 valid videos in the NYT VR platform, 436 videos presented textual elements that were included in the analysis. Therefore, 92.4% of the 360º videos generated by the NYT included titles. 1185 titles have been registered in the entire analysis process. From the 93 valid videos in the BBC YouTube channel, 37 videos included titles. Therefore, 39.8% of the 360º videos generated by the BBC presented textual elements. 74 titles have been registered in the entire analysis process.

[10] http://www.bbc.co.uk/programmes/genres

3.5.1. Results from NYT VR

Below, a summary of the results for each parameter is presented.

1. Function

The percentage of videos that included each function has been calculated (see Table 5). Please notice that the number of instances is not analysed but the fact the video included a function or not. Also, one video can include titles with more than one function.

Function	No. of videos	% (from total 436 videos with text on screen)
Location identification	400	91.7%
Additional explanatory information	277	63.5%
Speaker identification	231	53%
Explanatory titles	115	26.4%
Transcription	50	11.5%
Translation	46	10.5%
Directions	4	0.9%

Table 5. Information about function of titles in the videos from NYT VR app.

2. Placement

The percentage of videos that included each type of placement implementation has been calculated (see Table 6). One video can include more than one type of placement.

Placement	No. of videos	% (from total 436 videos with text on screen)
Evenly Spaced	420	96.3%
Fixed position	220	50.5%
Follow Head Immediately	0	0%

Other (to be defined)	0	0%

Table 6. Information about placement of titles in the videos from NYT VR app.

3. Position

The percentage of videos that included each type of positioning has been calculated (see Table 7). One video can include more than one type of placement.

Position	No. of videos	% (from total 436 videos with text on screen)
At the top	374	85.8%
Next to the speaker	150	34.4%
Middle of the image	111	25.5%
Next to an object or person	81	18.6%
At the bottom	61	14%
On top of character(s)	5	1.1%
Integrated in an area of the image	2	0.5%

Table 7. Information about position of titles in the videos from NYT VR app.

4. Speaker identification

In Table 8 below, the different speaker identification strategies found in the analysis are reported.

Speaker identification strategies	No. of videos	Description
Tags	2	The tag is displayed below the subtitle, indicating who is speaking, to differentiate the

		subtitle from others appearing before and after.
Placement	1	The subtitle is displayed next to the speaker.
Colours	0	
Other (to be defined)	0	

Table 8. Information about speaker identification strategies in the videos from NYT VR app.

5. Non-speech information

Only one occurrence from 1185 titles reads: “Crowd chanting”.

6. Font colour

From 1185 titles registered: 14 titles are black (1.2%); 1170 titles are white (98.7%); and 1 is white combined with yellow (0.08%).

7. Background box

From 1185 titles registered: 11 titles have background box (0.9%) and 1174 do not have background box (99.1%).

8. Graphic strategies

From 1185 titles, 53 (4.5%) present some graphic strategies. The variety of graphic effects implemented in the videos is considerable. There are some effects applied to the font type, such as the usage of customised font to highlight the title of the video, or fonts with metallic texture and sparkling effects. Also, different fading in and out effects are implemented. Some videos present creative or integrated titles in different forms, for example: titles integrated in the form of a newspapers layout; titles integrated in the image as if they were a poem; titles integrated inside a drawn map; titles integrated in the roof of the house that appears in the image; title integrated as if it was a monitor for the biker; titles integrated as if they were a recipe; titles integrated in rear and front parts of the plane and titles integrated in the form of a map. A video where three arrows are used to indicate who the speaker is was also found.

An interesting video in the corpus is titled "Sensations of Sound"[11]. The video features the story of Rachel, who gained partial hearing at age 20. She explains how she felt music, before and after receiving a cochlear implant. This video is about deafness, and they carefully created subtitles for it and all kind of integration strategies for titles can be found.

9. Font type

From 1185 titles, 26 (2.2%) use serif fonts, 1152 (97.2%) uses sans-serif fonts and 7 (0.6%) uses a combination of both.

10. Category

From the analysed 470 videos: 161 (34.3%) correspond to the category Life Stories; 67 (14.3%) fall under Science & Nature; 65 (13.8%) fall under the category Art, Culture & the Media; 52 (11.1%) videos correspond to the category Travel; 36 (7.7%) to News; 25 (5.3%) to Politics; 25 (5.3%) correspond to History; 17 (3.6%) to Music; 11 (2.3%) fall under Drama; 7 (1.5%) correspond to Sports; 3 (0.6%) are Promotional and 1 (0.2%) corresponds to Horror and Supernatural.

3.5.2. Results from BBC

Below, a summary of the results for each parameter is presented.

1. Function

The percentage of videos that included each function has been calculated (see Table 9).

Function	No. of videos	% (from total 37 videos with text on screen)
Explanatory titles	21	56.8%
Additional explanatory information	14	37.8%
Location identification	10	27%
Directions	7	18.9%

[11] https://www.youtube.com/watch?v=WOHFpm4w0Hc

Speaker identification	5	13.5%
Transcription	3	8.1%
Translation	0	0%

Table 9. Information about function of titles in the videos from BBC VR videos.

2. Placement

The percentage of videos that included each type of placement implementation has been calculated (see Table 10).

Placement	No. of videos	% (from total 37 videos with text on screen)
Fixed position	25	67.6%
Evenly Spaced	13	35.1%
Follow Head Immediately	0	0%
Other (to be defined)	0	0%

Table 10. Information about placement of titles in the videos from BBC VR videos.

3. Position

The percentage of videos that included each type of positioning has been calculated (see Table 11).

Position	No. of videos	% (from total 37 videos with text on screen)
Middle of the image	31	83.8%
Next to an object or person	12	32.4%
Next to the speaker	5	13.5%
At the bottom	2	5.4%

Integrated in an area of the image	2	5.4%
On top of character(s)	1	2.7%
At the top	0	0%

Table 11. Information about position of titles in the videos from BBC VR videos.

4. Speaker identification

In 2 out of 37 videos different colours (white, yellow, blue) have been used to differentiate the speakers.

5. Non-speech information

No strategy to include non-speech information in the text on screen has been used.

6. Font colour

From 74 titles registered: 13 titles are black (17.6%); 49 titles are white (66.2%); 7 (9.5%) combine different colours in the title: black/white, white/turquoise, yellow/grey, yellow/white, red/white/yellow, white/blue, white/grey; 4 are turquoise (5.4%) and 1 is yellow (1.3%).

7. Background box

From 74 titles registered: 30 titles have background box (40.5%) and 44 do not have background box (59.5%).

8. Graphic strategies

From 74 titles, 34 (45.9%) present some graphic strategies. Different strategies are implemented, for example: appearing and disappearing effects; fonts with a shadow effect; some appearing effects with boxes and lines pointing to the speaker and in one video the title follows a hawk.

9. Font type

From 74 titles, 5 (6.8%) use serif fonts and 69 (93.2%) uses sans-serif fonts.

10. Category

From the analysed 93 videos: 26 (28%) correspond to Science and Nature; 17 (18.3%) correspond to News; 15 (16.1%) correspond to Life Stories; 14 (15.1%)

correspond to Art, Culture and the Media; 6 (6.4%) correspond to Politics, 4 (4.3%) correspond to Sports; 4 (4.3%) correspond to Travel; 3 (3.2%) correspond to Music; 2 (2.1%) corresponds to Promotional; 1 (1.1%) correspond to Drama; and 1 (1.1%) correspond to History.

3.6. Current (sub)titling practices in immersive media

A discussion of the previous results can lead us to make some remarks regarding how titles in 360º videos are currently implemented. In general, titles have not been used to make the content accessible for viewers with hearing impairments or foreign language speakers. Textual elements have been mainly used to enhance the narrative of the videos in different ways.

As far as the genre of the videos, almost all videos were non-fiction. Fiction is not a widely explored genre for immersive content by NYT and BBC. In the NYT platform some examples are to be found: one video categorised as Horror and Supernatural called "Lincoln in the Bardo"; and a series of videos titled "Great Performers" categorised as Drama. In the BBC platform there is also one true crime video categorised as Drama called "360 murder scene in Tim Ifield's flat - Line of Duty: Series 4". The most recurrent genres are Life Stories, News and Science & Nature, followed by Art, Culture & the Media, Travel and Politics. Some videos about Music, History and Sports can also be found. These results seem to confirm that the most appealing genre so far for immersive content are those suitable for immersive journalism (De la Peña *et al.*, 2010), documentaries and potentially music and sport events.

Regarding function, location identification (91.7% for NYT and 27% for BBC), additional explanatory information (63.5% for NYT and 37.8% for BBC) and explanatory titles (26.4% for NYT and 56.8% for BBC) are the most common applications for titles in immersive content. Perhaps due to the importance of location in immersive videos, producers at NYT considered that it was relevant to indicate where the action takes place in 400 out of 470 videos, generally by including a title indicating the city or country at the beginning of the video. Also, the additional information and explanatory titles have been widely used in the analysed videos. This could be due to the fact that there is more space than in a 2D content to include textual elements, and editors felt tempted to add more information in the video to fill those blank spaces. Also, due to the fact that narrative in CVR is still at an early stage, using text to convey ideas or help narrative to make the story clearer to the audience might have been found as a useful strategy for content creators. The genre of the contents is mainly journalistic, which also explains that directors rely more on written text than image to express complex ideas, especially considering the restrictions in narrative strategies of CVR at this moment. Another possibility could be that because watching 360º content is a contemplative/immersive experience, including an off-screen voice sometimes could be distracting or even disruptive for immersion (the off-screen voice reminds the viewer that they are not actually there), and written text could be considered less intrusive to the experience.

Following with titles' function, it has also been noticed that when a translation is needed, NYT used (sub)titles and BBC used audio subtitles. NYT also included transcriptions of the speakers or narrators in form of titles, mainly when the speaker was off-screen. It was interesting to find titles that indicated the viewer where to look at; for example, 'Look up' or 'Look down'. This is a very innovative application for titles in audiovisual products and confirms that titles in immersive content present a different behaviour than in other audiovisual media.

Regarding placement, titles were mostly placed evenly spaced (97.5% for NYT and 35.1% for BBC) or in a fixed position (50.9% for NYT and 64.9% for BBC). Titles that followed viewers' head immediately were not found, which was the preferred option according to users in previous studies (Brown *et al.*, 2018). Regarding the position, surprisingly, at the top (86.5% for NYT and 0% for BBC), next to the speaker (34.4% for NYT and 13.5% for BBC) or next to an object or a person (19% for NYT and 32.4% for BBC) and in the middle (25.5% for NYT and 83.8% for BBC) of the image were the most used areas, instead of at the bottom (14% for NYT and 5.4% for BBC), as it usually happens with subtitles in 2D content. This could give some hints about the best position to integrate titles in 360º content. The reason for that could be to avoid the users looking down for the subtitles, which with the current heavy HMD equipment could provoke cervical discomfort. Also, it could be due to the type of content or the filmic nature of CVR. In the examples that were analysed, visual elements at the bottom area of the FoV are usually relevant in comparison to the top area, where background images (ceilings, sky, etc.) usually appear. Moreover, some contents are filmed in a first-person perspective and subtitles positioned at the bottom could be more intrusive to the experience, interfering with the perception of the viewers of themselves inside the virtual world. This would be even more relevant in the case of virtual reality in video games, where peripherals can be used to visualise the virtual hands in the virtual world.

SDH features such as speaker identification and non-speech information were only present in three videos of NYT and two videos of the BBC, probably because it was not the main intention of the producers to make the videos accessible.

Regarding font colour, white was the most used colour (98.7% for NYT and 66.2% for BBC), as it is for subtitles in most 2D subtitled content. The usage of background boxes was not very common in NYT videos (only 0.9% of the videos), which sometimes made reading the titles a cumbersome task. For BBC content, 40.5% of the titles included a background box.

As for graphic strategies, NYT used them mainly for aesthetical purposes in 4.5% of its videos, to improve the video content or the story. BBC used it more (45.9% of the videos), but apparently without a specific intention which might be considered distracting in some cases. In general, the integration of titles in the CVR content was more creative and aesthetical than in 2D content. An example is the graphic strategies found in the video "Sensations of Sound" by NYT, as an example of what could be done. Some strategies that were found in the video: (1) words appear one by one, and they are vertically positioned next to the speaker; (2) sets of words appear one by one as the speakers pronounce them, and they are vertically positioned next to the speakers; (3) some words (for example, 'blurred') are emphasised by adding extra spacing between characters; (4) some titles appear integrated with illustrations (a figure playing the piano or the guitar) that somehow represent what is being said; (5) some titles appear following the rhythm of a metronome; (6) the word 'vibration' is emphasised by being represented with a vibration effect; (7) some titles appear inside a comic bubble; (8) the word 'jump' is emphasised by using a zoom in/zoom out effect, like if the word was actually jumping; (9) some titles are integrated inside a stave for musical notes; (10) some titles appear integrated inside a stave, but vertically, as if each word were a musical note; (11) the word 'world' is emphasised with a flickering effect; (12) the word 'signing' is emphasised by appearing the letters one by one, probably for the viewer not to mix it with 'singing', because the context could be misleading.

Also, in general most titles were located in dynamic positions within the 360º sphere, or close to objects or persons. The potential to integrate titles in a more innovative way has been confirmed.

3.7. Conclusions

Filmmakers are experimenting with different techniques to create CVR content, and subtitle creators and researchers are trying to find the best way to implement (sub)titles in this new medium. In this article, the status of immersive content and the current challenges that filmmakers face when dealing with this new medium has been reviewed. It is important to understand how CVR content works in order to create usable, immersive and accessible subtitles. Researchers have started to design and test different subtitle modes for CVR content reaching different conclusions (Brown *et al.*, 2018; Rothe *et al.*, 2018). To complement their findings, the results of a corpus analysis to illustrate the current subtitle practices carried out by The New York Times and the BBC have been presented: to this end what has termed globally as 'titles' by the authors has been analysed. The results have shed some light on important issues such as the positioning of titles, innovative graphic strategies or the different functions.

The present findings have challenged the current practices of subtitling in other audiovisual content. For example, most titles are positioned at the top or in the middle of the image, contradicting standardised practices of positioning (sub)titles at the bottom-centred of the screen. Usually, subtitles are positioned at the bottom in 2D traditional contents because there are fewer visual elements that are relevant for the narrative and it is less disruptive. However, this is still to be decided in the case of CVR content. The scene compositions in CVR might differ from films or series in other media and, therefore, subtitles practices might need to be adapted depending on the new content. All in all, further testing is necessary to give an answer to all these questions.

The same happens when deciding between the two main subtitle modes that have been tested and designed: 'follow head immediately' or 'evenly spaced'. Results from previous tests (Brown *et al.*, 2018; Rothe *et al.*, 2018) are contradictory and therefore inconclusive. Both methods have advantages and disadvantages and the choice will probably depend on the type of content and the target audience. For example, if the action remains in a specific location in the sphere, 'evenly spaced' or 'fixed positioned' subtitles might be more immersive and integrated in the experience. However, if the action is fast or several characters are simultaneously speaking, subtitles that are always visible and 'follow head immediately' might be more suitable. It is still to be confirmed whether viewers would accept both methods or would rather prefer a consistent solution across all immersive content.

The impact of subtitles on the immersive experience also needs to be further tested. In previous studies, some differences have been found between static (subtitles that follow head movements) and dynamic (subtitles that are fixed in different positions within the video), achieving the latter higher levels of presence (Rothe *et al.*, 2018). As explained at the beginning, the main goal of CVR content is to create immersive experiences. Some researchers claimed that subtitles that are immersive and not disruptive to this experience should be implemented (Brown *et al.*, 2018). However, it can be argued that subtitles will never be disruptive for those who need them. It is much more disruptive not being able to hear what is being said or not being able to understand a foreign language. Therefore, the debate about the disruptiveness of subtitles is sterile, as some studies have shown (Wissmath *et al.*, 2009).

Further testing on the different subtitling modes, as well as positioning (top, bottom, in the middle) and speaker location methods for SDH are necessary to start creating guidelines. How the different modes impact on immersion also needs to be further researched. Other subtitling parameters such as reading speed or number of characters and lines per subtitle should be reconsidered again for CVR, because a new medium can bring new challenges for the viewers. Eye-tracking studies would also shed some light in this topic. VR is indeed a dangerous medium, for filmmakers and audiovisual translators. But one that is worth exploring.

References

AGULLÓ, Belén, & Anna MATAMALA, forthcoming: "The challenge of subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group", *The Journal of Specialised Translation* 32.

AGULLÓ, Belén, Anna MATAMALA & Pilar ORERO, 2018: "From disabilities to capabilities: testing subtitles in immersive environments with end users", *HIKMA* 17, 195-220.

AGULLÓ, Belén, 2018: "SDH in Immersive Environments: Can Subtitles be Immersive at all?", paper presented in Languages & The Media 2018.

ARNÁIZ-UZQUIZA, Verónica, 2012: *Subtitling for the deaf and the hard-of-hearing: Some parameters and their evaluation*. PhD dissertation, Universitat Autònoma de Barcelona.

BROWN, Andy *et al.*, 2018: "Exploring Subtitle Behaviour for 360° Video", White Paper WHP 330, BBC Research & Development, April 2018 [available at https://www.bbc.co.uk/rd/publications/whitepaper330, date of retrieval: March 23rd, 2019].

DE LA PEÑA, Nonny *et al.*, 2010: "Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News", *Presence: Teleoperators and Virtual Environments* 19 (4), 291-301.

DOOLEY, Kath, 2017: "Storytelling with virtual reality in 360-degrees: a new screen grammar", *Studies in Australasian Cinema* 11 (3), 161-171.

—, 2018: "Virtual reality: Moving beyond a cinema of attractions", *Screen Education* 88, 97-103.

EUROPEAN BROADCASTING UNION (EBU), 2017: "Virtual Reality: How are public broadcasters using it?" Report, July 2017 [available at https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it, date of retrieval: March 23rd, 2019].

FERRARI, Alex, 2016: "IFH 123: What is Cinematic Virtual Reality with Jason Diamond", December 15, 2016, in *Indie Film Hustle*, podcast, MP3 audio, 01:13:56 [available at https://indiefilmhustle.com/cinematic-virtual-reality/, date of retrieval: March 23rd, 2019].

FOERSTER, Anna, 2010: "Towards a creative approach in subtitling: a case study." in Jorge DÍAZ CINTAS, Anna MATAMALA & Josélia NEVES (eds.): *New Insights into Audiovisual Translation and Media Accessibility. Media for All 2*, Amsterdam: Rodopi: 81-98.

FOX, Wendy, 2016a: "Integrated titles: An improved viewing experience?" In Silvia HANSEN-SCHIRRA & Sambor GRUCZA (eds.): *Eyetracking and Applied Linguistics*, Berlin: Language Science Press: 5-30.

—, 2016b: "Should she really be covered by her own subtitle?", *Translation Spaces* 5 (2), 244-270.

FRAILE, Isaac *et al.*, 2018: "Personalized and Immersive Presentation of Video, Audio and Subtitles in 360º Environments: An Opera Use Case", paper presented in ACM TVX 2018 [available at https://figshare.com/articles/Demo_Personalized_and_Immersive_Presentation_of_Vide o_Audio_and_Subtitles_in_360o_Environments_An_Opera_Use_Case/6527531, date of retrieval: March 23rd, 2019].

GÖDDE, Michael *et al.*, 2018: "Cinematic Narration in VR – Rethinking Film Conventions for 360 Degrees". In Jessie Y. C. CHEN and Gino FRAGOMENI (eds.): *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry. VAMR 2018. Lecture Notes in Computer Science, vol 10910*, Cham: Springer: 184-201.

GRAMBART, Stefan, 2015: "Sleepy Hollow & Narrative in VR", December 10, 2015, in *SlideShare*, YouTube video, 34:15 [available at https://www.slideshare.net/fitc_slideshare/sleepy-hollow-narrative-in-vr, date of retrieval: March 23rd, 2019].

HEETER, Carrie, 1992: "Being there: The subjective experience of presence", *Presence: Teleoperators and Virtual Environments* 1 (2), 262-271.

LEE, Daniel G., Deborah I. FELS, & John Patrick UDO, 2007: "Emotive Captioning", *ACM Computers in Entertainment* 5 (2), 1-15.

MACQUARRIE, Andrew, & Anthony STEED, 2017: "Cinematic Virtual Reality: Evaluating the Effect of Display Type on the Viewing Experience for Panoramic Video", paper presented in 2017 IEEE Virtual Reality (VR).

MATEER, John, 2017: "Directing for Cinematic Virtual Reality: How the Traditional Film Director's Craft Applies to Immersive Environments and Notions of Presence", *Journal of Media Practice* 18 (1), 14-25.

MCCLARTY, Rebecca, 2012: "Towards a multidisciplinary approach in creative subtitling", *MonTI. Monografías de Traducción e Interpretación* 4, 133-153.

—, 2014: "In support of creative subtitling: contemporary context and theoretical framework", *Perspectives* 22 (4), 592-606.

MONTAGUD, Mario *et al.*, 2018: "ImAc: Enabling Immersive, Accessible and Personalized Media Experiences", paper presented in ACM TVX 2018 [available at https://dl.acm.org/citation.cfm?id=3213570, date of retrieval: March 23rd, 2019].

ROTHE, Sylvia, Heinrich HUSSMANN, & Mathias ALLARY, 2017: "Diegetic cues for guiding the viewer in cinematic virtual reality", paper presented in 23rd ACM Symposium on Virtual Reality Software and Technology.

ROTHE, Sylvia, Kim TRAN, & Heinrich HUSSMANN, 2018: “Dynamic Subtitles in Cinematic Virtual Reality”, paper presented in ACM TVX 2018 [available at https://dl.acm.org/citation.cfm?id=3213556, date of retrieval: March 23rd, 2019].

SHEIKH, Alia *et al*., 2017: “Directing attention in 360-degree video”, paper presented in IBC 2016 [available at https://www.ibc.org/production/directing-attention-in-360-degree-video/840.article, date of retrieval: March 23rd, 2019].

SPIELBERG, Steven, dir. *Schindler’s List*. 1993; USA. Universal Studios, 2013. Blu-ray Disc, 1080p HD.

VR INTELLIGENCE, 2017: “Virtual Reality Industry Survey”, report published in 2017.

WISSMATH, Bartholomäus, David WEIBEL, & Rudolf GRONER, 2009: “Dubbing or Subtitling? Effects on Spatial Presence, Transportation, Flow, and Enjoyment”, *Journal of Media Psychology* 21 (3), 114-125.

Chapter 4: Article 3. Technology for subtitling: a 360-degree turn

Technology for subtitling: a 360-degree turn

Belén Agulló, Universitat Autònoma de Barcelona

Subtitling has become one of the most important audiovisual translation modes and cannot be understood outside the context of the technology that makes it possible. New audiovisual media are appearing, such as 360º videos, and the necessity of subtitling this type of content to make it accessible is emerging. In this article, an updated review of current subtitling technology is presented to contextualise the study. Then, a review of main immersive environments (3D, augmented reality and virtual reality) and their implications for subtitling has also been introduced. The focus of the study is on virtual reality and, therefore, the main challenges of subtitling 360º content are presented. To respond to the needs of subtitling this type of content, a prototype version of a subtitle editor has been developed and presented to twenty-seven professional subtitlers who have tested the tool and reported the correspondent feedback on usability and preferences. This study has proven the importance of carrying out usability tests with end users when developing specific software. Finally, the challenges faced by subtitlers in new audiovisual media such as 360º content are presented.

subtitling, subtitling technology, 360º content, reception study, usability

4.1. Introduction

We are living a subtitling revolution. The distinction between subtitling and dubbing countries is becoming more and more outdated. Nowadays, on certain platforms the choice between subtitled or dubbed versions is made individually by each viewer and not imposed by distribution companies. The key to this revolution is the new distribution model for audiovisual content based on Video on Demand (VoD) platforms such as Netflix, Prime Video, HBO, Hulu and the ones to come, such as Disney+ or Apple TV+. In these platforms, the viewers can access content from all over the world, in different languages and with different audiovisual translation modes available, mainly subtitling and dubbing. Subtitling has become such an important element in audiovisual communication that has recently been the focus of a heated debate. Netflix decided to include Castilian Spanish subtitles for the film *Roma* (2018), which was filmed in Mexican Spanish, by the Mexican award-winning film director Alfonso Cuarón. Some viewers and linguists disagreed with this approach, and Cuarón himself considered the subtitles "provincial, ignorant and offensive to Spaniards themselves" (Cuarón quoted in Morales *et al.*, 2019). Due to social pressure and the comments from the director, Netflix ended up removing the Castilian Spanish subtitles for this film.

Changes are also being introduced in the way audiovisual stories are created. Recently, the first interactive series episode was launched by Netflix, accessible via computers, smartphones or Smart TV. The episode called 'Bandersnatch' is part of the science fiction series *Black Mirror*. The main character of the episode set in the 80s is an aspiring video game programmer who is working on coding a game based on a choose-your-own-adventure novel called 'Bandersnatch'. In this episode, the viewer has the control over the main character's decisions through the story, leading to a myriad of different endings (Harris *et al.*, 2019). Also, new technology such as virtual reality (VR) or augmented reality (AR) are inspiring content creators and storytellers to develop engaging stories. The European Broadcasting Union (EBU) issued a report on VR/AR (EBU, 2017) stating that 49% of its members are developing or planning to develop immersive content. Members believe that VR offers new opportunities to tell stories from a different perspective and could be more engaging for the audience. In a more recent survey on the industry insights of VR and AR, results show that the interest in VR/AR applications for films and series has dropped since last surveys (Perkins Coie LLP, 2018). According to the report, this could be due to the belief that shorter experiences work better for AR/VR. However, in the same report, they state that video games sector is still leading the list of the industries with more investment in VR/AR. This seems contradictory, because video games are long audiovisual experiences. Therefore, the reasons behind the drop in investments for VR/AR in audiovisual content other than video games might be different (for example, poor quality of user experience, lack of engaging content or hardware high prices). All in all, these new formats and technologies applied to audiovisual content creation may have an impact on how content is subtitled, and technology should be prepared to face the new challenges.

Academic studies on subtitling technology are limited with a few exceptions (Georgakopoulou, 2012; Díaz-Cintas, 2013, 2014). However, the importance of technology in the practice of subtitling is paramount, being technology necessary for its existence and further development. Subtitling software has rapidly evolved for the past few years, giving response to this new audiovisual and digital society, which has led to a "technology turn" in audiovisual translation (AVT) and, specifically, in subtitling (Díaz-Cintas, 2013). Many are the innovations introduced in subtitling software, mainly focused on automation to increase productivity and satisfy the ever-growing need for subtitles in different contexts and situations, as it will be detailed in the next section. Moreover, new audiovisual formats, such as VR or AR, introduce new dimensions and challenges for AVT that need to be tackled by the development of the appropriate technology. To that end, a prototype of a subtitle web editor for 360º videos has been developed in the framework of the European-funded project Immersive Accessibility (ImAc).[12] The main goal of ImAc is to make immersive content accessible to all kind of users. To do so, a first step was to develop different technologies to produce the access services for this new medium including subtitles, because to the best of our knowledge no commercial solutions were available at this point. A first version of a subtitle editor for 360º content has been developed. This prototype has been tested with twenty-seven professional users, with the aim of gathering their feedback regarding the technological needs of subtitlers working with immersive content.

In this article, an overview of current subtitling technology will be offered. Then, the challenges posed by immersive content in terms of subtitling are reviewed. In section 4, the prototype of the ImAc subtitle web editor will be presented and discussed. In section 5, the methodology and the results will be explained. Finally, some conclusions will be drawn.

4.2. Subtitling technology: overview

The change of analogue signal to digital broadcasting, the appearance of DVD and Blu-Ray and the universal access to internet have all contributed to the growth of subtitling and subtitling technology (Díaz-Cintas, 2014). Nowadays, both intralingual and interlingual subtitles are available a button away from the audience, which has caused what Díaz-Cintas labelled the "commoditisation of subtitling" (Díaz-Cintas, 2013, 2014). The consolidation of digital television and the proliferation of VoD platforms as a mainstream way of consuming audiovisual content and the accessibility approach taken by them, offering subtitles, subtitles for the deaf and hard-of-hearing (SDH), dubbing and audio description in several languages, has given a new boost to audiovisual translation. The demand for digital video content continues to increase and forecasts point to a growth

[12] www.imac-project.eu/

also in media localisation that will become a USD 2.5 billion industry by the end of 2020 (Estopace, 2017). The needs generated by this exponential increase in content and localisation demand will only be met with the support of technology that is able to adapt to the new challenges.

In previous studies, Díaz-Cintas (2013, 2014) offered an extensive review of subtitling technology advances up to that date. From the analogue subtitling practices from the 70s to the digital revolution that brought new advancements to this field, the practice of subtitling has evolved. Nowadays, with the professional software available, subtitlers are able "to create the timecodes, to respect shot changes, to control the reading speed of their subtitles, to translate and spell check their texts, to simulate their subtitles on screen, etc." (Díaz-Cintas, 2013: 124). Some examples of professional subtitling software are: EZTitles, FAB, Screen Subtitling, Swift, Spot or WinCAPS. However, the high prices of professional tools favoured the appearance of free software such as Subtitle Workshop, Aegisub or Subtitle Edit. The latter are mostly used by fansubbers to create their own subtitles for the fan community. Díaz-Cintas (2014) distinguishes between fansubs and crowdsourced subtitling. One of the main differences is that fansubbers produce subtitles without the consent of the content owners, and crowdsourced subtitles are distributed with the consent of the interested parties, without copyright infringements. Some platforms that encourage volunteers to create crowdsourced subtitles are Amara, TED or even YouTube. These platforms offer cloud-based subtitling platforms for the volunteers to create subtitles, but with very limited options compared to professional tools. Nonetheless, the trend of cloud-based professional subtitling tools to streamline the creation and distribution of subtitles is noticeable. For example, Zoosubs, imediatrans or OOONA offer an integrated, cloud-based workflow for localising audiovisual content.

The increase of audiovisual content demands subtitling as a commodity. Therefore, media companies are continuously looking for solutions to streamline their processes and be able to cope with the growing demand. Subtitling technology is including new features to respond to industry expectations, mainly based on automation. According to Díaz-Cintas (2014), some professional subtitling tools include an automatic detection of shot changes, facilitating the spotting task. Also, most programmes include an audio wave indicator that shows the audio behaviour in the video. Subtitlers can then skip the parts of the video where the dialogue is missing and go directly to the interest points, especially during quality assurance stage. Also, it is helpful for spotting, because subtitlers can visually see where the subtitle should start and end (Díaz-Cintas, 2014). Speech alignment technology can facilitate the task even more, automatically synchronising the transcription with the soundtrack and the video. Another step forward in automation comes from automatic speech recognition (ASR) software. Some subtitling software, such as Omniscien, can transcribe the audio, extract the dialogue and create a subtitle draft, even in the target language with machine translation (MT) technology (Dranch, 2018).

Several studies have been carried out with the aim of creating live subtitles automatically (Garcia *et al.*, 2009; Álvarez *et al.*, 2010, Mathur *et al.*, 2015). Some technologies that are key to the generation of automatic subtitles are: voice activity detection, ASR, discourse segment detection and speaker diarisation (Álvarez *et al.*, 2010). When this technology is put together in a single system, the creation of machine generated live subtitles is possible. A study on the reception of automatic generated and translated subtitles was carried out by Matamala *et al.* (2015). In the main experiment, 30 participants with different levels of English watched three randomised clips in English: one with no subtitles, one with automatic generated intralingual subtitles and another with automatic generated interlingual subtitles into Spanish. The conclusions of the study showed that intralingual subtitles can be beneficial to increase language comprehension for viewers with a medium level of English (B2). However, interlingual subtitles had a negative impact on C2 level viewers, probably due to a distracting effect.

The necessity of making live events, such as conferences, talks or even lessons, accessible to audiences with hearing impairments has promoted the implementation of ASR systems in several contexts. For example, about 1,500 deaf and hard-of-hearing (HoH) students are an integral part of the campus at the Rochester Institute of Technology. Apart from American Sign Language (ASL) interpreters, they have implemented Microsoft Translate, a communication technology that uses Artificial Intelligence technology to generate subtitles (Roach, 2018). This system uses advanced ASR technology to convert speech into fluent, correctly punctuated text, which can be automatically translated into the 60 languages that the system supports (with the quality that MT currently offers). The deaf and HoH students agreed that this was a useful tool to complement the ASL interpretation during lessons. Also, professors at Rochester Institute of Technology recognised the potential of this tool for education, because all students (deaf or hearing) used the transcripts as a learning and course material (Roach, 2018). Similar initiatives are being reproduced in other universities, such as the app poliSubs developed by the Universitat Politècnica de València.[13] Previous attempts have also been made at a European level, such as the EU-funded project transLectures[14] (2011-2014), which aimed at developing tools for automatic transcription and translation for online educational videos.

Open and free solutions to generate automatic subtitles with ASR technology can be found in YouTube[15] and the app Web Captioner.[16] YouTube integrates an option to generate automatic subtitles in English, Dutch, French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish. When a video is uploaded in YouTube, it can be edited, and transcriptions and translations can be added. Automatic transcriptions with timings are generated and can then be postedited through a basic interface. In the interface, the subtitler can add and delete subtitles, modify the time codes, modify subtitle content, previsualise subtitles in the video and use the audio wave for a more accurate spotting. A list of keyboard shortcuts is also provided to streamline the process. The transcribed subtitles can be then translated to as many languages as desired and be added to the video. Subtitles generated in YouTube can also be downloaded in different formats (.vtt, .srt, .sbv). Although the quality of the automatic subtitles varies depending on the type of content, the quality of sound, background noises, different accents, etc., this could be a powerful tool to increase accessibility. Web Captioner is a free real-time subtitling app for browsers based on speech-to-text transcription technology. It supports over 40 languages and dialects and only a computer with internet connection and a microphone is necessary. Subtitles can be then downloaded in a plain text format .txt or .doc (without time codes).

The use of computer-aided translation (CAT) tools, even if present in many translation fields such as technical or medical translation, is limited in the field of AVT. It is argued that CAT tools are not suitable for AVT due to the culturally-bond and creative types of content. However, CAT tools are widely used in game localisation (O'Hagan and Mangiron, 2013), which also includes dialogue text to be dubbed and subtitled and creativity and cultural references are involved in the translation process. CAT tools were introduced in game localisation to boost productivity due to the large

---

[13] https://www.elmundo.es/comunidad-valenciana/2018/12/26/5c235fdefc6c83c3018b456e.html
[14] https://www.mllp.upv.es/projects/translectures/
[15] https://support.google.com/youtube/answer/6373554?hl=en
[16] https://webcaptioner.com/

amount of words and the limited time to translate the content. Also, to maintain consistency across projects where teams of several translators work simultaneously thanks to translation memories (TM) and glossaries integrated in the tools. Consistency and intertextuality are also common and necessary in films and series. Different translators can work in different seasons of the same series without having the reference material. So far, the only way to ensure consistency is to watch previous episodes or look for previous subtitle files, but this could be very time-consuming. Having all the translations in a TM would significantly ease the translator's work. Anticipating industry's needs, some CAT tool companies such as memoQ have introduced some features to facilitate the task of the subtitlers. For example, the tool can process specific subtitle formats such as .srt. Also, the translator can previsualise the video with the translated subtitles in the same tool with a simple plugin, as can be seen in Figure 6.

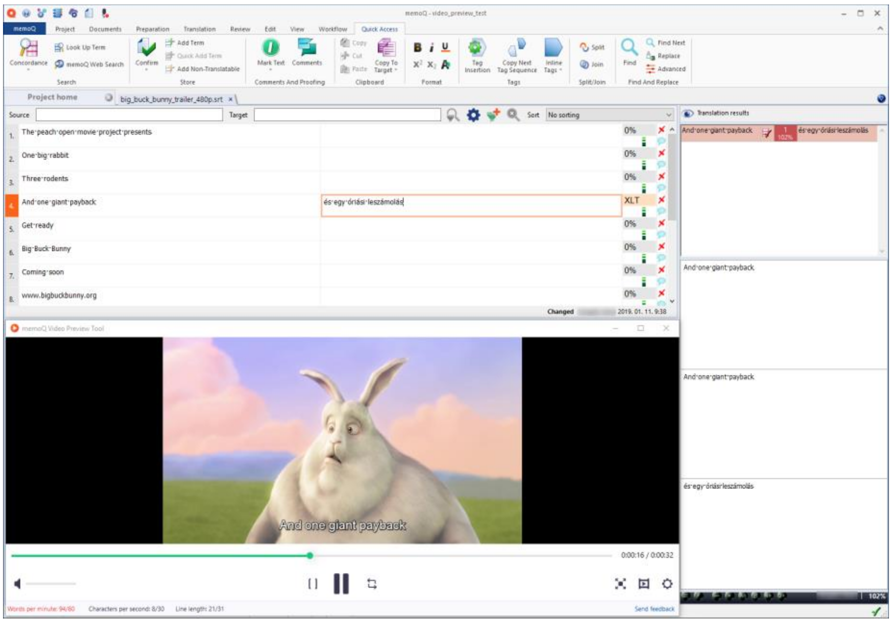


Figure 6. memoQ translation panel with video previsualisation.

The next step in automation would be to introduce (MT) in the subtitling workflow. Several funded projects aimed at achieving automatic subtitling with automatic translation have been developed in the last years such as MUSA[17] (2004-2007) and SUMAT[18] (2011-2014). The former's ambitious goal was to create a whole automatic system that converted audio into text, then generated the time-coded subtitles and, finally, translated the subtitles into other languages. However, no tangible results were ever materialised (Díaz-Cintas, 2014). The latter's main goal was "to use the archives of

[17] http://sifnos.ilsp.gr/musa/
[18] http://www.fp7-sumat-project.eu/

subtitle files owned by several subtitling companies to build a large corpus of aligned subtitles that will be used to train SMT [statistical machine translation] systems in various language pairs" (Bywood *et al.*, 2013: 596). An evaluation with trained subtitlers was carried out to test the quality of the MT-generated subtitles. Although the study has some limitations, according to their results, "subtitlers considered 56.79% of the subtitles [27,565 MT subtitles] they rated to be of quality 4 or 5, meaning that they required little to no post-editing to be of a publishable standard" (Bywood, Georgakopoulou and Etchegoyhen, 2017: 497). Even if results are far from optimal, they should not be disregarded. It is worth noticing that when SUMAT project was developed, the MT system was based on statistical MT and, nowadays, research on the MT field is focused on neural MT, which is the new paradigm for MT (Forcada, 2017). Applying neural MT might result into different results and further research is necessary to clarify the future of MT in AVT.

4.3. Immersive media and subtitling

Immersive media aim at making the audience feel immersed in a story, as if they were transported to another reality. One example of immersive medium is stereoscopic 3D, which has been present in homes and cinemas during the past decades. The blockbuster *Avatar* (2009) by James Cameron that was projected in cinemas in stereoscopic 3D all over the world caused a comeback of this technology that was already popular during the 50s (Agulló and Orero, 2017). However, accessibility was considered as an afterthought in this medium, and the implementation of subtitles was part of the postproduction stage. Due to that and the lack of specialised subtitling technology for stereoscopic 3D, some issues raised when integrating subtitles. Superimposing 2D subtitles on a 3D image provoked undesirable effects such as ghosting, hindering the readability of subtitles and causing fatigue and eye strain (Agulló and Matamala, 2019).

Other immersive medium still under research and development is AR, which combines real world objects with superimposed virtual objects (Agulló and Matamala, 2019). The applications of this technology are very promising in the field of accessibility, especially in theatres and cinemas. A project led by the researcher Romero-Fresco to use smart glasses to display live subtitling in theatre has been developed by the GALMA Research Group[19] together with the National Theatre in London.[20] They use AR and

[19] http://galmaobservatory.eu/projects/captioning-glasses-for-the-national-theatre/
[20] https://www.nationaltheatre.org.uk/your-visit/access/caption-glasses

speech recognition technology to implement live subtitles for their plays. Glasses are adaptable and subtitles customisable.[21] This is an ongoing project and evaluation is still not available, to the best of our knowledge. If this technology is fine-tuned and mainstreamed might become a revolution in accessibility.

Finally, the immersive medium under research in this study is VR, specifically 360º videos. This type of content can be watched in head-mounted displays (HMD), giving the viewer the freedom to look around, usually with 180º of freedom in the field of view (FoV). This new format poses different challenges for implementing subtitles. Firstly, unlike in traditional TV/cinema content, the frame for each scene depends on where the viewers decide to look at any moment, so there is no way to guarantee that a specific area of the 360º video is seen by them. Also, the background image where the subtitles will be displayed cannot be foreseen either, which can cause contrast issues and have a negative impact on readability if the subtitles are not well produced. Therefore, one of the main issues that need to be solved is where to locate the subtitles. Secondly, if the speaker is outside the FoV, this needs to be indicated somehow, to make the content accessible for deaf and HoH viewers as well. And thirdly, the implementation of subtitles must not disrupt the VR experience, that is, subtitles should not break immersion and should not worsen the VR sickness effect (i.e., viewers feeling dizziness, headache or eye fatigue because of the consumption of VR content).

Some studies have been already carried out regarding implementing subtitles in 360º content (Agulló *et al.*, 2018; Brown *et al.*, 2018; Rothe *et al.*, 2018; Agulló and Matamala, 2019). Some preliminary studies were carried out in order to gather feedback from end-users on how to implement subtitles in 360º contents (Agulló *et al.*, 2018; Agulló and Matamala, 2019). Results from a focus group in Spain showed that end-users would like to receive subtitles as similar as possible as shown in TV contents. For example, participants suggested using the current Spanish standard for SDH (AENOR, 2003). They also suggested locating subtitles in a fixed position in relation to the FoV and highlighted the importance of using a black background box in order to avoid readability issues. Regarding providing directions to enhance accessibility, participants suggested including icons (arrows) or text in brackets (to the left, to the right), as well as the possibility to include a compass or radar, in order to indicate where the speaker is located in the 360º video (Agulló and Matamala, 2019). In a different study, a prototype

[21] https://www.youtube.com/watch?v=Hdtf4qUWos4

was tested with a limited number of participants in order to gather information on the comfortable field of view for reading subtitles, and the speaker location identification. Also, a new methodological approach for accessibility studies is proposed, based on capabilities instead of disabilities (Agulló *et al.*, 2018). Both in Brown *et al.* and Rothe *et al.* studies, they compared different subtitling modes in order to evaluate which was the preferred solution. In the first study by the BBC team (Brown *et al.*, 2018), they compared four different modes: (1) Evenly spaced: subtitles equally spaced by 120º in a fixed position below the eye line; (2) Follow head immediately: subtitles follow the viewer as they look around, displayed always in front of them; (3) Follow with lag: the subtitle appears directly in front of the viewer and it remains there until the viewers look somewhere else; then, the subtitle rotates smoothly to the new position in front of the viewer; and (4) Appear in front, then fixed: subtitles appearing in front of viewers, and then fixed until they disappear (in this case, the subtitles do not follow the viewer if they look around). In their study conducted with 24 participants, they concluded that the (2) option was the preferred one. In the study by Rothe *et al.* (2018), they tested two types of subtitles with 34 participants: static subtitles (that is, subtitles that are always visible in the viewer's FoV) and dynamic subtitles (that is, subtitles that are in a fixed position close to the speaker). Even if the participants did not state a preference in the comparison part of the questionnaire, dynamic subtitles seemed to have better results in the questions about presence, sickness and workload.

Even if some previous studies have been carried out about the implementation of subtitles in 360º videos, to the best of our knowledge there is currently no specific subtitling software for this type of content, and subtitles so far have been created manually. Therefore, there is a need to develop a subtitling software that gives response to the current challenges that have been discussed and that is usable and accessible to professional subtitlers. For that reason, it was decided in the ImAc project to develop a prototype of a subtitle web editor that will be discussed in the following section.

4.4. ImAc web editor

The prototype of the ImAc subtitle web editor has been developed with the aim of producing accessibility services, specifically subtitles, in audiovisual content in 360º. The prototype version of the editor was created following the feedback from professional subtitlers that was gathered in a previous focus group (Agulló and Matamala, 2019). The tool has been developed in a collaboration between the different partners of the ImAc project [22] that includes broadcasters, researchers and technological companies. Specifically, the company Anglatècnic, experts in the engineering and development of software systems for the broadcast and IT sectors, was in charge of the technical development. End-users had a key role in the development, providing valuable feedback from the beginning of the project. It was decided to first develop a light version in web format as a prototype, and a desktop version would be created at a later stage of the project. The prototype editor, which is accessible via browser, supports 360º videos that can be uploaded and then previsualised for subtitling. Navigation in the 360º video with the mouse or the keyboard is possible. Most of the options available in the editor are similar to those in other commercial subtitle editors. In Figure 7, main options and sections can be seen: video controls (play, stop, pause, forward/backward options, etc.), subtitle controls (style, position of the subtitles, alignment options, colour coding for character identification, actions to navigate the subtitles, buttons for time code in and out, etc.), video preview, subtitle text box, subtitle list, among others.

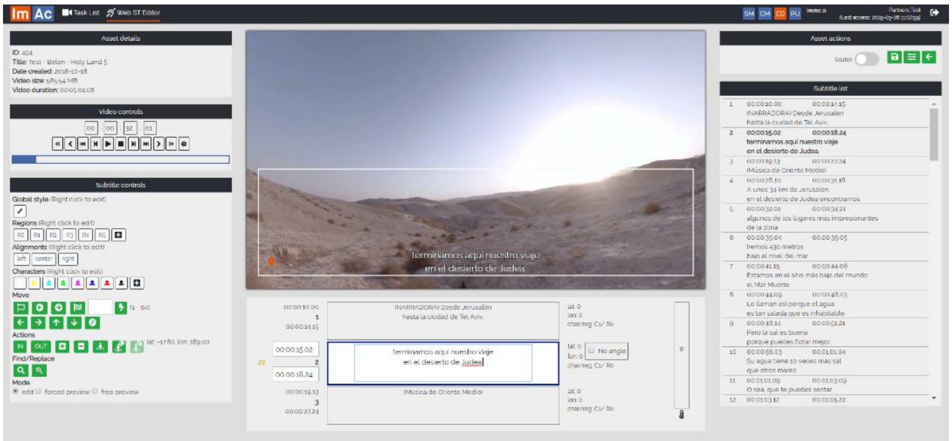


Figure 7. ImAc web subtitle editor interface.[23]

The prototype editor also includes a reading speed thermometer for character limitation. The thermometer is a visual guide to avoid exceeding the permitted words per minute. The default parameter is 120 words per minute. The thermometer becomes redder as the subtitles approach to the limit. Another option is the pre-set regions. That option allows the subtitler to set different regions for the subtitles (for example, up, down, left, right, etc.). The regions are highlighted with a white rectangle as can be seen in Figure 7, so that they can be spotted easily.

[22] http://www.imac-project.eu/consortium/
[23] The sample video in this image belongs to the video Holy Land (episode 5) created by RYOT.

Moreover, the prototype editor includes an option that had to be designed due to the nature of 360º content. This option is called “Set current angle”. As it has been explained in previous sections, the speakers are not always positioned within the viewers’ FoV, because they can move in the 360º space. Therefore, the position of the speaker for each subtitle needs to be indicated when creating the subtitles. To do so, the subtitler can navigate in the video and select the angle where the speaker is positioned. Then, when the option “Set current angle” is activated, the selected subtitle is anchored to the angle where the speaker is positioned. This information will then be translated into a metadata in the subtitle file that will be processed by the 360º video player and, when the speaker is out of the viewers’ FoV, a guiding mechanism will appear, for example, an arrow or a radar, to indicate the viewers where to find the speaker.

Finally, it is worth mentioning that the prototype editor includes three preview modes: (1) Edit mode: for editing the subtitles; (2) Forced preview: for previewing the subtitles (edit options are blocked) with the system forcing the selected angle for each subtitle; and (3) Free preview: for previewing the subtitles (edit options are blocked), being free to navigate the video as desired.

4.5. Evaluation

An experiment to test the prototype version of the ImAc web subtitle editor was conducted. The goal of this experiment was to test the usability of the editor and, most importantly, to gather participants’ feedback on the tool and on their needs when subtitling 360º content. Before carrying out the present test, a pilot was conducted with 3 users (one practitioner, one university lecturer and one engineer) to verify that the methodology worked properly. Pilot participants took the test online, timed themselves and provided feedback to the researcher. This section describes the evaluation methodology and presents the obtained results.

4.5.1. Evaluation setup

The test was carried out online and each participant took the test with their personal computers/laptops. The only technical requirements were to have a stable, high speed internet connection and to access the web editor with Google Chrome (recommended) or Firefox. The web editor includes a content management module that was used to set up the entire test. In the content management tool, users with the role of subtitler (for example, P1, P2, P3, etc.) were created and assigned to each participant.

Also, each participant was assigned an individual task for subtitling the same sample video. The video and the task were copied as per the number of participants, so they had access to an individual task only accessible by each of them. The videos were uploaded in low resolution to avoid overloading the server, causing a poor performance. The login information was provided by email to participants, together with the instructions. Questionnaires were administered using Google Forms. The test was designed in English, including the video to be subtitled, the instructions and the questionnaires.

4.5.2. Evaluation methodology

The test was sent to participants via email. It was divided in three parts and participants were required to perform it in just one session:

(1) Introduction, ethical clearance and demographic questionnaire

In the first part of the test, context information about the ImAc project, the goal of the test, the estimated duration (30 minutes) and the participant code was provided in the email to participants. Then, they were asked to give their consent to participate in the test via digital form, as requested by the ethical committee at Universitat Autònoma de Barcelona. Finally, they needed to fill in a demographic questionnaire in order to gather data about their profile, as it will be reported in section 5.4.

(2) Technical information and instructions

In this second part of the test, the participants were asked to first read a quick user guide that was created ad-hoc for the test. It was decided to provide a user guide instead of letting them figure out how the editor worked, because it was assumed that most participants were not familiar with 360º videos, nor with the potential issues of subtitling this type of content. In the user guide, an emphasis on the “Set the current angle” option was put, because this was the newest feature and the most difficult to understand, compared to current subtitling software. Also, giving participants more information about the tool would speed up the tasks and elicit more comprehensive and valuable replies in the open questions part of the test.

Then, instructions on how to access the web editor, as well as login information were provided. A transcription for the video was also provided to speed up the process. The tasks to be carried out were clearly written and shared with the participants as well, and were as follows:

The tasks that we kindly ask you to perform are:

1. Go to the subtitle editor using the login information. To access the subtitle web editor, you have to go to this address: XXXX and enter the login info that has been provided to you in the email.

2. Open the video that has been assigned to you for subtitling.

3. Subtitle the video into your native language from 00:00:00 to 00:01:11.

    a. Add subtitles with the correct timecodes.

    b. Assign the different colours to the different characters in the video.

    c. Set the angle for each subtitle.

    d. Set a second region for subtitles and apply it to one subtitle.

    e. Change the alignment to the left for one subtitle.

    f. Insert a subtitle between two existing subtitles.

    g. Delete two subtitles.

    h. Look for a subtitle by content.

4. Preview the video with the forced mode.

5. Save the subtitles and go back to the main window.

6. Open the video again.

7. Preview the video with free mode.

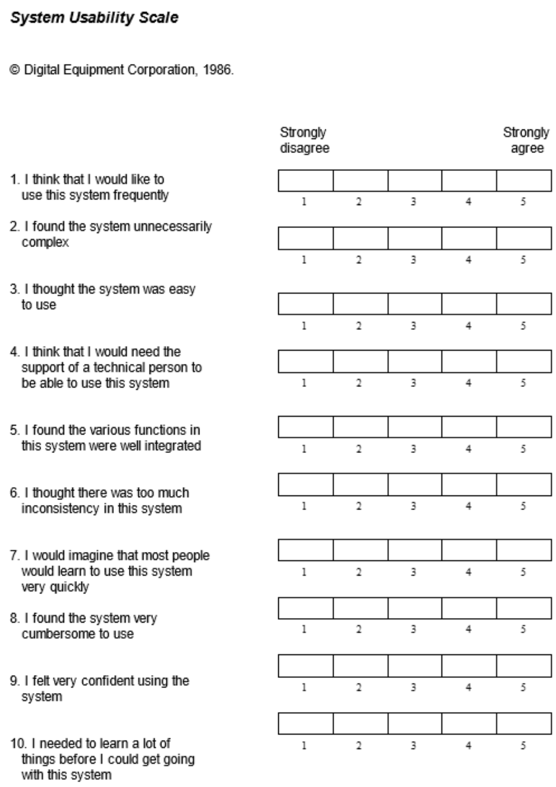
8. Save the subtitles and go back to the main window.

(3) Evaluation questionnaires

After performing all the tasks, participants were asked to reply to an online questionnaire that was divided into two parts: System Usability Scale (SUS) questionnaire and open questions.

Testing user experience is a widely researched field of study (Goodman *et al.*, 2012; Tullis and Albert, 2013; Pannafino and McNeil, 2017). There are several methods to test user experience in relation to effectiveness, efficiency and satisfaction. For example, cognitive walkthroughs, card sorting, icon usability testing, contextual inquiry or online surveys. In this test, the main focus is on usability as defined by Brooke (1996:

1): "Usability is not a quality that exists in any real or absolute sense. Perhaps it can be best summed up as being a general quality of the appropriateness to a purpose of any particular artefact." In this sense, the appropriateness of the prototype of ImAc web editor for subtitling 360º content was tested. For that purpose, one of the widest known and used scales, namely the System Usability Scale (SUS), was administered to participants. SUS is a ten-item Likert scale. Each item needs to be assessed from 0 to 5. This scale was chosen because it is easy to administer, provides reliable results with small sample sizes, and has been validated and used in many studies as a standard self-reported metrics (Brooke, 2013), becoming one of the most popular measurements for usability testing (Lewis, 2018).


System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Figure 8. System Usability Scale (Brooke, 1996).

To complement the quantitative data from SUS, open questions were also included in the post-questionnaire. The questions were aimed at gathering qualitative information on participants' feedback and impressions on the prototype and its specific functionalities. The questions were:

1. What did you like most about the subtitle editor?

2. What did you like less about the subtitle editor?

3. What do you think could be improved, and how?

4. Did you miss any functionality? If yes, can you tell us which?

5. Do you find the feature for setting the angle for the subtitle easy to use? Explain why.

6. Were the preview modes useful for you? Explain why.

7. Do you think it will take you longer to subtitle videos in 360º? Why?

8. Do you think 360º videos will impact your work as a subtitler?

9. Other comments.

4.5.3. Participants

Twenty-seven participants took part in the current study. The recruitment criterion was that participants had to be subtitlers who professionally subtitle audiovisual content. They were recruited via personal contacts and also by email and public posts on Twitter. The test was designed in English so that professionals from different countries could participate without language being a major limitation. The source language of the video was English. Participants could choose the language into which they preferred to create the subtitles. The quality of the subtitles was not observed; therefore, the participants did not need to master the target languages.

4.5.4. Evaluation results

The results from the different questionnaires are reported in this subsection.

Demographic questionnaire

Twenty-seven participants took part in the test (20 females and 7 males), with ages ranging 24-48 (mean=35.6, sd=6.9). Their main languages (that is, mother tongue) were Spanish (19), Catalan (3), English (3), Polish (2), Basque (1), Croatian (1) and Romanian (1) (two participants were bilingual Spanish/Catalan and one participant was bilingual Spanish/Basque). They describe their main jobs as AVT translators, subtitlers for different kind of products, university lecturers and researchers. Only one participant has subtitled a 360º video before. They presented a varying experience in the field of subtitling, varying from 1 month to 20 years (mean=8.1; sd=6.1). 16 participants have produced more than 300 hours of subtitled content, 3 participants have produced between

151 and 300 hours of subtitled content, 4 participants have produced between 51 and 150 hours and 4 participants have produced less than 50 hours. Participants usually subtitle in Spanish (21), English (10), Catalan (4), Polish (2), Basque (1), Croatian (1), French (1), Italian (1) or Romanian (1). Participants declared using different subtitling software (FAB, WinCAPS, Aegisub, VisualSubSync, Subtitle Workshop, EZTitles, Swift, Subtitle Edit, TED, Amara, YouTube, Spot, VICOM, Jayex, proprietary software from clients, among others). 26 participants have studies of university level and one participant has further education. 18 participants have a degree or master's degree on translation and interpreting studies (or languages degrees), 8 of them have PhD studies and 9 of them specialising in Audiovisual Translation. 24 participants have received specialised training on subtitling in MAs, specialised courses or training.

When asked about which devices they used on a daily basis, all participants agreed on using mobile phones; 23 participants use laptops; 21 participants use TVs, 17 participants use PCs; and 9 of them use tablets. When asked about how often they watch virtual reality content, none of the participants have watched VR content on a tablet, 23 participants have never watched VR content in a smartphone plugged to HMD or in HMD; some (14) occasionally watch VR content in a smartphone, 12 participants on a PC, 4 in a smartphone plugged to HMD and 3 in HMD; 1 participant watches VR content on a PC at least once a month, and 1 participant in an HMD; finally, 1 participant watches VR content in smartphone at least once a week. When asked to explain why they have never used virtual reality content such as 360º videos or only occasionally, 3 participants replied that they are not interested, 4 participants replied that it is not accessible, 16 participants replied that they have not had the chance to use it, and others gave other reasons regarding the expensive price, difficulties to use the technology or the lack of appealing content. When asked to state their level of agreement with the statement "I am interested in virtual reality content (such as 360º videos)", 3 participants replied that they strongly agree, 13 replied that they agree, 7 that they neither agree nor disagree and 4 of them disagree. Finally, when asked if they own any device to access virtual reality content, 15 participants replied that they do not own any, 5 replied that they do not know or prefer not to reply and 7 replied that they do (including BOBVR Z4, HTC Vive, PC, laptop, smartphone and PlayStation VR).

SUS

SUS scoring system is standardised and defined by its creators. It is considered that a score above 68 is above average and below 68 would be under average. The raw SUS score for this test was 59.5. The letter grade is D+, and the score corresponds to the percentile rank: 29-30% (Sauro and Lewis 2016: 203-204). The result could also be due to the lack of experience subtitling 360º content. However, this was a first prototype, so the most important part of the present study was the qualitative feedback that will be detailed below and that provides valuable input on how to improve this first version of the 360º video subtitle editor.

SUS statements	**1 (strongly disagree)**	**2**	**3**	**4**	**5 (strongly agree)**
1. I think that I would like to use this system frequently	3 (11.1%)	4 (14.8%)	11 (40.8%)	8 (29.6%)	1 (3.7%)
2. I found the system unnecessarily complex	5 (18.5%)	10 (37.1%)	7 (25.9%)	4 (14.8%)	1 (3.7%)
3. I thought the system was easy to use	1 (3.7%)	6 (22.2%)	4 (14.8%)	14 (51.9%)	2 (7.4%)
4. I think that I would need the support of a technical person to be able to use this system	8 (29.6%)	12 (44.5%)	5 (18.5%)	0 (0%)	2 (7.4%)
5. I found the various functions in this system were well integrated	0 (0%)	6 (22.2%)	9 (33.3%)	10 (37.1%)	2 (7.4%)
6. I thought there was too much inconsistency in this system	4 (14.8%)	10 (37.1%)	11 (40.7%)	2 (7.4%)	0 (0%)

7. I would imagine that most people would learn to use this system very quickly	2 (7.4%)	2 (7.4%)	7 (30%)	8 (29.6%)	8 (29.6%)
8. I found the system very cumbersome to use	2 (7.4%)	5 (18.5%)	8 (29.6%)	10 (37.1%)	2 (7.4%)
9. I felt very confident using the system	0 (0%)	8 (29.6%)	8 (29.6%)	11 (40.8%)	0 (0%)
10. I needed to learn a lot of things before I could get going with this system	9 (33.3%)	3 (11.1%)	10 (37.1%)	4 (14.8%)	1 (3.7%)

Table 12. SUS replies from participants.

Open questions

The analysis of replies to open questions follows a qualitative approach. Participants were asked to reply with their own words to the questions specified in section 5.2. All replies were thoroughly revised and tagged. Some of the questions were generic aimed at gathering general feedback on the ImAc web editor and others were more specific about functionalities. Finally, some questions were aimed at gathering feedback on how subtitlers perceived that the subtitling task could be impacted by immersive environments. The analysis allowed to define different areas that work well and others that need to be improved in the current ImAc editor, how the new functionalities were received and subtitlers' feedback on the future practice of subtitling immersive content. The most relevant aspects are described below.

(1) General feedback on web editor

As stated before, the open questions for the general feedback on the prototype were focused on what participants liked about the tool, what they did not like and what they would improve. After analysing the results, it can be stated that the prototype of the 360º subtitle editor was well received due to several reasons. Participants stated that what they liked the most was that the system and some of its features were easy to use (the word "easy" was used in 9 of the replies). Also, adjectives such as intuitive (2), userfriendly (1) and straightforward (1) were used to describe the tool, as well as characteristics such as simplicity (1), practicality (1) and versatility (1). The interface was referred by some respondents (4) as a positive part of the web editor, being considered clear and well designed. Some of the functions that were more praised and are related to the specific task of subtitling 360º content were: navigation and subtitling of 360º videos (5), set the angle option (4), assigning pre-set regions (4), assigning colours to different speakers (3) and reading speed thermometer (2). Two respondents stated that they liked the tool was cloud-based. Therefore, the main innovations introduced in this prototype such as previsualisation and navigation of 360º content and setting the angle for the speakers can be considered usable.

As for improvable features, two main blocks were detected: (a) general subtitling features and (b) specific 360º content subtitling features. The first block is less relevant for the current study, because the tested version was a prototype and general features were planned to be improved in future versions. The second block, however, is important to redefine the needs of subtitlers regarding 360º videos that might have not been considered when developing the tool. Regarding the first block, several improvements were suggested: customisable shortcuts; improved and clearer time coding options; freedom to break lines as desired (at the time the test was carried out, an automatic segmentation based on a default character limitation was implemented, not allowing customisation); transparency in the reading speed thermometer and clearer information about characters per second (cps) or words per minute (wpm); sound wave to ease spotting; more editing options for subtitles (including bold, italics, different colours); more quality assurance options such as spellcheck; among other minor suggestions.

In the second block, some interesting comments for improvement were spotted for the 360º content subtitling features. Firstly, two participants suggested to integrate editing and preview modes. Instead of changing from edit to preview mode each time the subtitler needs to previsualise their work, participants would prefer to have it integrated in a unique mode as in other 2D subtitling editors. This suggestion could ease the spotting process and quality assurance. Secondly, regarding the set the angle option, one participant suggested that it would be easier to right click on the video with the mouse to indicate where the speaker is or select with the mouse the area where the speaker is, instead of navigating with the mouse or arrows. This suggestion could improve accuracy for this feature. Finally, two respondents suggested to include an automatic shot detection system. However, this suggestion is not compatible with 360º videos. As it has been stated before, the viewer is in control of the camera for this type of content. Therefore, the changes of shots depend on viewers and are unpredictable. Automatic shot change options are not relevant in subtitling for immersive content.

(2) Specific functionalities

The second part of the open questionnaire was addressed to specific functionalities for 360º content: the “set the angle” option and the preview modes. Regarding the “set the angle” option, most participants (20) considered it was easy to use, describing the option as easy or very easy, relatively simple, straightforward, not complex, intuitive and logical. Five users, though, found it difficult. No correlation was found between previous experience or knowledge of immersive environments for this negative reply. Participants also made some recommendations to improve this option. Three of them suggested to have an option to apply the same angle to consecutive subtitles. This would definitely ease the task of assigning an angle for each character and scene. Also, one subtitler raised a concern about off-screen speakers (for example, a narrator), in which scenario no angle should be chosen. This was not considered when developing the tool, as it was assumed that the speaker was always on screen, but it is not always the case. Therefore, an option for indicating off-screen voices without angle needs to be implemented.

As far as the preview modes, most participants (23) considered these modes were useful. Three participants stated that forced mode was useful to check the subtitling work and the free mode was useful to experience the subtitles as a final user would do. Four participants considered that it was necessary to integrate edit and preview modes, so that they can edit subtitles while previewing the video.

(3) Impact of subtitling 360º content

The last part of the open questionnaire was aimed at gathering general feedback about the professional subtitlers’ impressions on this new medium and subtitling practice. When asked whether subtitling 360º video will take longer, most participants (22) responded affirmatively, and most of them (16) considered that having to set the angle would be more time-consuming. Two participants also stated that subtitling 360º would take longer, because the subtitler needs to check the whole 360º video in case there are some texts on screen or inserts that need to be translated in the subtitle. Three participants considered that it should not take longer as long as subtitlers have access to specific software for it. Finally, one participant was worried about shot changes and how they would work in 360º. As it has been stated before, this is not relevant in 360º videos.

When participants were asked about the impact of subtitling 360º videos on the job of a subtitler, different opinions were expressed. Three participants believed that 360º video will have no impact on the subtitler profession and five were not sure about it. Six participants thought that this will have an immediate impact on the profession and most of them (9) considered that 360º videos will impact subtitling, but in the future, if this medium is mainstreamed and the demand increases.

4.6. Conclusions

An updated review of current subtitling technology has been presented in this article in order to contextualise the study. Different innovative solutions are being developed to cope with current subtitling demands as the audiovisual translation industry grows, due to the digitalisation of communication and proliferation of VoD platforms. Efforts are put in automating and streamlining the workflow of subtitling, with technologies such as ASR, CAT tools or MT. However, new challenges for subtitling are being posed not only by the increasing demand but also by the appearance of new media, such as immersive environments. A review of main immersive environments (3D, AR and VR) and their implications for subtitling has also been introduced in this article. The focus of the study is on VR and, therefore, the main challenges of subtitling 360º content have been presented. Professional subtitlers now need to consider new aspects such as where to locate the subtitles for an optimal reading or how to indicate when a speaker is outside of the FoV of the viewer. To that end, a prototype of the ImAc web editor has been developed and presented to twenty-seven professional subtitlers who have tested the tool and reported the correspondent feedback on usability and preferences.

The test has shed some light on the possibilities of subtitling 360º content as well as the most important characteristics that a subtitling software should include to be usable for that task. New features designed considering the nature of 360º videos such as setting the angle, pre-set regions or preview modes have been well received by most participants. According to their feedback, these features are usable and suitable to generate subtitles for this kind of content, although some improvements have been suggested and will be implemented in future versions of the editor. For example, the integration of the edit and preview mode or easing the task of setting the angle or including an option for off-screen characters. Also, it is important to remark that traditional subtitling considerations such as shot changes do not apply in this new medium. Therefore, automatic shot changes options will not need to be implemented in this type of subtitle editors. As reported in the results section, most participants considered that this new medium will have an impact on the subtitling practice, mainly because of the new dimension brought by this medium: directions. Having to set the angle for the different speakers and having a 360º sphere to look around seems to be concerning for subtitlers. Therefore, the provided technology for this task needs to focus on simplifying and automating these additional tasks to make the subtitling process easier.

Some limitations could be identified in this test. The estimate duration of the test was not accurate. When the pilot for the test was carried out, pilot participants informed that the test would take around 30 minutes, but then this time was considered insufficient by some of the participants. This does not mean that participants had to stop the test, but some of them were frustrated that the estimation was not accurate enough. For future tests of this type, the duration should be longer, or participants should perform a previous session to get use to the tool before carrying out the test. Another limitation was the profile of subtitlers. Some subtitlers (especially for interlingual subtitles) are used to translate templates, not generate subtitles from scratch. Therefore, some of them might not be used to the process of setting timings, which was considered time-consuming and difficult by some participants. However, the participants of this test were familiar with subtitling software, because they replied to the question “What software do you normally use?” as reported in the demographic results section. For future tests, if possible, all participants should be accustomed to generating subtitles from scratch or at least a question in the demographic questionnaire should be included in this regard in order to find whether this has a correlation with the results.

This study has proven the importance of carrying out usability tests with end users when developing specific software. Subtitlers’ feedback has been essential to define the future of the ImAc editor and to develop a tool that would respond to professional needs and market demands. Future versions should be again tested and polished until professional subtitlers are satisfied with the results. Subtitling is a powerful tool for accessing information and is becoming an intrinsic part of the audiovisual communication. Therefore, the necessity to have appropriate software and technology to respond to the increasing demands is clear. A subtitling revolution will never be achieved without a revolutionary technology that complements it and subtitlers are the cornerstone of this revolution.

Bibliography

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*, 32.

Agulló, Belén and Pilar Orero (2017), "3D Movie Subtitling: Searching for the best viewing experience", *CoMe - Studi di Comunicazione e Mediazione linguistica e culturale*, 2, pp. 91-101.

Agulló, Belén, Anna Matamala and Pilar Orero (2018), "From disabilities to capabilities: testing subtitles in immersive environments with end users", *HIKMA*, 17, pp. 195–220.

Álvarez, Aitor, Arantza del Pozo and Andoni Arruti (2010), "APyCA: Towards the automatic subtitling of television content in Spanish", *Proceedings of the International Multiconference on Computer Science and Information Technology*, IEEE, pp. 567-574.

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. Madrid: Asociación Española de Normalización y Certificación.

Brooke, John (1996), "SUS-A quick and dirty usability scale", in Patrick W. Jordan *et al.* (eds.), *Usability evaluation in industry*, London, Taylor and Francis, pp. 189-194.

Brooke, John (2013), "SUS: A Retrospective", *Journal of Usability Studies*, 8(2), pp. 29-40.

Brown, Andy, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong and Maxine Glancy (2018), "Exploring Subtitle Behaviour for 360° Video." White Paper WHP 330, BBC Research & Development, on https://www.bbc.co.uk/rd/publications/whitepaper330 (consulted on 26/03/2019).

Bywood, Lindsay, Martin Volk, Mark Fishel and Panayota Georgakopoulou (2013), “Parallel subtitle corpora and their applications in machine translation and translatology”, *Perspectives*, 21(4), pp. 595-610.

Díaz-Cintas, Jorge (2013), “The technology turn in subtitling”, in Marcel Thelen and Barbara LewandowskaTomaszczyk (eds.), *Translation and meaning: Part 9*, Maastricht, Zuyd University of Applied Sciences, pp. 119-132.

Díaz-Cintas, Jorge (2014), “Technological Strides in Subtitling”, in Chan Sin-wai (ed.), *The Routledge Encyclopedia of Translation Technology Routledge*, Abingdon, Routledge, pp. 632-643.

Dranch, Konstantin (2018), “Translation tech goes media: Subtitling and dubbing won’t stay uncontested for long”. Nimdzi Blog, on https://www.nimdzi.com/translation-tech-goes-media/ (consulted on 18/01/2019).

Estopace, Eden (2017), “Audiovisual Translation Hits a Sweet Spot as Subscription Video On-Demand Skyrockets”, Slator, in https://slator.com/features/audiovisual-translation-hits-sweet-spot-subscription-video-on-demand-skyrockets/ (consulted on 15/01/2019).

European Broadcasting Union (EBU) (2017), “Virtual Reality: How are public broadcasters using it?”, on https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it (consulted on 11/01/2019).

Forcada, Mikel L. (2017), “Making sense of neural machine translation”, *Translation Spaces*, 6(2), pp. 291–309.

Garcia, Jose Enrique, Alfonso Ortega, Eduardo Lleida, Tomas Lozano, Emiliano Bernues and Daniel Sanchez (2009), “Audio and text synchronization for TV news subtitling based on Automatic Speech Recognition”, *2009 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, IEEE, pp. 1-6.

Georgakopoulou, Panayota (2012), “Challenges for the audiovisual industry in the digital age: the everchanging needs of subtitle production”, *The Journal of Specialised Translation*, 17, pp. 78-103.

Goodman, Elizabeth, Mike Kuniavsky and Andrea Moed (2012), *Observing the User Experience: A Practitioner's Guide to User Research*. Waltham, Morgan Kaufmann.

Harris, Aisha, Margaret Lyons and Maureen Ryan (2019), "'Bandersnatch' Has Many Paths, but Do Any of Them Add Up to Anything?" The New York Times, on https://www.nytimes.com/2019/01/04/arts/television/bandersnatch-black-mirror-netflix.html (consulted on 11/01/2019).

Lewis, James R. (2018), "The System Usability Scale: Past, Present, and Future", *International Journal of Human–Computer Interaction*, 34(7), pp. 577-590

Lindsay Bywood, Panayota Georgakopoulou and Thierry Etchegoyhen (2017), "Embracing the threat: machine translation as a solution for subtitling", *Perspectives*, 25(3), pp. 492-508.

Matamala, Anna, Andreu Oliver Moreno, Aitor Álvarez Muniain and Andoni Azpeitia Zaldua (2015), "The reception of intralingual and interlingual subtitling: an exploratory study within the HBB4ALL project", *Proceedings of the 37th Conference Translating and the Computer*, pp. 12-17. https://ddd.uab.cat/record/144868

Mathur, Abhinav, Tanya Saxena and Rajalakshmi Krishnamurthi (2015), "Generating Subtitles Automatically Using Audio Extraction and Speech Recognition", *2015 IEEE International Conference on Computational Intelligence & Communication Technology*, Ghaziabad, IEEE, pp. 621-626.

Morales, Manuel, Tommaso Koch and Luis Pablo Beauregard (2019), "Netflix removes Castilian Spanish subtitles from Alfonso Cuarón's 'Roma'", El País, on https://elpais.com/elpais/2019/01/10/inenglish/1547125380_758985.html (consulted on 11/01/2019).

O'Hagan, Minako and Carme Mangiron (2013), *Game localization: translating for the global digital entertainment industry*. Philadelphia, John Benjamins Publishing Company.

Pannafino, James and Patrick McNeil (2017), *UX Methods: A Quick Guide to User Experience Research Methods*. CDUXP LLC.

Perkins Coie LLP (2018), "2018 Augmented and Virtual Reality survey report: Insights into the future of AR/VR", on https://www.perkinscoie.com/images/content/1/8/v2/187785/2018-VR-AR-Survey-Digital.pdf (consulted on 11/01/2019).

Roach, Joan (2018). "AI technology helps students who are deaf learn", The AI Blog, on https://blogs.microsoft.com/ai/ai-powered-captioning/ (consulted on 29/01/2019).

Rothe, Sylvia, Kim Tran and Heinrich Hussmann. (2018), "Dynamic Subtitles in Cinematic Virtual Reality", *ACM TVX 2018*, ACM, DOI: https://doi.org/10.1145/3210825.3213556.

Sauro, Jeff and James R. Lewis (2016), *Quantifying the user experience: Practical statistics for user research*. Amsterdam, Morgan Kaufmann.

Tullis, Thomas and William Albert (2013), *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Waltham, Morgan Kaufmann.

Chapter 5: Article 4. Subtitles in virtual reality: a reception study

Subtitles in virtual reality: a reception study

Belén Agulló, Universitat Autònoma de Barcelona

Anna Matamala, Universitat Autònoma de Barcelona

Immersive content has become a popular medium for storytelling, with 360º videos offering a different perspective in the form of an immersive audiovisual experience. This type of content is typically accessed via a head-mounted visual display, within which the viewer is located at the center of the action, with the freedom to look around and explore the scene as desired. Unfortunately, 360º videos cannot yet be classified as accessible because subtitles are not always available and solutions are required. In this article the authors review this topic, the previous studies, and the main challenges faced when designing subtitles for 360º content: position and guiding methods. The criteria for subtitle position has still not been defined for immersive content as it has been for content viewed in a traditional context and solutions are required. Guiding mechanisms are necessary for circumstances in which the speakers are not visible and viewers lacking an audio cue need visual information to guide them through the virtual scene. For each area under investigation in this study, two methods have been compared. For position, always-visible position (when subtitles are consistently in front of the viewer) has been compared with fixed-position (when subtitles are fixed and evenly spaced every 120º in the 360º scene). For guiding methods, arrows (that are integrated into the subtitles and indicate where the viewers need to direct their gaze) are compared to a radar (located in a fixed position in the field of view, indicating where the speaker is located in relation to the position of the viewer). Feedback on preferences, immersion (using the IPQ questionnaire) and head movements was gathered from 40 participants (20 hearing and 20 hard of hearing). Results show that always-visible subtitles with arrows are the preferred option. Always-visible and arrows achieved higher scores in the IPQ questionnaire than fixed-position and radar. Head movement patterns show that participants move more freely when the subtitles are always-visible than when they are in a fixed position.

5.1. Introduction

Immersive media such as virtual reality (VR) is attracting the interest of academia and the industry. Content creators are developing 360º videos that provide an immersive and engaging audiovisual experience for their viewers. 360º videos are real-world recordings, captured using special camera sets and post-edited to create a 360º sphere (MacQuarrie & Steed, 2017). The most suitable devices for accessing this type of content are head-mounted displays (HMD). With HMD, the viewer finds itself in a 360º sphere, at the center of the scene, in the heart of the action. They have the freedom to look around and explore a virtual world, giving them the sensation that they are being physically transported to a different location. The potential for this technology in industries such as filmmaking, video games, and journalism has become evident in recent years.

The European Broadcasting Union (EBU) issued a report on the use of VR by public broadcasters in 2017. In this report, 49% of respondents stated that they offered 360º video content. Broadcasters see the potential of this medium because the audience can gain a better understanding of the content if they feel a sense of immersion, in particular with items such as the news or the narration of historical events (EBU, 2017). In that sense, *The New York Times* (*NYT*) and the British Broadcasting Corporation (BBC) have promoted the creation of 360º videos and launched specific platforms via which to display this content. *NYT* has its own app: the *NYT* VR[24], accessible in any smartphone with Google Cardboard or any other VR chassis. The BBC 360º content can be accessed via their website[25] or on their YouTube channel. Other broadcasters such as CNN or ABC have also produced immersive videos for sharing their news and events.

Fictional content is limited in the 360º filmmaking industry. The reasons for this could be due to a lack of resources and knowledge when it comes to filmmaking strategies in immersive media, along with certain factors holding back the mainstream adoption of this technology. Some of these factors include a lack of quality content, a lack of knowledge of VR technology among the general audience, the cost of hardware and the complexity of the user experience, which is still too arduous for the average user (Catapult Digital, 2018). However, the entertainment industry is taking advantage of this technology, with an anticipated growth in video games in the VR sector (Perkins Coie LLP, 2018). The VR technology used in video games is different from the technology

[24] http://www.nytimes.com/marketing/nytvr/
[25] https://www.bbc.com/reel/playlist/reel-world-360-videos-from-the-bbc

used in 360º content creation. In terms of creating fictional content, VR in video games is based on computer-generated images as apposed to real-life images and, therefore, offers more freedom to its content creators, without the need for expensive and complex recording equipment. In turn, video game users tend to be more familiar with adapting to new technologies, hence the success of this technology within the gaming industry.

Overall, it can be stated that immersive content is increasingly present in our society. However, it is still not accessible to all users. To address this issue, the Immersive Accessibility[26] (ImAc) project funded by the European Union's Horizon 2020 research and innovation program (under grant agreement number 761974) was formed. The aim of this project is to explore how access services such as subtitles for the deaf and hard of hearing (SDH), audio description and sign language interpreting can be implemented in immersive media. The present study is focused on the implementation of subtitles in 360º videos. Subtitles have become an intrinsic part of audiovisual content; what Díaz-Cintas (2013, 2014) has labeled as the commoditization of subtitling. Nowadays, subtitles can be easily accessed on digital televisions, video-on-demand platforms, video games and so on, with just one click. The number of subtitle consumers is also increasing and their needs and reasons for the use of the service are varied: they are learning a new language, watching television at night and do not want to disturb their children, commuting on a train and do not have headphones, they have hearing loss and are using subtitles in order to better understand the content, or as non-native speakers, they simply prefer to watch content in its original language but with subtitles to assist.

In recent years, extensive studies on subtitling have been conducted from various perspectives (D'Ydewalle, Pollet & van Rensbergen, 1987; Díaz-Cintas, Orero & Remael, 2007; Romero-Fresco, 2009; Bartoll & Martínez-Tejerina, 2010; Matamala & Orero, 2010; Perego, Del Missier, Porta & Mosconi, 2010; Szarkowska, Krejtz, Kłyszejko & Wieczorek, 2011; Arnáiz-Uzquiza, 2012; Mangiron, 2013; Romero-Fresco, 2015; Szarkowska, Krejtz, Pilipczuk, Dutka & Kruger, 2016). However, research into subtitles in immersive content is scarce and reception studies are needed in order to find solutions for its lack of accessibility. The aim of this study is to gather feedback regarding different subtitle solutions for the main issues encountered when implementing subtitles in immersive content i.e. position and guiding mechanisms. Feedback has been gathered from 40 participants regarding their preferences for the various options available and how

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

they impact on a sense of immersion, with additional data based on participants' head movements. In this article, a review of previous studies related to subtitling in immersive media is presented. The methodological aspects of the study are then introduced, such as the participants, the experimental design and the materials used. Results are presented in section 4, and in section 5 the study is discussed. Finally, conclusions are drawn in section 6.

5.2. Related work

Research on subtitles in immersive media is relatively recent. Researchers in this field have highlighted several of the challenges being faced when designing and implementing subtitles in immersive media (Rothe, Tran & Hussmann, 2018a; Brown et al., 2018). Firstly, the position of the subtitles needs to be defined. Subtitles on a traditional screen are already standardized and usually located at the bottom-center of the screen which is static. However, the field of view (FoV) in 360º content is dynamic and viewers can decide where to look at any time during the scene. Therefore, the position of the subtitles needs to be carefully defined so as to avoid loss of content during the experience. The subtitles need to be located in a comfortable field of view (CFoV), that is, a safe area that is guaranteed to be visible for its users. If subtitles overlay the CFoV, then they will be cropped and therefore unintelligible. In addition, a guiding mechanism needs to be included to enhance accessibility. If the speaker is outside the FoV, persons with hearing loss (for whom the audio cues are not always helpful) will need a guiding system which will indicate where to look. As 360º content aims to provide an immersive experience, subtitles must therefore be created in a way that does not disrupt this immersion. Finally, some viewers suffer from VR sickness or dizziness when watching VR content. The design of the subtitles should not worsen this negative effect and the subtitles should be easy to read.

In the ImAc project, some preliminary studies have approached this topic to gather feedback from users before developing a solution (Agulló, Matamala & Orero, 2018; Agulló & Matamala, 2019). In a focus group carried out in Spain, participants with hearing loss were asked how they would like to receive subtitles in 360º videos. They agreed that they would like them to be as similar as possible to those shown on traditional screens. They also stated that they would like the subtitles to be bottom-center of their vantage point and always in front of them. The participants also highlighted the importance of using the current Spanish standard for SDH (AENOR, 2003). Regarding directions, participants suggested the inclusion of arrows, text in brackets (to the left and to the right), and a compass or radar, to indicate where the speakers are in the scene (Agulló & Matamala, 2019). In a different preliminary study, feedback from a limited number of users was gathered regarding the CFoV and two guiding mechanisms (icons representing arrows and a compass). Results showed that the users preferred the arrows as a guiding mechanism and the largest font in the CFoV because it was easier to read (Agulló et al., 2018).

Some reception studies have already been conducted regarding subtitles in immersive media (Rothe et al., 2018a; Brown et al., 2018). The BBC Research & Development team proposed four behaviors for subtitles in 360º videos based on previous literature and design considerations. The four behaviors are 1) 120-Degree: subtitles are evenly spaced in the sphere, positioned 120º apart; 2) Static-Follow: subtitles are positioned in front of the viewer, responding immediately to their head movements; 3) Lag-Follow: the subtitles appear in front of the viewer and are fixed until the viewer rotates 30º, whereby it moves accordingly to its new position; 4) Appear: the subtitles appear in front of the viewer and are fixed in that position, even if the viewer moves around the scene (Brown et al., 2017). They tested the four different options with 24 hearing participants, using six clips which lasted 1 to 2 minutes. The Static-Follow behavior was preferred by participants because the subtitles were considered easy to locate and gave participants the freedom to move around the scene. Some issues were highlighted such as obstruction (a black background box was used) and VR sickness (Brown et al., 2018). Rothe et al. (2018a) conducted another study with 34 hearing participants comparing two positions: 1) static subtitles: always visible in front of the viewer, following their head movements; 2) dynamic subtitles: placed near the speaker in a fixed position. Participants did not state a clear preference in this study, however, dynamic subtitles performed better regarding workload, VR sickness, and immersion.

Current solutions carried out by broadcasters such as *NYT* or BBC are mainly burnt-in subtitles, spaced evenly in the 360º sphere[27], every 120º. In other media such as video games, subtitles that are always visible in front of the viewer are the most frequently used (Sidenmark, Kiefer & Gellersen, 2019). Different solutions are being tested and implemented by content creators, but a consensus on which option works best has not yet

[27] An example can be watched in the short documentary *The Displaced*, by *NYT*: https://www.nytimes.com/video/magazine/100000005005806/the-displaced.html

been reached. This study aims to further clarify which subtitles are more suitable for immersive environments and for all kinds of users, including hearing participants and participants with hearing loss. With this aim, the current solutions implemented by the main broadcasters (fixed subtitles located at 120º) are compared to the solutions developed in the ImAc project (always-visible subtitles and guiding mechanisms)[28]. This study contains a higher number of participants (40) than in previous research and includes a higher number of participants with hearing loss (20) in the sample. The subtitles were developed following SDH features (AENOR, 2003) and unprecedented research in the area of subtitle studies was made in the testing of guiding mechanisms in 360º videos. An additional contribution of this study is that longer content is used to better measure the preferences and immersion of its participants. In the following sections, the study and the results are presented.

5.3. Methodology

In this section, the methodology of the test is described, including aim and conditions, design and procedure, materials, and participants. Two pilot tests were carried out before the study to verify that the methodology was suitable for the test: a first pilot with eight participants (two participants with hearing loss and six hearing participants) (Agulló, Montagud & Fraile, 2019) and a second pilot with three participants (one with hearing loss and two hearing) to confirm that the changes from the first pilot were well implemented and that they were suitable for the experiment. The pilots were particularly necessary because of the complexity of the technical setup and the newness of the medium.

5.3.1. Aim and conditions

The main goal of the experiment was to test position and guiding methods for subtitles in 360º content in terms of participants' preferences, immersion, and head movement patterns, using the CVR Analyzer tool (Rothe, Höllerer & Hussmann, 2018b).

The two conditions for position were as follows: a) fixed-position subtitles - subtitles attached to three different fixed positions in the sphere, spaced evenly, 120º apart (see Figure 9); b) always-visible subtitles - always displayed in front of the viewer,

[28] Please note that fixed-position subtitles in ImAc terminology are referred to as 120-Degree by Brown et al. (2017). In Rothe et al. (2018a), dynamic subtitles are fixed in one position close to the speaker. Therefore, the implementation is different. Always-visible subtitles in ImAc terminology are equivalent to Static-follow by Brown et al. (2017) and to static subtitles by Rothe et al. (2018a).

following their movements, attached to their camera or CFoV (see Figure 10). The reasons for testing these two conditions were two-fold. Firstly, the results from previous studies (Brown at al., 2018; Rothe et al., 2018a) were inconclusive, and included a limited number of participants with hearing loss, as they were not the main target group for the study. Secondly, a comparison between the industry developments by broadcasters like *NYT* and the BBC and the solution developed in the ImAc project was desired in order to corroborate the relevance of the project.

Figure 9. Fixed-position subtitles attached to one position in the sphere, in Episode 4 of *Holy Land* created by Ryot.

Figure 10. Always-visible subtitles attached to the FoV, in Episode 5 of *Holy Land* created by Ryot.

The conditions for the guiding methods were as follows: a) arrows - intermittent arrows appear on the left or right side of the subtitle text depending on the location of the speaker. If the speaker is to the left of the viewer, the arrows appear on the left of the subtitle text; if the speaker is to the right (see Figure 11), the arrows appear on the right of the subtitle text. The arrows only appear when the speaker is outside the FoV. b) Radar - a radar appears in a fixed position in the FoV (towards the bottom, on the right-hand side, near the subtitle) with information about the speaker's location (see Figure 12). With the radar, the viewer is located at the center, with their gaze directed by three inverted triangles as seen in Figure 12. The speaker is represented by a small colored triangle. The color is assigned depending on the color of the subtitle for character identification purposes. The viewer needs to move their gaze in order to position the small triangle (speaker) in the center of their FoV. In this part of the test, always-visible subtitles were used. The reason for testing these two options was because they were suggested by participants in a focus group at the beginning of the project (Agulló & Matamala, 2019). Moreover, the radar option is used in other 360º content players such as *NYT* VR[29], and the intention was to compare the current solutions with the solutions developed in the ImAc project (arrows).

Figure 11. Subtitle with arrows in *I, Philip* created by ARTE.

[29] https://www.nytimes.com/video/360-video

Figure 12. Subtitles with radar in *I, Philip* created by ARTE.

5.3.2. Design and procedure

The experiment was carried out in one session divided into two parts. In the first part, the position of the subtitles was tested. In the second part, the guiding methods were tested. A within-subject design was used to test the different conditions. Each participant was asked to watch an acclimation clip along with four other clips (see section 3.3.2). The clips were randomly presented with different variables (fixed-position, always-visible, arrow and radar). Both the clips (except for the guiding methods part) and variables were randomized among the participants to avoid a learning effect. The video used for the guiding methods part was a short science fiction movie called *I, Philip*. The duration of this movie was around 12 minutes and was cut into two parts in order to test the two variables (arrows and radar). However, the order of the clips was not altered, as they were contained within the narrative of the movie and the participants would not have understood the story. For this reason, only the variables were randomized. Twenty participants watched each clip/variable combination.

Before conducting the test, ethical clearance was obtained. The methodology went through the ethical committee at Universitat Autònoma de Barcelona which approved the procedure and consent forms that would be signed by the participants.

During the test, participants were first welcomed by the facilitator, followed by an explanation about the ImAc project and the context for the test. The approved consent form was signed by the participants and they were then asked to complete the demographic questionnaire. Next, the acclimation clip was shown to familiarize the participants with the HMD. After that, the four clips were presented and they were asked to complete the IPQ questionnaire (see section 3.3.3) after each clip. Finally, after each part, the participants were asked to reply to a questionnaire on preferences.

The facilitator was present for the entire duration of the test; providing information as required, helping the participants to correctly place the HMD on their heads and assisting with the completion of the questionnaires (see Figure 12). The questionnaires were administered using Google Forms on a laptop that was provided by the facilitator.

Figure 13. The facilitator with a participant during the test.

5.3.3. Materials

In this section, the different materials used in the experiment such as technical equipment, stimuli, and measuring tools, will be explained.

5.3.3.1. Technical setup

An Apache webserver was installed on a PC and was set-up for the evaluation in order to host the player resources and media assets (360º videos and subtitles). A Samsung Gear VR with Samsung Galaxy S7 was used. The videos were accessible via a URL that was directed to the server resources.

5.3.3.2. Stimuli

Five clips were shown in total. All the clips were presented without sound. The tests were conducted without sound in order to control any external variables, such as the extent of the participants' hearing loss or their level of English (the video voiceover was in English). The participants with hearing loss presented varying levels of hearing loss and different types of hearing aid, creating a condition wherein its impact on immersion could not be controlled. The level of English among the participants with hearing loss and hearing participants was unpredictable and the results could have therefore been affected. The main focus of the study was to determine which subtitles were most suitable for this type of content and for any type of user. Due to current industry requirements and the unprecedented aspect of the research taking place, emphasis was placed on position and guiding mechanisms. The ecological validity was therefore sacrificed in favor of controlling the external variables in order to receive more consistent feedback regarding the suitability of the different subtitles. In other studies involving testing of different subtitle implementations, the audio was also muted (Kurzhals, Cetinkaya, Hu, Wang & Weiskopf, 2017) or manipulated (Rothe et al., 2018a).

Subtitles were produced in Spanish, a language spoken and understood by all participants. The font type selected for the tests was the open-source, sans-serif Roboto developed by Google because studies on typefaces indicated that sans-serif fonts appear more natural and are more suitable for VR environments[30]. This font was used in the study carried out by the BBC as well (Brown et al., 2018). Regarding font size in VR environments, rather than being calculated by pixels as it is for traditional screens it is based on the CFoV (see Figure 14). The largest font was used, that is, the font that took 100% of the CFoV. The use of the largest font was selected because participants in previous pilots stated that they were more comfortable reading larger fonts, ones that encompass the largest proportion of the CFoV (Agulló et al., 2018). A background box was not included because in previous studies it was considered obstructive (Brown et al., 2018). Furthermore, during the pilots, participants did not report the lack of background box as a negative aspect of the subtitles, nor a hindrance when reading. Subtitles had a maximum of 37 characters per line and paralinguistic information was indicated between

[30] https://www.monotype.com/resources/articles/the-virtual-frontier/

brackets. Colors were used for character identification when necessary and subtitles for the narrators were in italics, as indicated in the Spanish standard for SDH (AENOR, 2003). All clips began with a 6-second, black screen with the text, “The test is about to start for (selected variable)” so that participants had enough time to accommodate the HMD before the video started.

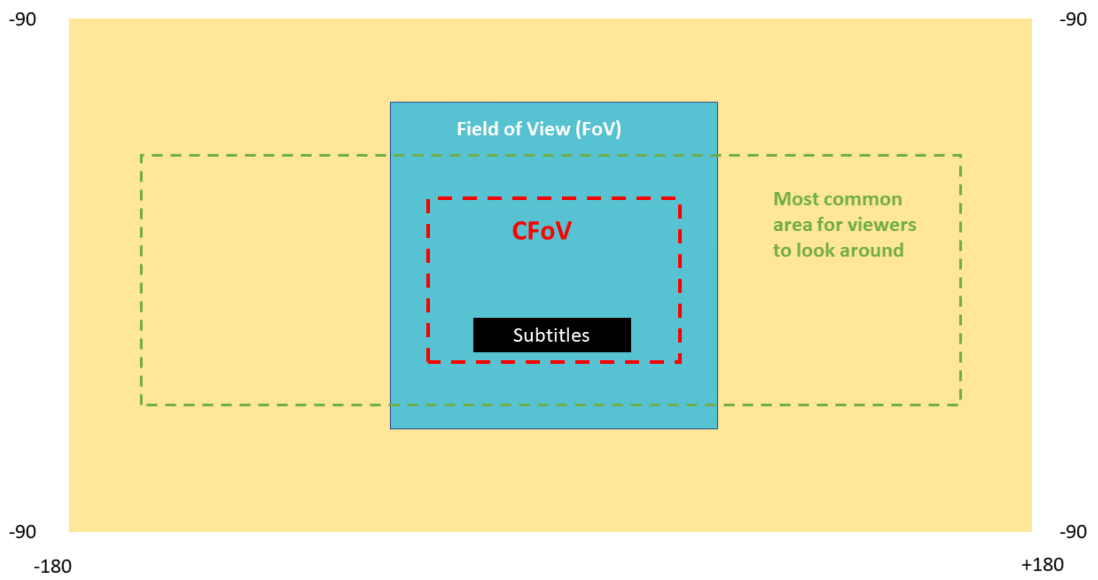


Figure 14. Diagram of 360º video view and its different areas.

An acclimation clip was prepared so that participants would become familiar with the HMD and 360º videos. The first video was 1-minute long and did not contain subtitles. It was assumed that most participants did not have extensive experience with the use of HMD and VR content and that the familiarization step was therefore necessary. The replies in the demographic questionnaire confirmed this assumption.

Two clips were used for the first condition (position). The videos were part of the series *Holy Land* created by Ryot in collaboration with Jaunt VR[31]. Specifically, episode 4 (4:13) and episode 5 (04:58) were used. Permission was given by the creators to use the videos in the study. The episodes depict Middle Eastern territories such as Jerusalem or Palestine. A narrator (voice-over) explains historical facts and the socio-cultural political situation. In the videos, different locations and landscapes are presented, inviting the viewers to explore the scenario. The hostess appears several times in the scene. However,

[31] https://www.vrfocus.com/2015/12/jaunt-and-ryot-announce-vr-documentary-series-holy-land/

her presence is not relevant to the narrative of the video. The clips were considered suitable for testing the first condition (position) due to the lack of multiple speakers and a range of landscapes, highlighting the exploratory nature of 360º videos. In this sense, participants could focus on exploring the scenarios in the clips and read the subtitles, without being required to make an effort to understand different speakers or narrative complexities.

For the second condition (guiding mechanisms), the award-winning short movie *I, Philip* created by ARTE[32] was used. The duration of the movie was 12 minutes and 26 seconds (excluding the credits) and was split into two parts to test both variables. In this case, permission to use the clip in the study was granted. The movie narrates the story of David Hanson who has developed the first-ever android, based on the personality and memories of the famous science fiction writer Philip K. Dick. It is a first-hand experience because the viewer sees the story through the eyes of Phil the robot. This clip was desirable for the test because it involved different speakers appearing in a variety of locations. This urged the participants to locate the speakers, allowing the different guiding methods to be tested.

5.3.3.3. Questionnaires

In order to measure the impact of conditions on immersion, the IPQ questionnaire was used. A review of presence questionnaires, such as Slater-Usoh-Steed Presence Questionnaire (Slater & Usoh, 1993), Presence Questionnaire (Witmer & Singer, 1998), ICT-SOPI (Lessiter, Freeman, Keogh & Davidoff, 2001) and IPQ Questionnaire (Schubert, 2003) was conducted in order to select the one most suitable for the study. IPQ Questionnaire[33] was chosen because it is a 14-item questionnaire based on a 7-point Likert scale and includes different subscales that can be analyzed separately if necessary. First, it includes an independent item not belonging to any subscale that reports on general presence. Then, there are three subscales: spatial presence (that is, how much the viewer is feeling physically present in the virtual world), involvement (that is, how much attention the viewer is paying to the virtual world), and realness (that is, to what extent the viewer feels that the virtual world is real). The questionnaire has also been validated in previous studies with similar technology (i.e. desktop VR, 3D games, VR or CAVE-like systems, etc.). Other questionnaires such as the Presence Questionnaire by Witmer

[32] https://www.arte.tv/sites/en/webproductions/i-philip/?lang=en
[33] http://www.igroup.org/pq/ipq/index.php

and Singer (1998) include questions on physical interaction with the virtual world, and this was not suitable for our 360º content because the interaction is non-existent. The IPQ questionnaire was translated into Spanish for the purpose of this study.

For each part of the test, an ad-hoc questionnaire in Spanish was created. In regard to position, participants were asked about their preferences, any difficulties faced when reading or locating the subtitles, and whether or not they were obstructive or distracting (see Appendix A, a translation into English from the original Spanish questionnaire). Regarding guiding mechanisms, participants were also asked about their preferences, any difficulties experienced when locating the speaker and whether or not the guiding mechanisms were distracting (see Appendix B, a translation into English from the original Spanish questionnaire). Closed and open questions were used, as well as 7-point Likert-scales. To ensure that the wording of the questions was suitable, all questionnaires were reviewed by a professional educational psychologist who specializes in teaching oral language to persons with hearing loss.

5.3.3.4. Head tracking – CVR Analyzer

Rothe et al. (2018b) developed a tracking tool for head and eye movements in 360º content. According to the authors:

> […] the CVR-Analyzer, which can be used for inspecting head pose and eye tracking data of viewers experiencing CVR movies. The visualized data are displayed on a flattened projection of the movie as flexibly controlled augmenting annotations, such as tracks or heatmaps, synchronously with the time code of the movie and allow inspecting and comparing the users' viewing behavior in different use cases. (Rothe et al., 2018b, p. 127).

This technology was used in the study to track the head position/orientation of the participants, with the aim of verifying whether a movement pattern could be linked to each specific variable in order to triangulate results with the data from the questionnaires. The results will be discussed in section 4.

5.3.4. Participants

Forty participants (14 male and 26 female), aged 18 to 70 (M=37.4, Mdn=37; SD=14.8), took part in the test. It was decided that the sample should be balanced, including participants ranging from 18 to 60 years of age or more. Participants from different generations are likely to have different levels of technical knowledge as well as different habits when it comes to consuming subtitled content. Therefore, representation from different age groups was obtained.

Twenty-seven participants defined themselves as hearing, six as hearing-impaired and seven as deaf. Twenty participants were identified as persons with hearing loss, which can be confirmed by the responses to the following question: “Age in which your disability began” (answered by 20 participants with hearing loss). However, the question about defining themselves was more subjective; some people with hearing aids considered themselves to be hearing. In two cases, the users’ disability started from birth, for one user it began from 0 to 4 years old; for five users, from 5 to 12 years old; for five users, from 13 to 20 years old; for five users, from 21 to 40 years old and for two users, from 41 to 60 years old. The hearing participants were recruited via personal contacts and the participants with hearing loss were recruited via the association APANAH[34]. The decision was made to include both hearing users and users with hearing loss because as explained in section 1, subtitles are beneficial for a wide range of viewers (persons with hearing loss, non-native speakers, persons who are in noisy environments, etc.). In this regard, if part of the spectrum of subtitled content users had been excluded, the study would have been incomplete. All of the participants were Spanish native speakers. It was decided that participants with hearing loss should be users of the Spanish oral language as well as hearing participants in order to have a sample that was as homogeneous as possible in terms of language proficiency.

Regarding use of technology, the devices found to be used most frequently on a daily basis were mobile phone (38), TV (31), laptop (20), PC (13), tablet (11), game console (4) and radio (1). Most users (30) had never watched VR content before. The reasons behind a lack of prior experience with VR content were: they are not interested (3), they have not had the chance to use it (23), it is not accessible (3) or other reasons (4). Ten participants had previously experienced VR content. When directly asked if they were interested in VR content some users (14) were strongly interested, some (14) were interested, some (11) were indifferent and one was not interested. Five participants owned a device to access VR content (Plastic VR Headsets for smartphone and PlayStation VR).

In terms of content preferences; most users liked news (29), fiction (36), and documentaries (33), while some also liked talk shows (14), sports (17) and cartoons (18).

[34] https://www.apanah.com/

The users activate subtitles depending on the content and their individual needs. Thirteen of them stated that they do not need subtitles, 15 of them stated that it depends on the content and five of them stated that they always activate subtitles. Seven participants stated other reasons for not using subtitles: bad quality, fast reading speed, issues activating subtitles, subtitles are distracting, difficulty reading the subtitles and simultaneously viewing the images, they only use subtitles when the hearing aid is not available. Regarding how many hours a day they watch subtitled content; 13 participants indicated zero; 11, less than one hour; seven, from one to two hours; eight, from two to three hours and one participant indicated four hours or more. When asked why they used subtitles (more than one reply was permitted), 22 indicated that they helped them understand, three participants indicated that it is the only way to access the dialogue, 11 participants said that they use them for language learning and 11 participants stated that they never use subtitles.

5.4. Results

The results from the test are reported in this section.

5.4.1. Demographics

In this section, some correlations found in the demographic questionnaire are reported. The Spearman test was used for calculating correlations.

Different correlations were found with age and the use of certain devices. Younger participants used the laptop (r=-.360, p=.023) and the game console (r=-.350, p=.027) more than older participants. From the four participants that played games on a daily basis, three had previous experience using VR which could indicate that VR content is mainly consumed in gaming platforms. Regarding the previous experience with VR contents, it was found that younger participants (M=27.8) had previous experience (r=-.503, p=.001). A correlation was found regarding interest in VR content and the type of participant. Participants with hearing loss were more interested in VR content (M=4.3) than hearing participants (M=3.75) (r=-.314, p=.049). This was not related to age (participants with hearing loss age M= 42.55; hearing participants age M=37.55).

Of the 27 participants consuming subtitled content on a daily basis, 12 had hearing loss and 15 were hearing. Of the 13 participants that did not consume subtitles on a daily basis, eight were persons with hearing loss and five were hearing. A correlation was found between age and the number of hours of subtitled content watched by participants. The younger the participant, the more subtitled content they watched on a daily basis (*r*=-.592, p=.000). Participants who watched subtitled content every day presented a mean of 34.2 years old and participants who never watched subtitled content had a mean of 52.2. Also, there was a correlation between the use of a laptop and the subtitled content that participants watched on a daily basis (*r*=.374, p=.018). From the 20 participants using a laptop every day, 17 watched subtitled content on a daily basis. From the 20 participants who indicated that they did not use the laptop every day, 10 stated that they never watched subtitled content.

5.4.2. Preferences

In this section, results from the preferences questionnaires in the two parts of the test are reported. Statistical comparisons were calculated using a Wilcoxon signed-rank test. Correlations were calculated using the Spearman test and are reported when statistically relevant.

Always-visible vs fixed-position

Thirty-three participants preferred the always-visible subtitles and seven participants preferred the fixed-position subtitles. According to the participants, with always-visible subtitles they felt more freedom to explore and could comfortably access the content of the subtitles and video scenes without missing details. Participants in favor of fixed-position subtitles stated that they could read them more easily when comparing them to always-visible subtitles. Moreover, they attested that the dizziness effect was minimized with fixed-position subtitles. Some users suggested making the always-visible subtitles more static and less bumpy because the movement hindered their readability and was sometimes distracting.

A difference was found regarding the ease of locating the subtitles in the videos based on a 7-point Likert scale (7 being “the easiest”, 1 being “the most difficult”). Always-visible subtitles (M=6.32; SD=1.5) were considered easier to find than fixed-position subtitles (M=4.25; SD=1.61). This difference is statistically significant (Z=-3.986, *p*=.000, ties=5). A correlation was found with the demographic profile: participants with hearing loss (M=4.85) considered the fixed-position subtitles easier to find than the hearing participants (M=3.65) (*r*=-.368, p=.019). Again, a difference was encountered when asking about the ease of reading the subtitles in the videos, based on a 7-point Likert scale (7 being “the easiest”, 1 being “the most difficult”). Always-visible subtitles

(M=5.72; SD=1.88) were considered easier to read than fixed-position subtitles (M=4.77; SD=1.91). However, this difference is not statistically significant (Z=-1.919, *p*=.055, ties=9).

	Easy to find (7 "the easiest", 1 "the most difficult")	Easy to read (7 "the easiest", 1 "the most difficult")	Level of obstruction (7 "not at all", 1 "yes, very much")	Level of distraction (7 "not at all", 1 "yes, very much")
Always-visible	6.32	5.72	5.5	5.4
Fixed-position	4.25	4.77	5.87	3.77

Table 13. Summary of results (mean) in closed questions for the preference questionnaire comparing always-visible and fixed-position subtitles.

When asked whether always-visible subtitles were obstructing important parts of the image based on a 7-point Likert scale (7 being "not at all", 1 being "yes, very much"), participants felt that fixed-position subtitles (M=5.87; SD=1.28) were slightly less obstructive than the always-visible subtitles (M=5.5; SD=1.9). However, the difference is not statistically significant (Z=-1,123, *p*=.261, ties=23). Two correlations were found: participants with hearing loss (M=6.2) considered always-visible subtitles to be less obstructive than hearing participants (M=4.8) (*r*=-.452, p=.003). Participants with hearing loss (M=6.5) also considered fixed-position subtitles less obstructive than the hearing participants (M=5.25) (*r*=-.552, p=.000). A difference was found when participants were asked if the subtitles were distracting them from what was happening in the video, based on a 7-point Likert scale (7 being "not at all", 1 being "yes, very much"). The always-visible subtitles (M=5.4; SD=1.98) were considered to be less distracting than the fixed-position subtitles (M=3.77; SD=2.33). This difference is statistically significant (Z=-2,696, *p*=.007, ties=13). Another correlation was found here: participants with hearing loss (M=4.8) considered the fixed-position subtitles to be less distracting than the hearing participants (M=2.75) (*r*=-.397, p=.011). In Table 13, a summary of the results for preference regarding always-visible and fixed-position subtitles can be found.

Arrow vs radar

Thirty-three participants preferred arrows and seven participants preferred the radar as a guiding method. More participants with hearing loss (six out of seven) preferred the radar as a guiding mechanism. Participants who favored the arrows argued that this guiding method is more intuitive, direct, comfortable, less invasive and less distracting. Furthermore, during the test several participants reported that they did not understand the radar. Participants who preferred the radar argued that it provides more spatially accurate information. One participant suggested that a short explanation about how the radar works before using it could improve the user perception of this mechanism. Two participants suggested moving the radar from the right-hand side to the center (maybe directly below the subtitle) in order to improve usability. Four participants, despite having chosen one method or the other, stated that they would prefer to not use guiding mechanisms and that the color identification was sufficient.

	Easy to find the speaker (7 “the easiest”, 1 “the most difficult”)	Level of distraction (7 “not at all”, 1 “yes, very much”)
Arrows	6.12	6.42
Radar	3.9	3.9

Table 14. Summary of results (mean) in closed questions for the preference questionnaire comparing subtitles with arrows and radar.

A difference in the results was found when asking about how easy it was to find the speaker using the arrow guiding method, based on a 7-point Likert scale (7 being “the easiest”, 1 being “the most difficult”) (see Table 14). Participants felt that the speaker was easier to find with the arrows ($M=6.12$; $SD=1.26$) than with the radar ($M=3.9$; $SD=2.24$). The difference is statistically significant ($Z=-4.166$, $p=.000$, ties=10). A difference was also found regarding the level of distraction with each method, based on a 7-point Likert scale (7 being “not at all”, 1 being “yes, very much”). Arrows ($M=6.42$; $SD=1.03$) were considered to be less distracting than radar ($M=3.9$; $SD=2.46$). The difference is statistically significant ($Z=4.125$, $p=.000$, ties=12). A correlation was found here: the four participants that played video games on a daily basis found the radar less distracting ($M=1$) than the rest of the participants ($M=4.22$) ($r=-.460$, $p=.003$).

5.4.3. Presence

In this section, results from the IPQ questionnaires for each video are analyzed and compared. Statistical comparisons were calculated using a Wilcoxon signed-rank test.

Always-visible vs fixed-position

A comparison between always-visible and fixed-position subtitles was made in the IPQ and the results are as follows (see Table 15): spatial scale (Z=-1.791, p=.073, ties=7), involvement scale (Z=-1.229, p=.219, ties=8), realness scale (Z=-.064, p=.949, ties=14). The test indicated that the differences between the results were not statistically significant. However, for the general presence item, the test indicated that the difference between the results is statistically significant (Z=-2.694, p=.007, ties=17). This means that the fixed-position subtitles had a negative impact on the presence of the participants.

	General Presence	Spatial Presence	Involvement	Experienced Realism
Always-visible	4.7	3.74	3.31	2.36
Fixed-position	3.95	3.43	3.38	2.49

Table 15. IPQ results (mean) comparing always-visible and fixed-position subtitles.

Arrow vs radar

A comparison between arrow and radar methods was made in the IPQ and the results are as follows (see Table 16): general presence item (Z=-1.852, p=.064, ties=21), spatial scale (Z=-1.000, p=.317, ties=13), realness scale (Z=-1.430, p=.153, ties=7). The test indicated that the differences between the results were not statistically significant. However, for the involvement scale, the test indicated that the difference between the results is statistically significant (Z=-2.138, p=.033, ties=12). In this case, the radar had a negative impact on the involvement of the participants.

	General Presence	Spatial Presence	Involvement	Experienced Realism
Arrow	4.6	3.66	3.5	2.49
Radar	4.27	3.49	3.27	2.47

Table 16. IPQ results (mean) comparing subtitles with arrows and radar.

5.4.4. Head tracking

The qualitative results from the CVR Analyzer shed some light on the participants' viewing patterns. The data gathered from this tool is visual, using heat maps to represent the participants' head movements. A difference in viewing patterns was observed between the always-visible and fixed-position subtitles. With the always-visible subtitles, the head movements were more scattered, showing that participants explored the scenes with more freedom (see Images 16 and 18). In the case of fixed-position subtitles, it can be observed that the participants were more concentrated on one specific point, that is, the central subtitle (see Images 15 and 17).

Figure 15. The pattern of head movements from *Holy Land* with fixed-position subtitles.

Figure 16. The pattern of head movements from *Holy Land* with always-visible subtitles (exactly the same scene as Figure 15).

Figure 17. The pattern of head movements from *Holy Land* with fixed-position subtitles.

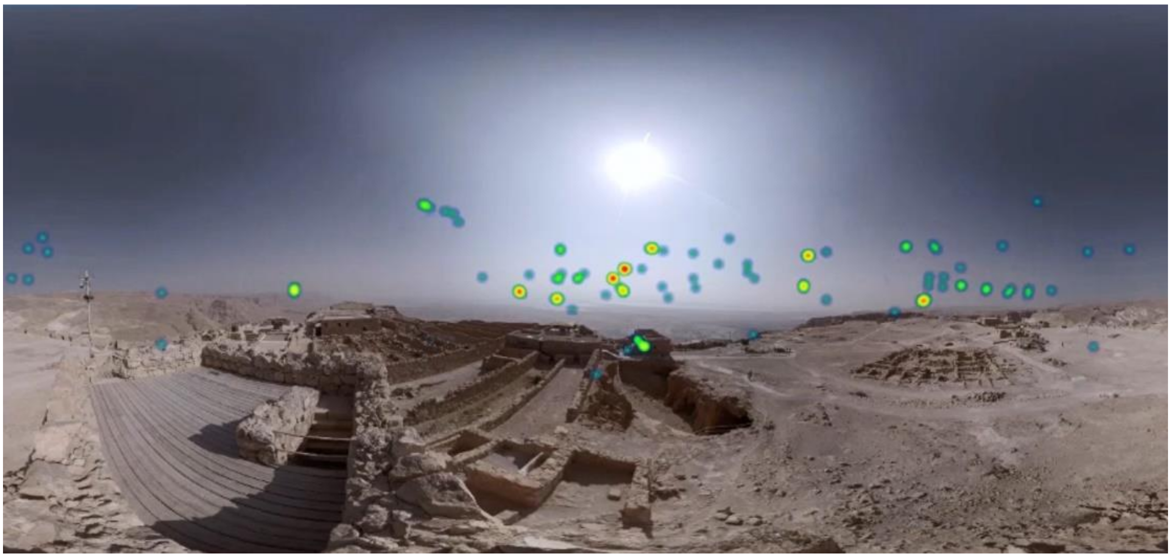


Figure 18. The pattern of head movements from *Holy Land* with always-visible subtitles (exactly the same scene as Figure 17).

Significant differences were not found when comparing viewing patterns for arrows and the radar. It was observed that participants directed their attention to the speakers and their faces (see Figure 19 and 20), as is usual in 2D content (Kruger, 2012). No significant differences were observed between hearing participants and participants with hearing loss for any of the variables.

Figure 19. *I, Philip* scene with arrows.

Figure 20. *I, Philip* scene with radar.

5.5. Discussion

Always-visible subtitles are the preferred option by 82.5% of participants. Subtitles in this position were considered to be easier to read and less distracting than fixed-position subtitles. Participants stated that with these subtitles, they had more freedom to look around the scenes without feeling restricted. In general, always-visible subtitles performed better in the IPQ questionnaire whereas fixed-position subtitles had a negative impact on presence. According to the comments made by participants in the open questions, this could be because they felt less free to explore the 360º scene and claimed to have missed parts of the subtitle content. Moreover, as reported above, participants encountered more difficulty when locating and reading the subtitles in this mode, and also considered them to be more distracting. This extra effort could have therefore caused a negative impact on the feeling of being present in the scene.

Head tracking data also supported the qualitative data, with the participants showing more exploratory behavior in the always-visible condition compared to the fixed-position condition. Results from the head-tracking tool were similar to the results for the study by Rothe et al. (2018a). They also found that with always-visible subtitles, the head movements are more scattered. In this study, unlike for others before it (Brown et al., 2018), font with an outline was used instead of a background box which resulted in less obstructive subtitles. However, the readability of the subtitles can still be hindered, due to and depending on the video background. It would be interesting to carry out further research using artificial intelligence technology that could potentially add a background box automatically when the contrast is poor, as explained by Sidenmark et al. (2019). Also, always-visible subtitles were considered slightly more obstructive than fixed-position. Further studies could be carried out to test different font sizes or different positions (for example, lower in the FoV, or above the FoV).

The arrows were the preferred option for guiding mechanisms with 82.5% of participants favoring it over radar. Most people considered them to be more intuitive and comfortable than the radar, which required an extra cognitive effort (understanding how it works), something that was not called for in the arrows condition. The radar was also considered more distracting than the arrows. In this regard, it can be asserted that in 360º content, the simpler solution is preferred because moving and interacting with the virtual world is already complex for the viewers and the cognitive workload should therefore be reduced. The sample of participants who preferred the radar happened to also be those who played video games on a daily basis (four participants, three of whom were participants with hearing loss). Previous experience with video games and the use of a radar for spatial orientation might have an impact on the users' preferences in terms of guiding mechanisms for VR content.

It was found that the radar had a negative impact on the involvement of the participants. Participants reported that the radar was difficult to understand and some of them felt frustrated by it when watching the stimuli at the beginning. At certain points during the story, the participants were busy trying to understand how it worked as opposed to following the plot, which therefore led to a loss of content. As stated before, one of the prerequisites for implementing subtitles in immersive content is that they do not disrupt the immersive experience. In this case, the complexity of the radar interfered with users' immersion. Some participants claimed that they would prefer no-guiding mechanism at all. If 360º videos are post-produced to a high standard from a filmmaking perspective, it is possible that there will be no need for guiding mechanisms because when changing from one scene to another, speakers will be located in the center of the FoV. Therefore, non-intrusive mechanisms such as arrows might be more advisable for that type of content. Guiding methods in this medium such as arrows and radar could be used to improve not only subtitles but also narrative. For example, radar or arrows could indicate that the action or point of interest is located in a specific location outside the FoV. In general, the most optimal solution can vary depending on the type of content.

Despite the fact that it was not the aim of the study, replies from the demographic questionnaire also shed some light on the profile of subtitled content users. Hearing participants (15) consumed more subtitled content than participants with hearing loss (12) on a daily basis. Moreover, younger participants (M=34.2) watched subtitled content every day, and a correlation was found between the use of a laptop and the consumption of subtitled content (r=.374, p=.018). Younger (hearing or with hearing loss) viewers who use advanced technology and devices on a daily basis are more active users of subtitles than older users, who mainly consume audiovisual content on TV sets. This indicates that subtitle consumption habits are evolving. Consumption habits do not necessarily depend on the level of hearing loss but on age or familiarity with the technology (Ulanoff, 2019). This might be for consideration when profiling participants for future research in subtitle studies.

5.6. Conclusions

In this article, the potential of immersive content and the need to make it more accessible with subtitles has been highlighted. A review of the main studies regarding subtitles in immersive content has been presented, showing that further research is needed. The study has been presented as well as the results. The primary aim was to compare different subtitling solutions in two main areas of 360º content: position and guiding mechanisms. Feedback on preferences, immersion and head movements from 40 participants was gathered to clarify the open questions. Results have shown that always-visible subtitles, that is, subtitles that are always in front of the viewers, are the preferred option and in turn, more immersive. Regarding guiding mechanisms, arrows were preferred over radar due to their simple and efficient design that was easier to understand.

The present study also had some limitations regarding the type of content. For each variable, a specific type of content was used: a documentary and a short science-fiction movie. This might have had an impact on the results and a replication of this study with different content is encouraged. However, the agreement of participants on the preferred option was clear. It can be concluded from the results that the subtitles that best perform in terms of usability and immersion are always-visible subtitles with arrows. From a production point of view, this type of subtitles is more scalable and easier to produce than burnt-in subtitles created manually. A 360º subtitle editor is being developed in the ImAc project for this purpose (Agulló, in press).

Always-visible subtitles with arrows could be considered as good practice guidelines for subtitles in 360º videos, especially for exploratory content such as documentaries or fast action and interactive content such as video games. However, results from the present study are not final and further research is encouraged. Depending on the content, subtitles located close to the speaker in a fixed position could perform better. They could also be considered more aesthetically suitable and better integrated within the scenes, as reported by Rothe et. al (2018a). Fixed-position subtitles could be created manually in the creative part of the production of the film, as suggested in the accessible filmmaking theory by Romero-Fresco (2019). Even a combination of both methods (always-visible and fixed-position subtitles) could work, as suggested by Rothe et al. (2018). For exploratory content or fast action, fixed-position subtitles might not be the most suitable because fixed-position subtitles could hinder the immersive experience. Finally, studies into the reading speed of subtitles in immersive content are required as user interaction with content in this new medium is different from that of traditional screens and the current rules therefore do not necessarily apply. The processing of subtitles may be affected when exploring VR scenes with an HMD as it may entail a heavier cognitive workload compared to traditional media. Further studies with eye-tracking could clarify current and future questions.

Subtitles are an important part of audiovisual media and beneficial to a wide range of viewers. Virtual and augmented reality applications in our society are numerous, so further studies are necessary if user experience and access to immersive content are to be improved. In this study, it has been proven that subtitles can be implemented in this type of content and the preferred solution (always-visible with arrows) has been provided.

5.7. Acknowledgments

Authors would like to thank ImAc project partners, especially i2Cat members, Mario Montagud and Isaac Fraile, for their technical contributions; to the content creators Ryot and ARTE for giving us permission to use their fantastic 360º content; also, to APANAH for their inestimable help carrying out the tests (thanks to Anto, Rafa, Isabel, Graciela, Vicente, and everyone); thanks to Sylvia Rothe for sharing her expertise and for the CVR tool; thanks to Irene Tor-Carroggio and José Ramón Calvo-Ferrer for their help with the statistical analysis; and, especially, to all participants, without their involvement, this study would not have been possible.

This work has been funded by the European Union's Horizon 2020 program, under agreement nº 761974 (ImAc project).

Belén Agulló and Anna Matamala are members of TransMedia Catalonia, an SGR research group funded by "Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya" (2017SGR113). This article is part of Belén Agulló'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 the Universitat Autònoma de Barcelona.

References

Agulló, B. (in press). Technology for subtitling: a 360-degree turn. *Hermeneus*, *22*.

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

Agulló, B. & Matamala, A. (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*, *32*, 217–235. Retrieved from: http://www.jostrans.org/issue32/art_agullo.pdf

Agulló, B., Montagud, M., & Fraile, I. (in press). Making interaction with virtual reality accessible: rendering and guiding methods for subtitles. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *33*(4).

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

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*. Retrieved from: https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?Tipo=N&c=N0029761

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). Bern, Switzerland: Peter Lang.

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2017). Subtitles in 360-degree video. *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video, Hilversum, Netherlands*, 3–8. doi: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) Retrieved from BBC Research & Development: https://www.bbc.co.uk/rd/publications/whitepaper330

Catapult Digital (2018). *Growing VR/AR companies in the UK: A business and legal handbook*. Retrieved from: https://www.pwc.co.uk/intelligent-digital/vr/growing-vr-ar-companies-in-the-uk.pdf

D'Ydewalle, G., Pollet, J., & van Rensbergen, J. (1987). Reading a message when the same message is available auditorily in another language: the case of subtitling. In J.K. O'Regan, & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition*, (pp. 313–321). North Holland, Netherlands: Elsevier Science Publishers.

Díaz-Cintas, J. (2013). The technology turn in subtitling. In M. Thelen and B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: Part 9*, (pp. 119–132). Maastricht, Netherlands: Zuyd University of Applied Sciences.

Díaz-Cintas, J. (2014). Technological strides in subtitling. In Chan Sin-wai (ed.), *The Routledge encyclopedia of translation technology*, (pp. 632–643). Abingdon, UK: Routledge.

Díaz-Cintas, J., Orero, P., & Remael, A. (Eds.) (2007). *Media for all: subtitling for the deaf, audio description, and sign language*. Amsterdam, Netherlands: Rodopi.

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

Kruger, J. L. (2012). Making meaning in AVT: eye tracking and viewer construction of narrative. *Perspectives: Studies in Translation Theory and Practice*, *20*(1), 67–86. doi: 10.1080/0907676X.2011.632688

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, USA*, 6559–6568. doi:10.1145/3025453.3025772

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators, and Virtual Environments 10*(3), 282–297. doi:10.1162/105474601300343612

MacQuarrie, A., & Steed, A. (2017). Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. *Proceedings of 2017 IEEE Virtual Reality (VR), USA*, 45–54. doi:10.1109/VR.2017.7892230

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

Matamala, A., & Orero, P. (Eds.) (2010). *Listening to subtitles: Subtitles for the deaf and hard of hearing*. Bern, Switzerland: Peter Lang.

Perego, E., Del Missier, F., Porta, M., & Mosconi. M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, *13*(3), 243–272. doi:10.1080/15213269.2010.502873

Perkins Coie LLP (2018). *2018 augmented and virtual reality survey report: Insights into the future of AR/VR*. Retrieved from https://www.perkinscoie.com/images/content/1/8/v2/187785/2018-VR-AR-Survey-Digital.pdf

Romero-Fresco, P. (2009). More haste less speed: edited vs. Verbatim respeaking. *Vigo International Journal of Applied Linguistics*, *6*, 109–133. Retrieved from: https://dialnet.unirioja.es/servlet/articulo?codigo=3207248

Romero-Fresco, P. (Ed.) (2015). *The reception of subtitles for the deaf and hard of hearing in Europe*. Bern, Switzerland: Peter Lang.

Romero-Fresco, P. (2019). *Accessible filmmaking: Integrating translation and accessibility into the filmmaking process*. Abingdon, UK: Routledge.

Rothe, S., Höllerer, T., & Hussmann, H. (2018b). CVR-analyzer: a tool for analyzing cinematic virtual reality viewing patterns. *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, Cairo, Egypt*, 127–137. doi:10.1145/3267782.3274688

Rothe, S., Tran, K., & Hussmann, H. (2018a). Dynamic subtitles in cinematic virtual reality. *ACM TVX 2018, Seoul, Republic of Korea*, 209–214. doi:10.1145/3210825.3213556

Schubert, T. (2003). The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, *15*, 69–71. doi:10.1026//1617-6383.15.2.69

Sidenmark, L., Kiefer, N., & Gellersen, H. (2019). Subtitles in interactive virtual reality: Using gaze to address depth conflicts. *Proceedings of Workshop on Emerging Novel Input Devices and Interaction Techniques*, Osaka, Japan. Retrieved from: https://eprints.lancs.ac.uk/id/eprint/132411/

Slater, M., & Usoh, M. (1993). Representations systems, perceptual position, and presence in immersive virtual environments. *Presence* *2*(3), 221–233. doi:10.1162/pres.1993.2.3.221

Szarkowska, A., Krejtz, I., Kłyszejko, Z., & Wieczorek, A. (2011). Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American Annals of the Deaf*, *156*(4), 363–378. doi:10.1353/aad.2011.0039

Szarkowska, A., Krejtz, I., Pilipczuk, Dutka, L., & Kruger, J.L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures*, *17*(2), 183–204. doi:10.1556/084.2016.17.2.3

Ulanoff, L. (2019, January 17). Why Gen Z loves closed captioning [Web log post]. Retrieved from https://onezero.medium.com/why-gen-z-loves-closed-captioning-ec4e44b8d02f

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence Teleoperators & Virtual Environments*, *7*(3), 225–240. doi:10.1162/105474698565686

Appendices

Appendix A

1. With which system do you prefer to read subtitles in 360º videos?

a) Always visible	b) Fixed position

2. Please, explain why you prefer this option.

3. What do you think could be improved, and how?

4. Did you find it easy to find the always-visible subtitles?

1- very difficult	**2**	**3**	**4**	**5**	**6**	**7- very easy**

5. Did you find it easy to find the fixed-position subtitles?

1- very difficult	**2**	**3**	**4**	**5**	**6**	**7- very easy**

6. Did you find it easy to read the always-visible subtitles?

1- very difficult	**2**	**3**	**4**	**5**	**6**	**7- very easy**

7. Did you find it easy to read the fixed-position subtitles?

1- very difficult	**2**	**3**	**4**	**5**	**6**	**7- very easy**

8. Do you think always-visible subtitles obstruct important parts of the image?

1- yes very much	**2**	**3**	**4**	**5**	**6**	**7- not at all**

9. Do you think fixed-position subtitles obstruct important parts of the image?

1- yes very much	**2**	**3**	**4**	**5**	**6**	**7- not at all**

10. Do you think that always-visible subtitles distract you from the video?

1- yes very much	**2**	**3**	**4**	**5**	**6**	**7- not at all**

11. Do you think that fixed-position subtitles distract you from the video?

1- yes very much	**2**	**3**	**4**	**5**	**6**	**7- not at all**

12. Other comments:

Appendix B

1. Which system do you prefer to indicate to you the location of the character speaking?

a) Arrow	b) Radar

2. Please, explain why you prefer this option.

3. What do you think could be improved, and how?

4. Did you find it easy to find the person who speaks with the arrows?

1- very difficult	**2**	**3**	**4**	**5**	**6**	**7- very easy**

5. Did you find it easy to find the person who speaks with the radar?

1- very difficult	**2**	**3**	**4**	**5**	**6**	**7- very easy**

6. Did the arrows seem to distract you from what was happening in the video?

1- yes very much	**2**	**3**	**4**	**5**	**6**	**7- not at all**

7. Did the radar seem to distract you from what was happening in the video?

1- yes very much	**2**	**3**	**4**	**5**	**6**	**7- not at all**

8. Other comments:

Chapter 6. Summary

This dissertation analyzed the presentation of subtitles in 360° videos from a user-centered perspective. In particular, it evaluated different subtitling strategies in 360° content. Three main goals were established in this dissertation: 1) To identify challenges and subtitling strategies for immersive content; 2) To identify professional subtitlers' needs for immersive content; and 3) To evaluate different subtitling strategies (related to position and directions) in immersive content (360° videos) among subtitle users (both hearing and with hearing loss), comparing levels of presence and user preferences. To fulfill these goals, one descriptive study and three experimental studies were carried out.

The first qualitative experimental study was a focus group, which was conducted with both subtitle consumers and professional subtitlers in order to gather feedback about how they were expecting to see subtitles in immersive content and what they would need in a subtitle editor for this type of content. Subtitle consumers, on the one hand, agreed that they would like to see subtitles as similarly as possible to how they appear in traditional audiovisual media such as TV, following the current standards. However, they were also open to modifications, such as the use of emojis to represent sounds. Moreover, they highlighted the importance of including direction indications in 360° videos, such as arrows, radar or text in brackets. Professional subtitlers, on the other hand, stated that they would need an editor similar to existing ones, but enabling 360° videos and immersive audio to be played. They agreed that they would like to access the 360° videos in a player showing an angle of the 360° view such that they could navigate through the 360° angles using their mouse or keyboard. They also underlined the importance of being able to test the results with an HMD, as well as seeing the preview on a flat screen. The feedback gathered from the focus group was used to create both subtitling solutions and a usable subtitle editor for 360° videos. The subtitling solutions were then further tested and redefined until acceptable strategies were put in place.

Secondly, a descriptive study was carried out which consisted in analyzing a 360° video multimedia corpus to explore commercially available subtitling solutions. The videos were selected from the NYT and BBC platforms, then carefully analyzed in search of subtitling solutions that could be implemented in the study. The corpus analysis showed that subtitles were not only being used the way they are in traditional media (i.e., as a translation or transcription of the original soundtrack), but also to contribute to the narrative of the stories. The position of the subtitles was mainly found to be evenly spaced every 120° or in a fixed position, rather than always visible. Moreover, subtitles were not located at the bottom of the FoV as in traditional media. They were mainly found at the top, next to a speaker or object, or in the middle. The information gathered from this analysis was used to design the subtitling strategies that were subsequently tested in the reception study.

Thirdly, a reception study was conducted to test the prototype of the subtitle editor and gather in-depth feedback from 27 professional subtitlers. The aim of the study was to establish the priorities and needs of professional subtitlers when generating subtitles in 360° videos. The SUS questionnaire was administered to participants to measure the usability of the prototype, along with open questions to gather qualitative feedback from them. The study led to findings that will help define the features of 360° subtitle editors, such as the need to indicate the angle where the speaker is located (or the possibility of selecting an off-camera speaker), the fact that shot detections do not work as they do in traditional media given that shot changes depend on where the viewer decides to look, or the importance of enabling different preview modes for this type of content.

Finally, another reception study was carried out to evaluate different subtitling strategies in 360° content, which involved comparing levels of presence and user preferences. The following strategies were compared: 1) position: always-visible subtitles versus fixed-position subtitles every 120°; and 2) directions: arrows versus radar. The strategies were tested on 40 participants (20 hearing and 20 with hearing loss). This reception study had the aim of comparing commercially available subtitling solutions (found on the NYT and BBC platforms) with solutions that were developed based on user feedback. The results showed that the solutions based on user feedback – that is, with always-visible subtitles and arrows – were the preferred option among users, and these solutions reported higher levels of presence in the IPQ questionnaire. Moreover, the head tracking analysis confirmed that users moved more freely with always-visible subtitles

.

Chapter 7. Discussion and conclusions

Subtitling has become one of the most prolific fields of research within AVT Studies. According to the Bibliography on Translation and Interpreting (BITRA) database[35], there are 5,341 entries featuring the keyword “audiovisual.” Out of that number, publications with the subfield “accessibility” amount to 989 entries, followed by “dubbing” with 1,136 entries, and “subtitling” with 1,893 entries. Thus, according to this database, subtitling is the most researched area within AVT Studies. However, no studies can be found therein addressing the topics of “virtual reality,” “immersive content” or “360° videos” together with “subtitling.” VR technologies, as well as immersive content, are becoming gradually more present in our society. Immersive content can be found in different fields such as video games, museums, education, or broadcasting, as is the case of the 360° videos by the NYC and the BBC analyzed and discussed in Chapter 4.

Subtitles are a necessary instrument for accessing audiovisual content. Viewers with hearing loss, viewers who cannot understand the language of the content, or viewers who simply like to use subtitles must be given the option of choosing subtitles in immersive content, as they would in other mainstream media such as television programs, video games or VoD platforms. This dissertation aimed to fill the gap between the needs of the viewers and the current technological shortcomings that fail to provide standardized subtitling practices in immersive content. The lack of subtitling standards in immersive content and of commercially available technologies to facilitate their creation proved to be a challenge. The main question posed by this dissertation was how to present subtitles in immersive content in a way that is suitable for end users and does not disrupt their sense of presence. To find an answer to that question, various goals were established:

1. **To identify challenges and subtitling strategies for immersive content.** With the following specific objectives:
    1.1. Gathering feedback from subtitle users about their expectations and recommendations regarding the consumption of subtitles in immersive content (360° videos).
    1.2. Describing the state of the art of immersive technology in regard to media applications.

[35] https://aplicacionesua.cpd.ua.es/tra_int/usu/buscar.asp?idioma=en (consulted on 12.03.2019)

2. **To identify professional subtitlers' needs for immersive content.** With the specific objectives of:

    2.1. Gathering feedback from professional subtitlers about their expectations and recommendations regarding the creation of subtitles for immersive content (360° videos).

    2.2. Testing a 360° content subtitle editor prototype with professional subtitlers in order to gather their feedback in regard to the needs for subtitling 360° videos.

3. **To evaluate different subtitling strategies in immersive content (360° videos).** Specifically:

    - Comparing levels of presence and preferences among subtitle users (both hearing and with hearing loss) when the position of subtitles is either (a) always visible or (b) fixed position every 120°.
    - Comparing levels of presence and preferences among subtitle users (both hearing and with hearing loss) when directions are transferred using either (a) arrows or (b) a radar.

The third goal involved defining the subtitling strategies to be evaluated, but it was not a straightforward task because we were lacking a) commercially available solutions for subtitles in 360° videos, and b) a subtitle editor to create subtitles for 360° content. Therefore, the first two main goals were initially established. Overcoming those obstacles was only possible thanks to: 1) collaboration with the ImAc Project partners who provided the technical setup to develop (subtitle editor) and integrate (ImAc Player) subtitles in immersive content, as well as the content itself; 2) a global user-centered approach that involved users in the study from the start; and 3) an exploratory approach to the state of the art of immersive technologies. After the first two goals were accomplished, we were then able to define and evaluate different subtitling strategies for 360° videos in order to find out which option was preferred by users and achieved the highest levels of presence.

This concluding chapter will review the most relevant findings and main contributions of this study, as well as the limitations encountered and future lines of research.

8.1. Theoretical framework: Exploring a user-centered approach to AVT Studies

This section is related to the following objectives:

1.1. Gathering feedback from subtitle users about their expectations and recommendations regarding the consumption of subtitles in immersive content (360° videos).

2.1. Gathering feedback from professional subtitlers about their expectations and recommendations regarding the creation of subtitles for immersive content (360° videos).

Research in AVT Studies has practical applications in our daily lives because audiovisual content is an important part of present-day society. There are several undoubtedly user-centered AVT studies, in particular those involving reception studies. The focus of the scholars in this field is often the audience, how they react to certain translation modes, their comprehension, and enjoyment or feelings towards different translation modes in audiovisual media. Reception studies with end users are, therefore, common in this field (Romero-Fresco, 2015; Di Giovanni & Gambier, 2018).

Even though reception studies within AVT are prolific, a systematic and mature theoretical framework with theories to support the research questions and development of such studies is still work in progress. Pérez-González (2014, p. 97) argues that "a fully fledged theory is yet to emerge within and for audiovisual translation theories." The author differentiates between "allochthonous models of translation (approaches imported from the wider context of translation studies or beyond)" and "autochthonous counterparts (developed from within and for audiovisual translation studies)." (Pérez-González, 2014, p. 97) He states that most AVT studies have relied on allochthonous models at the expense of autochthonous models.

Faced with this lack of an autochthonous and established theoretical framework for AVT studies, we decided to adopt the UCT model. The reasons behind adopting this allochthonous model were manifold. Firstly, the new medium under research (i.e. immersive media, in particular 360° videos) was not mainstream yet. Therefore, standard subtitling solutions were unavailable at the time the study was carried out. Designing subtitling solutions for an audiovisual medium that differs from traditional media such as

TV content, cinema, or video games cannot be accomplished in isolation. This study required collaboration with other fields of research and professionals, as well as the involvement of end users. The process of designing the subtitles entailed a comprehensive study of the state of the art of subtitles in immersive content (related studies, current applications in media, etc.), a multidisciplinary team of professionals who could envisage the possible solutions and define how to implement them with the right technology and, most of all, end users as well as professional subtitlers to guide the creation of new subtitling strategies and a dedicated 360°-video subtitle editor, contributing their suggestions, needs, and expectations regarding subtitles in immersive content.

Secondly, it was important to pay special attention to the evolving and complex needs of subtitle consumers. Nowadays, profiling these users is as complex as ever. The responses to the demographic questionnaire of the reception study (see Chapter 5) showed that hearing participants (15) watched more subtitled content every day than participants with hearing loss (12). Previous studies found that some users with complete hearing loss would prefer sign language interpreting rather than subtitles (Arnáiz-Uzquiza, 2012). A younger hearing audience is using subtitles to better understand certain content and help them focus while multitasking (Ulanoff, 2018). All in all, viewing habits are changing and customization seems to be the natural evolution for subtitling practices, supported by technological advances. As put by Neves (2018b, p. 92),"[a]s user-centred technological environments become ever more ubiquitous, viewers will be able to choose specific formats that suit their personal needs," and we will move towards a "more encompassing framework characterized by the use of 'enriched (responsive) subtitles'."

Consequently, the approach to user profiling was updated for this subtitling reception study. Experimental design imposes some limitations. For example, in order for results to be comparable, the experiment's conditions must be the same for all subjects. To achieve that aim, apart from the experimental design itself, we need to control external variables. One external variable in subtitling reception studies is the language proficiency of participants. That is the reason why we decided to include hearing participants and participants with hearing loss whose main language was spoken Spanish, as opposed to including sign language speakers for whom spoken Spanish is a second language and, therefore, proficiency is not always ensured. We also included hearing participants because they too are subtitle consumers, and it would not have been a complete user-centered study if we had excluded a significant segment of end users.

The UCT model proved suitable for the abovementioned reasons and contributed to overcoming the challenges thanks to:

1. **A user-centered approach**: At the start of the project, we did not have a finalized product to test. Instead of creating a subtitling solution or subtitle editor in isolation, we decided to include users so that they could tell us how the product should be, according to their expectations, preferences, and needs. During the first focus group (Chapter 2), participants provided valuable feedback that was later implemented in the ongoing design of subtitles for 360° videos as well as in the subtitle editor. For example, subtitle users indicated that subtitles should be placed at the bottom of the FoV, in a fixed position and be visible at all times, and include a guiding method to indicate directions when the speaker is out of sight, such as arrows, a radar or text between brackets (to the left or to the right). In the reception study, participants chose the subtitling option that was designed following user feedback during the focus group, thus proving the suitability of the user-centered approach. Professional subtitlers also highlighted the importance of including a proper preview in the subtitle editor so that they could easily spot potential errors, and the need to have access to the original 360° immersive audio to identify where the sound comes from.
2. **A trial-and-error approach to gather iterative feedback**: Subtitle prototypes were designed based on user feedback and current solutions found in 360° videos. However, the validity of the prototypes could not be tested in a large reception study because we were not sure whether the solutions would work in a real scenario due to the newness of the technology. Therefore, usability testing was needed. Solutions were developed and tested, then redefined and tested until a usable solution was found (see Annex 1). This approach was adequate and prevented further issues. For example, the first prototypes developed for subtitles in 360° content were deficient because some people with astigmatism experienced issues like double vision and eye strain. The first prototype for the radar was considered insufficient by testers and was, therefore, redesigned in subsequent versions. Professional subtitlers also provided valuable iterative feedback (first with the focus group, then with the usability test for the subtitle editor prototype), which was implemented by the development team in order to improve the final editor.

The UCT model filled the current gap in AVT Studies, where technology and technological advances are not considered in any other theoretical AVT models. It is important to highlight that technology is an intrinsic part of AVT Studies. Subtitle editors are needed to create subtitles; audio processing tools are necessary to generate dubbed or voice-over versions; coding in video games is designed to trigger certain localized strings depending on the sex of the speaker. When we study AVT modes, it is important to focus not only on the translation itself, but also on how it is delivered or presented to the audience. We believe that, considering the nuances of the presentation of subtitles, this discipline falls within the field of AVT Studies. There are few studies on creative subtitling, for example (Foerster, 2010; McClarty, 2012, 2014; Fox, 2016a, 2016b, 2018). The focus of these studies is to understand which style of subtitles are preferred by the audience. It is important to understand the impact of subtitle presentation on the target audience. To that end, a user-centered approach can provide the tools to deliver a usable final product accepted by viewers, as has been proven by this dissertation.

This approach is innovative because, to the best of our knowledge, only two other studies have implemented UCT within AVT studies (Tuominen, 2016, 2018) (and the author of both studies is the co-author of the *User-centered translation* monographic). According to Tuominen (2018, p. 83), "it is possible to examine subtitles from the perspective of usability, seeking to find out how efficiently and successfully they allow the viewing event to proceed." That is precisely what has been done in this study: examining different subtitling solutions to determine which ones were more efficient and successful in delivering an adequate viewing experience in immersive content. Implementing usability tools in AVT Studies, subtitling in particular, and involving users from the very first steps of the research has proven to be an effective approach.

End users have been at the core of this study from the beginning. Firstly, they helped define the different subtitling solutions by providing their preferences and opinions in a focus group (see Chapter 2). The feedback from the focus group was used by the ImAc team to start drafting the first strategies for subtitles in 360° content, as well as for the subtitle editor prototype to create the subtitles (see Chapter 4). Then, the first strategies were tested on a limited group of hearing users with the aim of refining and redefining the provided solutions (see Annex 1). Improved versions of different subtitling solutions were once again tested on a limited group of users, including hearing participants and participants with hearing loss (see Annex 1). The iterative feedback gathered during the process was then used to generate functional and usable subtitles for 360° content. In the final reception study, the subtitles that had been designed following user feedback were compared to the subtitles that were used in commercial solutions found on the NYT and BBC 360° platforms. Those designed following user feedback were preferred and achieved a higher level of presence in the IPQ questionnaire (see Chapter 5). This proves that a user-centered approach provides valuable insights for the development of AVT translation modes such as subtitling, particularly when implementing subtitles before a medium becomes mainstream. The same user-centered approach was applied to develop the subtitle editor. Professional subtitlers' feedback was gathered in the focus group (Chapter 2) and then the first prototype based on that feedback was tested on a larger group of professional users in order to evaluate the usability of the subtitle editor (Chapter 4).

We believe that this approach could be applied to similar studies involving interactive media such as VR, AR or video games. Nonetheless, the UCT approach could also be applied to other types of research involving AVT translation modes and technology. For example, smartphone applications aiming to provide services such as SDH, AD, or sign language at cinemas, theaters or other events, voice recognition software used to generate subtitles, or new software for AVT practitioners and trainees, to name just a few.

8.2. Subtitling in immersive media: Setting the stage for further research

> This section is related to the following objective:
>
> 1.2. Describing the state of the art of immersive technology in regard to media applications.

Subtitling in immersive media appears to be an underresearched topic. To the best of our knowledge, this subject has not been addressed in TS or AVT Studies, hence its groundbreaking nature. Although TS and AVT Studies provided the general framework to start working on this study, the lack of previous related studies and sound literature within our field proved a hindrance. We had to turn to other fields such as engineering to find related studies. In fact, some of the key papers (Brown et al., 2018, Rothe et al., 2018) were published during the course of our study, proving the newness of the subject. Reviewing the literature and analyzing the state of the art helped us highlight the key elements that needed to be considered when tackling the topic of subtitling in immersive media.

Firstly, it was essential to understand immersive media and the technologies that enable this type of content. In Chapter 2, immersive technologies were reviewed, including fulldomes, 3D content, VR, AR, and mixed reality. It is important to be familiar with the different technologies, available devices and how users interact with each of them. Subtitling strategies are different for a 3D film in a movie theater with 3D glasses and for an immersive video game to be played on a PlayStationVR. In the former, the main goal is to integrate the subtitles with the image so that they can be properly read with 3D glasses and avoid causing eye strain. In the latter, the challenge lies in implementing subtitles in an interactive medium where the user will move, look around, and interact with the virtual space. Before carrying out a study related to immersive media and AVT, it is advisable to first study and understand the medium and experience the immersive content on commercially available devices. Then, the different challenges posed by the specific technology in relation to the AVT mode should be specified and be considered when developing the solutions.

We narrowed down the scope of the study and decided to focus on 360° videos that can be watched with an HMD. This type of content was discussed at length in Chapter 3. During the literature review, it was found that the most widespread term to refer to 360° videos is “cinematic virtual reality” or CVR. The different aspects of CVR were discussed, including technical aspects, average duration, narrative strategies and challenges, and the most common genres. Gaining a deep understanding of the product is important to later define and design subtitle solutions. For that reason, we considered that it was fitting to carry out a descriptive study and analyze a selected number of 360° videos, as explained in Chapter 3. The corpus analysis not only allowed us to better understand the nature of 360° content, but also to discover that text was already being integrated in the videos.

The corpus analysis shed some light regarding the integration of titles in 360° videos, which were mainly non-fiction. First, we found that titles were not used to make the videos accessible for foreign language speakers or viewers with hearing loss, but rather to contribute to the narrative of the videos. Regarding the function of titles in this type of videos, they were mostly used to indicate the video location or to include additional information for the viewers. The position of the titles was mainly evenly spaced every 120° or in a fixed position, and we did not find always-visible subtitles. Most titles were located at the top of the scene (86.5% for NYT and 0% for BBC), next to a speaker (34.4% for NYT and 13.5% for BBC) or next to an object or a person (19% for NYT and 32.4% for BBC), and in the middle of the image (25.5% for NYT and 83.8% for BBC). Surprisingly, titles were not mainly located at the bottom (14% for NYT and 5.4% for BBC), as is common practice in traditional screens. The preferred font color was white, and most videos did not include a black background box. The corpus analysis provided valuable insights regarding the possibilities of implementing subtitles in this type of content. Therefore, before conducting research on subtitling in immersive or any other emerging media, it is advisable to undertake an exploratory study with a view to gaining an overview of the type of content and the possibilities offered for subtitling.

Secondly, we defined the challenges posed by the medium in regard to subtitling. The comprehensive literature review and the analysis of current solutions, as well as user feedback, provided the necessary information to define a list of elements to be considered. According to related studies (Brown et al., 2018, Rothe et al., 2018) and our own conclusions, the main aspects to be considered when implementing subtitles in 360° videos are:

1. Position: Unlike traditional screens (TV, cinema, computer), 360° videos do not offer a static framed image, but rather it depends on the viewer to decide where to look within the 360° sphere. With that in mind, subtitles need to be defined in a way that they are available, accessible and readable for the viewer at all times.
2. Directions: The viewer has the freedom to look at any section of the 360° sphere. Therefore, any character speaking in the video may be out of sight at any given time. A system needs to be developed to show the viewers where the speaker is located in the 360° sphere.
3. Presence: The assumption is that 360° content delivers an immersive experience to viewers. The goal is, therefore, to design subtitles that do not disrupt that sense of presence. For example, we should avoid subtitles that are hard to read because the font is too small, or subtitles that are located in a position that is distracting for the viewers.

4. Avoiding VR sickness: Some users may experience VR sickness or dizziness when watching 360° content. Subtitles should be designed in a way that eases reading, rather than generating additional discomfort.

Finally, presence is a key concept in immersive media research. One of the main differences between traditional audiovisual media such as TV or cinema and immersive media such as 360° content is that the latter aims to provide a full immersive experience. In particular, immersive content strives to make viewers feel physically transported to the virtual world. Therefore, it is evident that the concept of presence needs to be integrated in any research related to AVT and immersive media. The most widespread tool for measuring presence is questionnaires. Choosing one of the validated presence questionnaires or another depends on how we understand presence. As explained in subsection 1.3.5, we construe presence as defined by Schubert et al. (2001). We considered that the authors' definition and the questionnaire that is based thereon (IPQ) were suitable for the product under research, namely 360° videos.

There is no standard process for measuring presence, and several possibilities can be found. Scholars interested in measuring presence in relation to AVT are encouraged to go through the literature that has been presented in this dissertation, and beyond if necessary, so as to decide which presence definition and measurement tools are more suitable for the immersive content being researched, given that a one-size-fits-all approach is not attainable. Nevertheless, it is undeniable that presence needs to be considered when studying immersive media. In relation to AVT studies, the main question is how the different AVT modes and strategies impact the user experience and, therefore, presence. The validity of presence questionnaires has received some criticism (Slater, 1999). Triangulating results with other methods is advisable in order to overcome the limitations posed by the usage of questionnaires. In our study, we also used questionnaires that included open questions, as well as a head movement tracking tool. The results of the IPQ questionnaire were triangulated with the other tools in order to better understand and interpret presence results. Eye-tracking, for example, could be used to scrutinize viewers' behavior in immersive media when using different translation modes. Other elements such as postural response (for example, a ball being thrown towards the viewers in the virtual world and them dodging in the real world) and physiological measures (such as heart rate and skin conductance response) have also been suggested as a way of objectively corroborating presence (Ijsselsteijn et al., 2000).

8.3. Study of subtitling technology as part of AVT Studies

> This section is related to the following objective:
>
> 2.2. Testing a 360° content subtitle editor prototype with professional subtitlers in order to gather their feedback in regard to the needs for subtitling 360° videos.

Translation technology has been thoroughly discussed and researched in TS over the last three decades. The keyword "machine translation" has been included in the BITRA database since its creation in 2001. That same year saw the birth of the academic e-journal *Revista Tradumàtica*, specializing in translation technologies. Other relevant translation journals such as *Perspectives: Studies in Translation Theory and Practice* (Oncins, 2015; Moorkens, 2017; Martín Mor, 2019; Sakamoto, 2019) or *The Journal of Specialised Translation* (Fulford & Granell Zafra, 2005; Garcia, 2005) have also been publishing articles related to translation technologies, to name just a few examples. More recently, a monographic exclusively dedicated to translation and technologies has been published (O'Hagan, 2019), evincing the importance of technology in our field of study.

However, studies focused on translation technology for AVT and subtitling in particular are scarce. Some general and descriptive papers on subtitling technology can be found (Díaz-Cintas, 2013, 2014; Georgakopoulou, 2012), as well as studies exploring the use of MT in the subtitling practice (Bywood et al., 2013; Bywood et al., 2017). Aside from this, studies focused on usability and computer-human interaction in relation to translation technologies are generally limited, with a few exceptions (O'Brien, 2012; Krüger, 2016; Witczak & Jaworski, 2018). Since technology is an intrinsic part of translation practice, it would make sense for more studies to be aimed at understanding how translators interact with commercially available tools and the impact of this interaction on translation, as well as revealing the usability of such tools from the translators' perspective.

We decided to carry out a usability test with a subtitling tool, as explained in Chapter 4, to aid in filling the current gap in AVT studies. In our study, we considered that usability was a key aspect of subtitling technology, as in any other software. Moreover, we were given the unique opportunity of contributing to the development of a subtitle editor for 360° videos from scratch in collaboration with engineers and developers. It was the perfect scenario to integrate a user-centered approach and involve end users from the beginning. In that sense, our contribution to the field of AVT has been manifold: 1) we have placed subtitling technology at the center of the study; 2) we have adopted a user-centered approach in the development and evaluation of subtitling technology; 3) we have integrated usability studies in the research of subtitling technology; and 4) we have highlighted the needs of subtitling technology for immersive content.

In our study, we investigated the features and characteristics that a subtitle editor for 360° videos requires. To do so, we firstly organized a focus group with professional subtitlers in order to gather their feedback, as explained in Chapter 2. After that, the development team of the ImAc project developed a first prototype integrating the feedback from the focus group. The first prototype was then tested with 27 professional subtitlers, as explained in Chapter 4. We conducted a widely adopted and validated usability questionnaire (namely, the SUS questionnaire) to measure the usability of the subtitle editor along with open questions to better understand the quantitative results drawn from the SUS questionnaire. The subtitling tool was then refined and improved with the feedback gathered and tested again in a second round of usability testing. Although the second test is not part of this dissertation[36], we can confirm that the tool received a better score in the SUS questionnaire after the improvements, which showed that having involved users from the outset was beneficial for the development of the tool.

This study shed some light on the needs that subtitling technology must meet for immersive content. Among other contributions, we found that:

- A new feature is necessary in order to indicate where the speaker is located in the 360° sphere. To help viewers navigate the virtual space, when a speaker is out of sight, a system needs to be put in place to redirect the viewers to the center of the action (for example, arrows or a radar). For that system to work, the subtitler needs to inform the tool where the speaker is. In this case, the developer created a functionality labelled "Set current angle" for that purpose. It was also important to consider the possibility of an off-camera speaker, for which no angle needs to be set. An option for off-camera speakers was also integrated in the tool.
- In 360° videos, shot change detection is not necessary because there are no shot changes *per se*, as it depends on where the viewer is looking.

[36] The results can be found on the ImAc Project website: https://www.imac-project.eu/documentation/deliverables/ (consulted on 03.24.2020).

- It is important for the subtitler to preview 360° videos to ensure that the subtitles are displayed correctly in the video – not just that the subtitles are well synchronized, but also that they are placed at the correct angle. In the subtitle editor, two types of previewing have been implemented: 1) forced preview: the video is previewed with the system forcing the selected angle for every subtitle; and 2) free preview: the subtitler can freely navigate the video as desired (as if they were the end viewer of the product). The first option is quicker and more straightforward, while the second option is intended for tasks where testing is required.
- For interlingual subtitles, it is also important to subtitle text shown on screen (for example, if a sign or a text message appear on the screen). Subtitlers will need to look for text on screen in the entire 360° sphere, which can be more time-consuming.

Technology is an essential part of AVT, especially of subtitling. Subtitling tools constrain and have a direct impact on the way professionals perform their job. It is, therefore, necessary to involve subtitling professionals in the development or evaluation of subtitling tools. The evaluation process is intended to improve the usability of the tools, but also to comprehend how subtitlers interact with the software and the cognitive processes that originate from that interaction. This could help researchers to better understand the process behind subtitling and how it can be improved with enhanced technology.

8.4. Evaluating subtitling strategies in immersive media

This section is related to the following goal:

3. To evaluate different subtitling strategies in immersive content (360° videos). Specifically:

- Comparing levels of presence and preferences among subtitle users (both hearing and with hearing loss) when the position of subtitles is either (a) always visible or (b) fixed position every 120°.
- Comparing levels of presence and preferences among subtitle users (both hearing and with hearing loss) when directions are transferred using either (a) arrows or (b) a radar.

One of the main goals of this dissertation was to evaluate different subtitling strategies in immersive media (360° content), comparing levels of presence and user preferences. After reviewing the literature, gathering feedback from end users via focus groups (Chapter 2), analyzing the multimedia corpus (Chapter 3), validating a subtitle editor for 360° videos (Chapter 4), and carrying out some additional usability testing (Annex 1), we decided to compare the following strategies (Chapter 5):

1. Position: always-visible subtitles versus fixed-position subtitles every 120°.
2. Directions: arrows versus radar.

These subtitling strategies were tested in a study with 40 participants: 20 hearing and 20 with hearing loss. The purpose of the test was to compare commercially available subtitling solutions (found on the NYT and BBC platforms, as explained in Chapter 3) with solutions developed within the ImAc project, which were based on user feedback (as explained in Chapter 2), in order to find out which solutions are more suitable according to end users. The duration of the four stimuli shown was longer than in previous studies (Rothe et al., 2018; Brown et al., 2018), with videos lasting around five minutes, thus overcoming one of the limitations highlighted in previous studies (Brown et al., 2018). The subtitles were developed including features generally associated with subtitles for persons with hearing loss, such as character identification or non-speech information, unlike previous studies. In this study, directions were tested for the first time. The measurement methods for testing were questionnaires (IPQ for presence and closed and open questions for preferences) and a head movement tracking tool to analyze user behavior while watching the stimuli.

We established two hypotheses in regard to this goal, and both were confirmed by the results.

H1: Users report higher levels of presence when viewing subtitles in always-visible mode compared to fixed-position mode. Always-visible mode is also their preferred option.

This hypothesis was made based on the idea that viewer preferences are frequently influenced by habit (Bartoll & Martínez-Tejerina, 2010). Always-visible subtitles behave similarly to those found in traditional media (that is, subtitles that are placed always in the same position, in front of the viewer, and at the bottom of the screen or the FoV). Habit also affects the usability of a product. If a product is intuitive, then it will be easier to use. Furthermore, the always-visible subtitles were designed following user feedback gathered in previous stages. The fixed-position subtitles, on the contrary, were based on current solutions found on commercially available platforms such as NYT or BBC 360° content.

The results of the preference questionnaire in this part of the test showed that 82.5 percent of participants preferred always-visible subtitles over fixed-position ones. Always-visible subtitles were considered easier to find and read and less distracting. However, fixed-position subtitles were considered slightly less obstructive (although the difference was not statistically significant). This might suggest that always-visible subtitles interfere slightly with the viewing experience, so it would be advisable to test different positions or different font sizes for this type of subtitles so as to keep researching and optimizing subtitles for immersive media. The presence questionnaire showed that always-visible subtitles performed slightly better than fixed-position subtitles. As regards overall presence (general presence item of the IPQ), the test indicated that the difference between the results was statistically significant, showing that fixed-position subtitles had a negative impact on the participants' sense of presence.

The responses to the open questions and the head tracking tool analysis reinforced and clarified the abovementioned results. With always-visible subtitles, participants agreed that they felt more freedom to look around the scene without missing the subtitles, as was then corroborated by the head tracking analysis. Fixed-position subtitles were sometimes frustrating for participants because they missed parts of the subtitles if they wanted to further explore the video. As was subsequently observed in the head tracking analysis, participants tended to be more static with fixed-position subtitles.

H2: Users report higher levels of presence when visualizing directions with arrows compared to radar. Arrows are the preferred option.

We believed that arrows would be the preferred option because they are less intrusive to the viewing experience. Arrows only appear when the speaker is out of sight, while the radar is always present in the FoV. The idea behind the radar was to provide accurate spatial information to viewers, allowing them to be aware of their position in relation to the video at all times. However, as the results showed, understanding the radar proved to be challenging for participants.

In this case, 82.5 percent of participants preferred arrows over radar. Arrows were considered more intuitive and direct compared to radar, which required a higher effort to process the given information. Participants who preferred radar suggested some improvements to make it more usable. For example, including a short explanation before the video started or locating the radar at the center and directly below the subtitle, instead of on the right-hand side. Four participants stated that they would prefer to watch the video without directions. Arrows were considered to be an easier method to find the speaker as well as being less distracting compared to radar. The IPQ questionnaire showed that arrows performed better on the involvement scale, which means that radar had a negative impact on the involvement of participants. The replies to the open questions showed that this could be caused by the fact that participants struggled to understand how the radar worked and, therefore, were paying less attention to the content of the video. The head tracking analysis did not show any significant differences between the two methods.

We can, therefore, conclude that for directions in 360° videos, the desirable solution should be simple, direct and intuitive. At the same time, the solution should not interfere with the immersive experience, requiring additional effort for viewers who are already overexposed to the input of the 360° video. According to the results of the reception study, arrows meet the abovementioned requirements.

8.5. An updated concept of subtitling: moving away from the distinction between subtitling and SDH

When approaching the topic of this dissertation, we encountered some hindrances that made us rethink the whole concept of subtitling. We started using the terms subtitling and SDH as if they were two completely different concepts, but as we moved forward in the study, we realized that these two terms became interchangeable. The reason for this progression was that we were creating a new type of subtitles from scratch for a new medium and, therefore, it became necessary to include all end users of subtitles, which means both hearing persons and persons with hearing loss. At the same time, this new medium introduced new dimensions that did not fit in the previous formal definitions of subtitling or SDH (for example, new elements such as directions were necessary). Therefore, we needed to consider not just the textual dimension of subtitles, but also the way they are presented in immersive media.

We then tackled the concept of subtitling from a user-centered perspective and considered the opportunities brought about by new technologies. The heterogeneous nature of the considered target audience of SDH had already come to light in the past

(Neves, 2008; Arnáiz-Uzquiza, 2012). Persons with hearing loss may approach subtitles in a different way, which is not necessarily based on their disability, but rather on their langue proficiency and preferences. Therefore, it became clear that the segregation between these two artificial groups of subtitling users (hearing and with hearing loss) was no longer valid for the purpose of our study. Consequently, we decided to propose a new definition for subtitles, as indicated in subsection 1.3.1:

> 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 and extralinguistic aspects of the content for its understanding) that are legible, readable, comprehensible and accessible for the intended users.

This definition avoids any reference to disabilities or to the purely formal aspects of subtitling. It was built based on an inclusive user-centered approach. Developing new subtitle solutions for new media called for a user-centered approach involving end users from a very early stage in the study. This specific setup together with the new technology forced us to reconsider the way we understand subtitling and viewers. We were unsure as to whether we were creating subtitles only for persons with hearing loss or for any type of viewer. The latter was considered more appropriate and accurate because hearing viewers are also subtitle consumers. Therefore, there was no reason to think that hearing participants could not contribute to the development of a product that was ultimately also intended for them. The type of subtitles used is not necessarily based on viewers' disabilities, but rather on the preferences and needs of each user, which may or may not be related to their disabilities. For example, the user's proficiency in relation to written language may have an impact on how they assess subtitles.

Some scholars have argued that persons with hearing loss read slower than hearing persons (Neves, 2008). However, this generalization is not completely true because it depends on the viewers' language proficiency rather than on their capacity to hear. For example, hearing participants who are not familiar with subtitling may complain about subtitles being too fast (as occurred in the reception study we carried out). And some persons with complete hearing loss may be perfectly proficient in written language and, therefore, their reading speed would be the same as that of a hearing person. In short, the type of subtitles should not depend on the disability of the person, but rather on their capabilities to, for example, read quickly or understand the written language, or even on their preferences or habits. We believe that this approach is more inclusive and reflects the current reality more accurately.

This reasoning led us to include both hearing viewers and viewers with hearing loss in the study. The results showed two facts that are worth mentioning. Firstly, we found that more hearing participants (15) watched subtitled content on a daily basis compared to participants with hearing loss (12). Also, participants who watched subtitled content every day were younger (M=34.2) and used more technology (namely, laptops) than participants who did not watch subtitled content. This shows that subtitling consumption is not directly related to the capability of hearing or not, but rather to the users' preferences, or even age and technology usage habits. Secondly, it is worth noting that hearing participants did not complain about subtitles including non-speech or extralinguistic information such as character identification (colors) or sound representation.

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

In this last chapter, we have explained the manifold contributions of this dissertation. On one hand, the theoretical implications that have been discussed include applying a user-centered approach to studies in the field of AVT, laying the theoretical foundations for subtitling research in immersive media, including technology reception studies in the framework of AVT Studies, and proposing an updated definition of subtitling and subtitling viewers. On the other hand, the practical implications have mainly focused on subtitles in 360° videos. We have covered different angles from which subtitling can be studied: we have gathered feedback from subtitling users and professional subtitlers; we have analyzed the current commercially available subtitling solutions provided by the NYC and the BBC; we have tested a prototype of a subtitle editor for 360° videos; and we have tested different subtitle solutions in 360° content with subtitling consumers. Our reception study has been the largest so far concerning this topic (40 participants). It is the first study to include hearing participants and a large number of participants with hearing loss. The stimuli used were the longest so far in this type of studies, which meant that the sense of presence and viewer preferences could be measured more accurately. Finally, we have presented preliminary recommendations specifying how viewers prefer to see subtitles in 360° videos (i.e., always visible with arrows). We have also presented the features that a subtitle editor for immersive content would require.

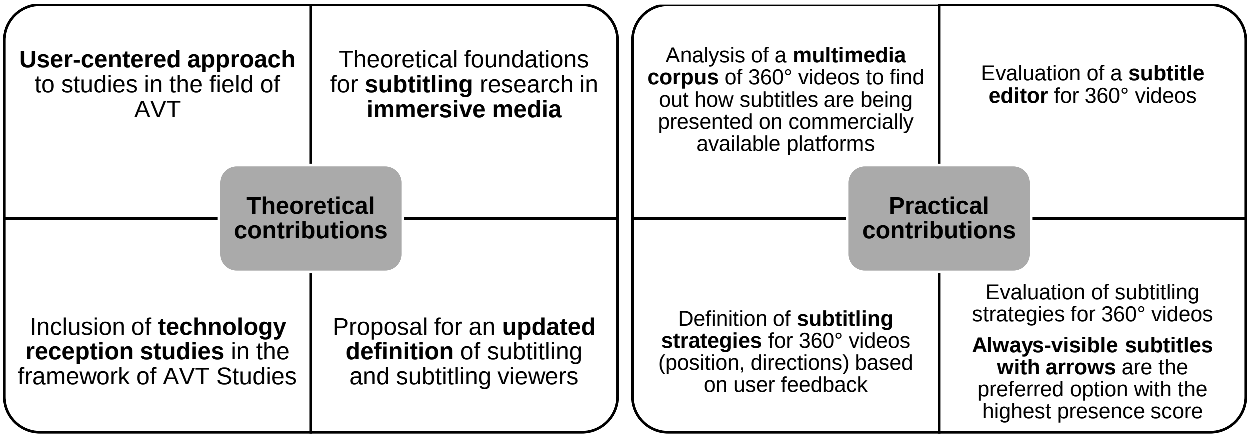


Figure 21. Summary of theoretical and practical contributions of this dissertation.

The study also had its limitations. Firstly, due to the newness of this technology, we were faced with several technical constraints that could not be overcome at the time of the study. For example, video quality could be improved using 4K resolution, so the videos could have delivered a more immersive experience contributing to higher levels of presence. The always-visible subtitles shown had a slight flickering effect due to the limitations of the HMD that was used in the test. Perhaps, if the quality of the videos and the technology were optimal, the results would have more clearly favored always-visible subtitles. In any case, the consensus among users regarding their preferred option was strong enough to consider the always-visible solution with arrows to be the best option for this medium for the time being. Similarly, the newness of the technology may have had an impact on the reaction of participants. For most of them, it was their first time using VR technology. It may be that some of them felt highly present because they were impressed by the technology, while others did not feel so present because they were uncomfortable with the HMD. It would be advisable to carry out related studies involving participants who have more experience with VR to see if the levels of presence are similar or if the level of VR expertise has an impact on the results.

Secondly, the type of content could also be considered a limitation. For the main reception study, we used two self-contained documentary clips and two not self-contained science fiction clips. The travel documentary type of content may have favored always-visible subtitles because viewers wanted to explore the scenes. Perhaps if the chosen video were a more static setup, they would have favored fixed-position subtitles. Replicating the test with different types of content would be advisable to further explore the possibilities. For the science fiction content, we had to split the clip in two parts and, therefore, the content was not self-contained. In order to test immersion, self-contained content is ideal (Orero et al., 2018). However, we did not have any other stimuli at our disposal that could be used at the time of the study. The *I, Philip* movie was considered suitable for many other reasons. For future research, it would, therefore, be advisable to use self-contained videos. Finally, we believe that eye-tracking technology could have contributed to a better understanding of user behavior in relation to subtitling in 360° videos. At the time of the study, we did not have eye-tracking technology for HMD at our disposal. We implemented head tracking software to solve this problem, which at least gave us an idea of participants' movement patterns. However, we strongly recommend carrying out similar tests with eye-tracking technology.

We believe this dissertation helps pave the way for future studies regarding subtitling in immersive environments, and we encourage researchers to further test different aspects of subtitles in immersive content. For example, the number of characters per line and reading speed. Current subtitling standards on characters per line and reading speed should be further tested in immersive media to corroborate whether they are still valid. Our current hypothesis is that viewers would prefer shorter and slower subtitles for this type of media, but that assumption needs to be validated. Testing different font sizes and locations is also necessary. During the reception study, participants noted that always-visible subtitles were slightly more obstructive than fixed-position subtitles. Perhaps, subtitles located at the top of the FoV would be less intrusive. A study comparing subtitles at the bottom and subtitles at the top could clarify this question. Subtitle position in immersive content should, therefore, be further researched. Although always-visible subtitles have been chosen as the preferred option, different integrations are possible depending on the type of content. Creative approaches could also be implemented and tested in this medium. Again, it is important to test subtitles with different types of content – featuring more or less action, for instance – to see which options work better depending on genre. It would be interesting to carry out reception studies with subtitles in more interactive VR setups, such as video games.

Regarding directions, there is still room for improvement. For example, an improved version of the radar could be developed following user feedback (locating the radar at the center, below the subtitle, and including a short explanation before the video starts). Maybe other methods could be developed to further explore user preferences and behavior in immersive environments. Moreover, it could be interesting to use directions not only for speakers but also to indicate the main action in the scene. For example, the scene might not focus on a speaker at a certain moment, but on someone dancing. Arrows or radar could be used to tell viewers where the main action is taking place.

Regarding subtitling studies in general, we encourage AVT researchers to take a user-centered approach and carry out similar studies following the UCT methodology. We are referring, for instance, to subtitling studies in different audiovisual media where innovation is needed, such as video games, VR and AR, vertical videos on smartphones, or smartphone apps for subtitling at live events such as cinema or theaters. Involving users from a very early stage of the product development and carrying out usability testing can help to generate a higher-quality and more adequate final subtitling solution. Following the results from the reception study, it would be interesting to test traditionally SDH features such as character identification or non-speech information on hearing viewers from a user-centered perspective in order to explore whether these attributes enhance the usability of the subtitles.

The importance of carrying out studies in relation to subtitling technology has been made clear. It is crucial to understand how subtitling technologies impact the work of professional subtitlers, the cognitive processes behind their interaction with the software, or the process of subtitling with a specific tool. AVT scholars should be more involved in the development of these tools and proactive in analyzing how technology constrains or enhances a subtitler's work. Research methods such as eye-tracking, keylogging or screen recording could be used in addition to focus groups or usability questionnaires.

Finally, the AVT scholars' community should reconsider the way we understand subtitling and accessibility, and further discuss this topic to provide theories that will drive our field of research forward. Only by questioning ourselves and discussing relevant topics, as well as carrying out reception studies with end users, can we have a wider impact on society.

Updated bibliography

Updated bibliography

This is a unified bibliography for the dissertation, following referencing rules from APA 6th Edition. When necessary, references included in the articles have been updated here.

Agulló, B. (2018). SDH in immersive environments: can subtitles be immersive at all? Paper presented in Languages & The Media 2018. Retrieved from https://ddd.uab.cat/record/196360

Agulló, B. (in press). Technology for subtitling: a 360-degree turn. *Hermeneus*, *22*.

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

Agulló, B., Matamala, A., & 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., & 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., 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., & Matamala, A. (in press). (Sub)titles in cinematic virtual reality: a descriptive study. *Onomazéin*.

Álvarez, A., del Pozo, A., & Arruti, A. (2010). APyCA: towards the automatic subtitling of television content in Spanish. In M. Ganzha & M. Paprzycki (Eds.), *Proceedings of the International Multiconference on Computer Science and Information Technology, Wisla, Poland*, 567-574. Piscataway, NJ: IEEE. https://doi.org/10.1109/IMCSIT.2010.5680055

Antonini, R. (2005). The perception of subtitled humor in Italy. *Humor: International Journal of Humor Research, 18*(2), 209-225.

Arias Badia, B. (2017). *Television dialogue and subtitling: a corpus-driven study of police procedurals* (Doctoral dissertation). Retrieved from E-Repositori of Universitat Pompeu Fabra https://repositori.upf.edu/handle/10230/32615

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

Arnáiz-Uzquiza, V. (2015). Eye tracking in Spain. In P. Romero-Fresco (ed.), *The reception of subtitles for the deaf and hard of hearing in Europe* (263-287). Bern: Peter Lang.

Arrés, E. (2017). Traducir en entornos de realidad virtual: cómo garantizar la inmersión del usuario. In A. Bécart, V. Merola, & R. López-Campos Bodineau (Eds.), *New*

Technologies Applied to Translation Studies: strategies, tools and resources (217-134). Sevilla: Editorial Bienza.

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*.

Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355-385. https://doi.org/10.1162/pres.1997.6.4.355

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, *21*(6), 34-47. https://doi.org/10.1109/38.963459

Báez Montero, I. C., & Fernández Soneira, A. M. (2010). Spanish deaf people as recipients of closed captioning. In A. Matamala, & P. Orero (Eds.), *Listening to subtitles: Subtitles for the deaf and hard of hearing* (25-44). Bern: Peter Lang.

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* (69-86). Bern: Peter Lang.

Belton, J. (2012). Digital 3D cinema: digital cinema's missing novelty phase. *Film History: An International Journal*, *24*(2), 187-195. Retrieved from https://muse.jhu.edu/article/479267

Biocca, F. (1992). Virtual reality technology: a tutorial. *Journal of Communication 42*(4), 23-72. https://doi.org/10.1111/j.1460-2466.1992.tb00811.x

British Broadcasting Corporation (BBC) (2012). *Binaural sound: Immersive spatial audio for headphones*. Retrieved from http://www.bbc.co.uk/rd/projects/binaural-broadcasting

British Broadcasting Corporation (BBC) (2014). *360 video and virtual reality: Investigating and developing 360-degree video and VR for broadcast-related applications*. Retrieved from http://www.bbc.co.uk/rd/projects/360-video-virtual-reality

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P.W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (189-194). London: Taylor and Francis. https://doi.org/10.1201/9781498710411

Brooke, J. (2013). SUS: a retrospective. *Journal of Usability Studies*, *8*(2), 29-40. Retrieved from https://uxpajournal.org/sus-a-retrospective/

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2017). Subtitles in 360-degree video. In J. Redi, O. Niamut, & D. Bulterman (Eds.), *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video, Hilversum, Netherlands*, 3-8. New York: ACM. 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). Retrieved from https://www.bbc.co.uk/rd/publications/whitepaper330

Bruti, S. & Zanotti, S. (2018). Representations of stuttering in subtitling: a view from a corpus of English language films. In I. Ranzato & S. Zanotti (Eds.), *Linguistic and*

Cultural Representation in Audiovisual Translation (228-262). London: Routledge.

Bryman, A. (2004). *Social Research Methods*. New York: Oxford University Press.

Bywood, L., Volk, M., Fishel, M., & Georgakopoulou, P. (2013). Parallel subtitle corpora and their applications in machine translation and translatology. *Perspectives: Studies in Translation Theory and Practice*, *21*(4), 595-610. https://doi.org/10.1080/0907676X.2013.831920

Bywood, L., Georgakopoulou P., & Etchegoyhen, T. (2017). Embracing the threat: machine translation as a solution for subtitling. *Perspectives: Studies in Translation Theory and Practice*, *25*(3), 492-508. http://dx.doi.org/10.1080/0907676X.2017.1291695

Caffrey, C. (2009). *Relevant abuse? Investigating the effects of an abusive subtitling procedure on the perception of TV anime using eye tracker and questionnaire* (Doctoral dissertation). Retrieved from DCU Online Research Access Service (Doras) http://doras.dcu.ie/14835/

Cambra, C., Silvestre, N., & Leal, A. (2015). How useful are television subtitles in helping deaf children to interpret cartoon programmes? In J. Díaz-Cintas & J. Neves (Eds.), *Audiovisual Translation: Taking Stock* (245-260). Newcastle upon Tyne: Cambridge Scholars Publishing.

Catapult Digital (2018*). Growing VR/AR companies in the UK: A business and legal handbook*. Retrieved from https://www.pwc.co.uk/intelligent-digital/vr/growing-vr-ar-companies-in-the-uk.pdf

Cavaliere, F. (2008). Measuring the perception of the screen translation of Un posto al sole: A cross-cultural study. In D. Chiaro, C. Heiss, & C. Bucaria (Eds.), *Between text and image: Updating research in screen translation* (165-180). Amsterdam and Philadelphia: John Benjamins.

Chaume, F. (2013). *Audiovisual translation. Dubbing*. Manchester: St. Jerome Publishing.

Chen, Y. L. (2016). The effects of virtual reality learning environment on student cognitive and linguistic development. *The Asia-Pacific Education Researcher*. *25*(4), 637-646. https://doi.org/10.1007/s40299-016-0293-2

Chong, H. & Wang, K. (2014). Subtitling swearwords in reality TV series from English into Chinese: A corpus-based study of The Family. *Translation & Interpreting*, *6*(2), 1-17. Retrieved from http://www.trans-int.org/index.php/transint/article/view/271

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* (149-162). Bern: Peter Lang.

Conroy, A. (2017, March 14). The BBC and Virtual Reality [Web log post]. Retrieved from http://www.bbc.co.uk/rd/blog/2016-06-the-bbc-and-virtual-reality

D'Ydewalle, G., Pollet, J., & van Rensbergen, J. (1987). Reading a message when the same message is available auditorily in another language: the case of subtitling. In J.K. O'Regan, & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition*, (313-321). North Holland: Elsevier Science Publishers.

D'Ydewalle, G., Praet, C., Verfaille, K., & Van Rensbergen, J. (1991). Watching subtitled television: automatic reading behavior. *Communication Research, 18*, 650-666.

De la Peña, N., Weil, P., Llobera, J., Giannopoulos, E., Pomés, A., Spanlang, B., Friedman, D., Sanchez-Vives, M. V., & Slater, M. (2010). Immersive journalism: immersive virtual reality for the first-person experience of news. *Presence: Teleoperators and Virtual Environments, 19*(4), 291-301. https://doi.org/10.1162/PRES_a_00005

De Linde, Z., & Kay, N. (1999). *The semiotics of subtitling*. Manchester: St. Jerome Publishing.

De Meao, M. (2012) Subtitling dialects: strategies of socio-cultural transfer from Italian into English. In S. Bruti & E. Di Giovanni (Eds.), *Audiovisual Translation across Europe: An Ever-Changing Landscape* (79-96). Bern: Peter Lang.

Deloitte (2016). Technology, Media & Communications Predictions. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Technology-Media-Telecommunications/gx-tmt-prediction-2016-full-report.pdf

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

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

Díaz-Cintas, J. (2013). The technology turn in subtitling. In M. Thelen, & B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: Part 9* (119-132). Maastricht: Zuyd University of Applied Sciences.

Díaz-Cintas, J. (2014). Technological strides in subtitling. In C. Sin-wai (ed.), *The Routledge encyclopedia of translation technology*, (632-643). Abingdon: Routledge.

Díaz-Cintas, J., Orero, P., & Remael, A. (Eds.) (2007). *Media for all: subtitling for the deaf, audio description, and sign language*. Amsterdam: Rodopi.

Díaz-Cintas, J., & Remael, A. (2007). *Audiovisual Translation. Subtitling*. London: Routledge.

Di Giovanni, E., & Gambier, Y. (Eds.) (2018). *Reception Studies and Audiovisual Translation*. Amsterdam and Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/btl.141

Dooley, K. (2017). Storytelling with virtual reality in 360-degrees: a new screen grammar. *Studies in Australasian Cinema, 11*(3), 161-171. https://doi.org/10.1080/17503175.2017.1387357

Dooley, K. (2018). Virtual reality: moving beyond a cinema of attractions. *Screen Education, 88*, 97-103.

Dranch, K. (2018). Translation tech goes media: Subtitling and dubbing won't stay uncontested for long [Web log post]. Retrieved from https://www.nimdzi.com/translation-tech-goes-media/

Eardley-Weaver, S. (2015). Opera (sur)titles for the deaf and hard-of-hearing. In J. Díaz-Cintas & J. Neves (Eds.), *Audiovisual translation. Taking stock* (261-276). Newcastle upon Tyne: Cambridge Scholars Publishing.

Estopace, E. (2017, November 23), Audiovisual translation hits a sweet spot as subscription video on-demand skyrockets [Web log post]. Retrieved from https://slator.com/features/audiovisual-translation-hits-sweet-spot-subscription-video-on-demand-skyrockets/

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

Ferrari, A. (2016, December 15). IFH 123: What is cinematic virtual reality with Jason Diamond. *Indie Film Hustle*. Podcast retrieved from https://indiefilmhustle.com/cinematic-virtual-reality/

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* (81-98). Amsterdam: Rodopi.

Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces*, *6*(2), 291-309. https://doi.org/10.1075/ts.6.2.06for

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

Fox, W. (2016b). "Should she really be covered by her own subtitle?" Text elements in film and their graphical translation. *Translation Spaces*, *5*(2), 244-270. https://doi.org/10.1075/ts.5.2.05fox

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*. Berlin: Language Science Press. Retrieved from https://langsci-press.org/catalog/book/187

Fraile, I., Gómez, D., Núñez, J. A., Montagud, M., & Fernández, S. (2018). Personalized and immersive presentation of video, audio and subtitles in 360º environments: an opera use case. *ACM TVX 2018*. Retrieved from https://figshare.com/articles/Demo_Personalized_and_Immersive_Presentation_of_Video_Audio_and_Subtitles_in_360o_Environments_An_Opera_Use_Case/6527531

Franco, E., Matamala, A., & Orero, P. (2010). *Voice-over translation: An overview*. Berna: Peter Lang.

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, *47*(14), 2393-2400. https://doi.org/10.1017/S003329171700040X

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

Fulford, H., & Granell Zafra, J. (2005). Translation and technology: a study of UK freelance translators. *The Journal of Specialised Translation*, *4*, 2-17. Retrieved from http://www.jostrans.org/issue04/art_fulford_zafra.php

Gambier Y. (2003) Screen transadaptation: perception and reception. *The Translator: Screen Translation, 9*(2), 17-189.

Garcia, I. (2005). Long term memories: Trados and TM turn 20. *The Journal of Specialised Translation*, *4*, 18-31. Retrieved from http://www.jostrans.org/issue04/art_garcia.php

Garcia, J. E., Ortega, A., Lleida, E., Lozano, T., Bernues, E., & Sanchez, D. (2009). Audio and text synchronization for TV news subtitling based on Automatic Speech Recognition. *2009 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, Bilbao, Spain*, 1-6. Piscataway, NJ: IEEE. http://doi.org/10.1109/ISBMSB.2009.5133758

Georgakopoulou, P. (2012). Challenges for the audiovisual industry in the digital age: the everchanging needs of subtitle production. *The Journal of Specialised Translation*, *17*, 78-103. Retrieved from https://www.jostrans.org/issue17/art_georgakopoulou.pdf

Gerber-Morón, O., Szarkowska, A. & Woll, B. (2018). The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research*, *11*(4), 1-18. https://doi.org/10.16910/11.4.2

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gödde, M., Gabler, F., Siegmund, D., & Braun, A. (2018). Cinematic narration in VR – Rethinking film conventions for 360 degrees. In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry. VAMR 2018. Lecture Notes in Computer Science, vol 10910* (184-201). Cham: Springer. http://doi.org/10.1007/978-3-319-91584-5_15

González-Zúñiga, D., Carrabina, J., & Orero, P. (2013). Evaluation of depth cues in 3D dubtitling. *Online Journal of Art and Design*, *1*(3), 16-29. Retrieved from http://www.adjournal.net/articles/13/132.pdf

Goodman, E., Kuniavsky, M., & Moed, A. (2012). *Observing the User Experience: A Practitioner's Guide to User Research*. Waltham: Morgan Kaufmann.

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. In Y. Gambier, E. Salo, E. Satopää, & R. Haeseryn (Eds.), *Communication audiovisuelle et transferts linguistiques: Translatio – Nouvelles de la FIT*, *14*(3-4), 388-409.

Grambart, S. (2015, December 10). *Sleepy Hollow & Narrative in VR*. Retrieved from https://www.slideshare.net/fitc_slideshare/sleepy-hollow-narrative-in-vr

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* (11-33). London: Palgrave Macmillan. http://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, vol 11786, 15-27. Cham: Springer. http://doi.org/10.1007/978-3-030-30033-3_2

Guillot, M.-N. (2008) Orality and film subtitling: the riches of punctuation. *The Sign Language Translator and Interpreter, 2*(2), 127-147.

Guillot, M. N. (2018). Subtitling on the cusp of its future. In L. Pérez-González (ed.), *The Routledge Handbook of Audiovisual Translation* (31-47). London: Routledge.

Hall, C. (2017, January 24). Watch VR's first Oscar-nominated short film in Polygon [Web log post]. Retrieved from https://www.polygon.com/2017/1/24/14370892/virtual-reality-first-oscar-nominated-short-film-pearl

Haluck, R. S., & Krummel, T. M. (2000). Computers and virtual reality for surgical education in the 21st century. *Archives of Surgery*, *135* (7), 786-792.

Harris, A., Lyons, M., & Ryan, M. (2019). 'Bandersnatch' has many paths, but do any of them add up to anything? *The New York Times*. Retrieved from https://www.nytimes.com/2019/01/04/arts/television/bandersnatch-black-mirror-netflix.html

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, *1*(2), 262-271. http://doi.org/10.1162/pres.1992.1.2.262

Holliman, N. S., Dodgson, N. A., Favalora, G. E., & Pockett, L. (2011). Three-dimensional displays: a review and applications analysis. *IEEE Trans. Broadcast*, *57*(2), 362-371. http://doi.org/10.1109/TBC.2011.2130930

Hu, Y., Kautz, J., Yu, Y., and Wang, W. (2015). Speaker-following video subtitles. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *11*(2), 1-17. http://doi.org/10.1145/2632111

Hughes, C. J., Armstrong, M., Jones, R., and Crabb, M. (2015). Responsive design for personalised subtitles. In L. Carriço & S. Mirri (Eds.), *W4A '15 Proceedings of the 12th Web for All Conference*, 1-4. New York: ACM. https://doi.org/10.1145/2745555.2746650

Ijsselsteijn, W.A., de Ridder, H., Freeman, J., & Avons, S.E. (2000). Presence: concept, determinants, and measurement. *Human Vision and Electronic Imaging V, 39*, 59-76. https://doi.org/10.1117/12.387188

Jakobson, R. (1959). On linguistic aspects of translation. In V. Lawrence (ed.), *The translation studies reader* (113-118). London & New York: Routledge.

Jayaraj, L., Wood, J., & Gibson, M. (2017). Improving the immersion in virtual reality with real-time avatar and haptic feedback in a cricket simulation. In W. Broll, H. Regenbrecht, J. E. Swan II, G. Bruder, M. Servières, & M. Sugimoto (Eds.), *Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, Nantes, France*, 310-314. http://doi.org/doi:10.1109/ISMAR-Adjunct.2017.95

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

Jensema, C. (1998). Viewer reaction to different television captioning speeds. *American Annals of the Deaf*, *143*(4), 318-324.

Jiménez Hurtado, C., Seibel, C., & Soler Gallego, S. (2012). Museos para todos. La traducción e interpretación para entornos multimodales como herramientas de accesibilidad universal. *MonTI. Monografies de Traducció i Interpretació*, *4*, 349-383. https://doi.org/10.6035/MonTI.2012.4.15

Jones, S. (2017). Disrupting the narrative: immersive journalism in virtual reality. *Journal of Media Practice*, *18*(2–3), 171-185. http://doi.org/10.1080/14682753.2017.1374677

Jung, T., tom Dieck, M.C., Lee, H., & Chung, N. (2016). Effects of virtual reality and augmented reality on visitor experiences in museum. In A. Inversini & R. Schegg (Eds.), *Information and Communication Technologies in Tourism* (621-635). Cham: Springer. https://doi.org/10.1007/978-3-319-28231-2_45

Kalantzi, D. (2008). *Subtitling for the deaf and hard of hearing: a corpus-based methodology for the analysis of subtitles with a focus on segmentation and deletion* (Unpublished doctoral dissertation). The University of Manchester, Manchester.

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

Kruger, J. L.(2012). Making meaning in AVT: eye tracking and viewer construction of narrative. *Perspectives: Studies in Translation Theory and Practice*, *20*(1), 67-86. http://doi.org/10.1080/0907676X.2011.632688

Krüger, R. (2016). Contextualising computer-assisted translation tools and modelling their usability. *Trans-Kom-Journal of Translation and Technical Communication Research*, 9(1), 114-148. Retrieved from http://www.trans-kom.eu/bd09nr01/trans-kom_09_01_08_Krueger_CAT.20160705.pdf

Kurzhals, K., Cetinkaya, E., Hu, Y., Wang, W., & Weiskopf, D. (2017). Close to the action: eye-tracking evaluation of speaker-following subtitles. In G. Mark & S. Fussell (Eds.), *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, USA*, 6559-6568. New York: ACM. http://doi.org/10.1145/3025453.3025772

Lambooij, M., Murdoch, M. J., Ijsselsteijn, W. A., & Heynderickx, I. (2013). The impact of video characteristics and subtitles on visual comfort of 3D TV. *Displays*, *34*(1), 8-16. http://doi.org/10.1016/j.displa.2012.09.002

Lee, K. M. (2004). Presence, explicated. *Communication theory*, 14(1), 27-50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lee, D. G., Fels, D. I., & Udo, J. P. (2007). Emotive captioning. *ACM Computers in Entertainment*, *5*(2), 1-15. http://doi.org/10.1145/1279540.1279551

Lee, N. (2017, November 15). Pixar's VR debut takes you inside the entrancing world of 'Coco' [Web log post]. Retrieved from https://www.engadget.com/2017/11/15/pixar-coco-vr/

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators, and Virtual Environments*, *10*(3), 282-297. http://doi.org/10.1162/105474601300343612

Lewis, J. R. (2018). The System Usability Scale: past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577-590. http://doi.org/10.1080/10447318.2018.1455307

Lombard, M., & Ditton, T. B. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2), n. p. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Longo, A. (2009) Subtitling the Italian South. In J. Díaz Cintas (ed.), *New Trends in Audiovisual Translation* (99-108). Clevedon: Multilingual Matters.

MacQuarrie, A., & Steed, A. (2017). Cinematic virtual reality: evaluating the effect of display type on the viewing experience for panoramic video. In E. S. Rosenberg et al. (Eds.), *Proceedings of 2017 IEEE Virtual Reality (VR), USA*, 45-54. Piscataway, NJ: IEEE. http://doi.org/10.1109/VR.2017.7892230

Mangiron, C. (2013). Subtitling in game localisation: a descriptive study. *Perspectives: Studies in Translation Theory and Practice*, *21*(1), 42-56. http://doi.org/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

Martín Mor, A. (2019). Do translation memories affect translations? Final results of the TRACE project. *Perspectives: Studies in Translation Theory and Practice*, *27*(3), 455-476. https://doi.org/10.1080/0907676X.2018.1459753

Maszerowska, A., Matamala, A., & Orero, P. (2014). *Audio description. New perspectives illustrated*. Amsterdam and Philadelphia: John Benjamins.

Matamala, A. (2019). Accessibilitat i traducció audiovisual. Barcelona: Eumo Editorial.

Matamala, A., & Orero, P. (Eds.) (2010). *Listening to subtitles: Subtitles for the deaf and hard of hearing*. Bern: Peter Lang.

Matamala, A., Moreno, A. O., Álvarez Muniain, A., & Azpeitia Zaldua, A. (2015). The reception of intralingual and interlingual subtitling: an exploratory study within the HBB4ALL project. *Proceedings of the 37th Conference Translating and the Computer*, 12-17. Retrieved from https://ddd.uab.cat/record/144868

Matamala, A., Orero, P., Rovira-Esteva, S., Casas-Tost, H., Morales, L. F., Soler-Vilageliu, O., Agulló, B., Fidyka, A., Segura, D., & Tor-Carroggio, I (2018). User-centric approaches in Access Services Evaluation: Profiling the End User. *Proceedings of LREC Workshop "Improving Social Inclusion using NLP: Tools, Methods and Resources", ISINLP2 2018*. Retrieved from https://ddd.uab.cat/record/189707

Mateer, J. (2017). Directing for cinematic virtual reality: how the traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice*, *18*(1), 14-25. http://doi.org/10.1080/14682753.2017.1305838

Mathur, A., Saxena, T., & Krishnamurthi, R. (2015). Generating subtitles automatically using audio extraction and speech recognition. *2015 IEEE International Conference on Computational Intelligence & Communication Technology, Ghaziabad*, 621-626. Piscataway, NJ: IEEE. http://doi.org/10.1109/CICT.2015.46

Mattsson, J. (2009). *The subtitling of discourse particles. A corpus-based study of well, you know, I mean, and like, and their Swedish translations in ten American films*. (Doctoral Dissertation). Retrieved from Göteborgs Universitetbiblitek https://gupea.ub.gu.se/handle/2077/21007

McGreevy, M.W. (1993) The presence of field geologists in a Mars-like terrain. *Presence: Teleoperators and Virtual Environments*, *1*(4), 375-403. https://doi.org/10.1162/pres.1992.1.4.375

McClarty, R. (2012). Towards a multidisciplinary approach in creative subtitling. *MonTI. Monografías de Traducción e Interpretación*, *4*, 133-153. http://doi.org/10.6035/MonTI.2012.4.6

McClarty, R. (2014). In support of creative subtitling: contemporary context and theoretical framework. *Perspectives: Studies in Translation Theory and Practice*, *22*(4), 592-606. http://doi.org/10.1080/0907676X.2013.842258

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems*, *77*, 1321-1329.

Montagud, M., Fraile, I., Nuñez, J. A., & Fernández, S. (2018). ImAc: enabling immersive, accessible and personalized media experiences. *ACM TVX 2018*. Retrieved from https://dl.acm.org/citation.cfm?id=3213570

Moorkens, J. (2017). Under pressure: translation in times of austerity. *Perspectives: Studies in Translation Theory and Practice*, *25*(3), 464-477. https://doi.org/10.1080/0907676X.2017.1285331

Morales, M., Koch, T., & Beauregard, L. P. (2019). Netflix removes Castilian Spanish subtitles from Alfonso Cuarón's 'Roma'. *El País*. Retrieved from https://elpais.com/elpais/2019/01/10/inenglish/1547125380_758985.html

Morgan, D. L. (1998). *Planning Focus Groups*. Thousand Oaks: Sage.

Neves, J. (2005). *Audiovisual translation: subtitling for the deaf and hard of hearing*. (Doctoral dissertation). Retrieved from IC-Online http://hdl.handle.net/10400.8/409

Neves, J. (2008). 10 fallacies about subtitling for the d/Deaf and the hard of hearing. *The Journal of Specialised Translation*, *10*, 128-143. Retrieved from https://www.jostrans.org/issue10/art_neves.pdf

Neves, J. (2018a). Cultures of accessibility: translation making cultural heritage in museums accessible to people of all abilities. In S. A. Harding, & O. Carbonell Cortes (Eds.), *The Routledge Handbook of Translation and Culture* (415-430). London and New York: Routledge.

Neves, J. (2018b). Subtitling for the deaf and hard of hearing audiences: moving forward. In L. Pérez-González (ed.), *The Routledge handbook of audiovisual translation* (82-95). London & New York: Routledge.

Nielsen, J. (1993). *Usability Engineering*. Boston: Academic Press.

Nornes, A. M. (1999). For an abusive subtitling. *Film Quarterly*, *52*(3), pp 17-34. http://doi.org/10.2307/1213822

O'Brien, S. (2012). Translation as human–computer interaction. *Translation Spaces*, 1(1), 101-122. https://doi.org/10.1075/ts.1.05obr

Oculus Team (2016, September 8). Oculus film short "Henry" wins an Emmy! [Web log post] Retrieved from https://www.oculus.com/blog/oculus-film-short-henry-wins-an-emmy/

O'Hagan, M., & Mangiron, C. (2013). *Game localization: translating for the global digital entertainment industry*. Philadelphia: John Benjamins Publishing Company.

O'Hagan, M. (Ed.) (2019). *The Routledge Handbook of Translation and Technology*. London and New York: Routledge.

Oncins, E. (2015). The tyranny of the tool: surtitling live performances. *Perspectives: Studies in Translation Theory and Practice*, *23*(1), 42-61. http://doi.org/10.1080/0907676X.2013.793374

Oncins, E., Orero, P., Lopes, O., & Serrano, J. (2013). All together now. A multi-language and multi-system mobile application to make live performing arts accessible. *The Journal of Specialised Translation, 20*, 147-164. Retrieved from http://www.jostrans.org/issue20/art_oncins.pdf

Orero, P. (ed.) (2004). *Topics in Audiovisual Translation*. Amsterdam and Philadelphia: 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. http://doi.org/10.1080/13670050802326766

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. *The Journal of Specialised Translation*, *30*, 105-126. Retrieved from https://www.jostrans.org/issue30/art_orero_et_al.pdf

Pannafino, J., & McNeil, P. (2017). *UX Methods: A Quick Guide to User Experience Research Methods*. CDUXP LLC.

Pedersen, J. (2011). *Subtitling norms for television: an exploration focusing on extralinguistic cultural references*. Amsterdam and Philadelphia: 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. Retrieved from https://www.jatjournal.org/index.php/jat/article/view/46

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

Perego, E., Del Missier, F., & Bottiroli, S. (2015). Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects. *Perspectives: Studies in Translation Theory and Practice* 23(1), 1-21. http://doi.org/10.1080/0907676X.2014.912343

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. (2014). *Audiovisual Translation: Theories, Methods, and Issues*, London: Routledge.

Perkins Coie LLP (2018). *2018 Augmented and Virtual Reality survey report: Insights into the future of AR/VR*. Retrieved from https://www.perkinscoie.com/images/content/1/8/v2/187785/2018-VR-AR-Survey-Digital.pdf

Reeves, B. (1991). "Being there": television as symbolic versus natural experience. Unpublished manuscript, Institute for Communication Research, Stanford University, Stanford, CA.

Remael, A., Orero, P., & Carroll, M. (Eds.) (2014). *Audiovisual translation and media accessibility at the crossroads*. Amsterdam and New York: Rodopi.

Roach, J. (2018, April 5). AI technology helps students who are deaf learn [Web log post]. Retrieved from https://blogs.microsoft.com/ai/ai-powered-captioning/

Romero-Fresco, P. (2009). More haste less speed: edited vs. Verbatim respeaking. *Vigo International Journal of Applied Linguistics*, *6*, 109-133. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=3207248

Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of the Specialised Translation*, *20*, 201-223. Retrieved from https://www.jostrans.org/issue20/art_romero.pdf

Romero-Fresco, Pablo (ed.) (2015). *The reception of subtitles for the deaf and hard-of-hearing in Europ*e. Bern: Peter Lang.

Romero-Fresco, P. (2019). *Accessible filmmaking: Integrating translation and accessibility into the filmmaking process*. Abingdon: Routledge.

Rothbaum, B., Hodges, L., Alarcon, R., Ready, D., Shahar, F., Graap, K., Pair, J., Hebert, P., Gotz, D., Wills, B., & Baltzell, D. (1999). Virtual reality exposure therapy for PTSD Vietnam veterans: a case study. *Journal of Traumatic Stress*, *12*(2), 263-71. https://doi.org/10.1023/A:1024772308758

Rothe, S., Hussmann, H., & Allary, M. (2017). Diegetic cues for guiding the viewer in cinematic virtual reality. *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden*. New York: ACM. http://doi.org/10.1145/3139131.3143421

Rothe, S., Tran, K., & Hussmann, H. (2018). Dynamic subtitles in cinematic virtual reality. *ACM TVX 2018, Seoul, Republic of Korea*, 209-214. New York: ACM. http://doi.org/10.1145/3210825.3213556

Rothe, S., Höllerer, T., & Hussmann, H. (2018). CVR-analyzer: a tool for analyzing cinematic virtual reality viewing patterns. *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, Cairo, Egypt*, 127-137. New York: ACM. http://doi.org/10.1145/3267782.3274688

Sakamoto, A. (2019). Unintended consequences of translation technologies: from project managers' perspectives. *Perspectives: Studies in Translation Theory and Practice*, *27*(1), 58-73. http://doi.org/10.1080/0907676X.2018.1473452

Sala Robert, E. (2016). *Creactive subtitles. Subtitling for all* (Doctoral dissertation,). Retrieved from E-Repositori of Universitat Pompeu Fabra https://repositori.upf.edu/handle/10230/27737

Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research*. Amsterdam: Morgan Kaufmann.

Schubert, T. (2003). The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, *15*, 69-71. https://doi.org/10.1026//1617-6383.15.2.69

Schubert, T., Friedmann, F., & Regenbrecht, H. (1999). Embodied presence in virtual environments. In R. Paton & I. Neilson (Eds.), *Visual Representations and Interpretations* (269-278). London: Springer-Verlag.

Schubert, T., Friedmann, F., & 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

Sheikh, A., Brown, A., Watson, Z., & Evans, M. (2017). Directing attention in 360-degree video. *IBC 2016*. Retrieved from https://www.ibc.org/production/directing-attention-in-360-degree-video/840.article

Sherman, W. R., & Craig, A. B. (2003). *Understanding Virtual Reality: Interface, Application, and Design*. San Francisco: Morgan Kaufmann Publishers.

Sidenmark, L., Kiefer, N., & Gellersen, H. (2019). Subtitles in interactive virtual reality: using gaze to address depth conflicts. *Proceedings of Workshop on Emerging Novel Input Devices and Interaction Techniques, Osaka, Japan*. Retrieved from https://eprints.lancs.ac.uk/id/eprint/132411/

Sketchfab (2017). *VR Industry Trends: A comprehensive overview on the state of the VR Industry in Q2-2017*. Retrieved from https://sketchfab.com/trends/q2-2017

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. Preprint manuscript retrieved from http://www.macs.hw.ac.uk/~ruth/year4VEs/Resources/pq.pdf

Slater, M., & Usoh, M. (1993). Representations systems, perceptual position, and presence in immersive virtual environments. *Presence*, *2*(3), 221-233. https://doi.org/10.1162/pres.1993.2.3.221

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Preprint manuscript retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.622&rep=rep1&type=pdf

Statista (2018). *Virtual reality (VR) video gaming sales revenue worldwide from 2015 to 2020*. Retrieved from https://www.statista.com/statistics/499714/global-virtual-reality-gaming-sales-revenue/

Statista (2019). *Forecast augmented (AR) and virtual reality (VR) market size worldwide from 2016 to 2023*. Retrieved from https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/

Steuer, J. (1992). Defining virtual reality: dimensions determining telepresence. *Journal of Communication*, *42*(4), 73-93. http://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Suojanen, T., Koskinen, K., & Tuominen, T. (2015). *User-centered translation*. London and New York: Routledge.

Szarkowska, A. (2013). Towards interlingual subtitling for the deaf and the hard of hearing. *Perspectives: Studies in Translation Theory and Practice*, *21*(1), 68-81. http://doi.org/10.1080/0907676X.2012.722650

Szarkowska, A., Krejtz, I., Kłyszejko, Z., & Wieczorek, A. (2011). Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American Annals of the Deaf*, *156*(4), 363-378. http://doi.org/10.1353/aad.2011.0039

Szarkowska, A., Krejtz, I., Krejtz, K., & Duchowski, A. (2013). Harnessing the potential of eyetracking for media accessibility. In S. Grucza, M. Płużyczka, & J. Zając (Eds.), *Translation studies and eye-tracking analysis* (153-183). Frankfurt: Peter Lang.

Szarkowska, A., Krejtz, I., Pilipczuk, Dutka, L., & Kruger, J.L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures*, *17*(2), 183–204. http://doi.org/10.1556/084.2016.17.2.3

Szarkowska, A., Jankowska, A., Krejtz, K., & Kowalski, J. (2016). Open Art: designing accessible vontent in a multimedia guide app for visitors with and without sensory impairments. In A. Matamala, & P. Orero (Eds.), *Researching Audio Description. New Approaches* (301-320). London: Palgrave Macmillan.

Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: evidence from eye movements. *PLOS ONE*, *13*(6). http://doi.org/10.1371/journal.pone.0199331

Tullis, T., & Albert, W. (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Waltham: Morgan Kaufmann.

Tuominen, T. (2016) Työkaluina käyttäjäkeskeinen kääntäminen ja käytettävyystestaus = [User-centered translation and usability testing as research tools]. In E. Pitkäsalo, & N. Isolahti (Eds.) *Kääntäminen, tulkkaus ja multimodaalisuus: Menetelmiä monimuotoisten viestien tutkimiseen = [Translation, Interpreting and Multimodality: Methods for the Study of Multimodal Messages]* (27-32). Tampere: Tampere Studies in Language, Translation and Literature.

Tuominen, T. (2018) Multi-method research: reception in context. In: E. Di Giovanni, & Y. Gambier (Eds.), *Reception Studies and Audiovisual Translation* (69-90). Amsterdam: John Benjamins Publishing Company.

Udo, J. P., & Fels, D. I. (2011). From the describer's mouth: reflections on creating unconventional audio description for live theatre. In A. Serban, A. Matamala, & J. M. Lavaur (Eds.), *Audiovisual translation in closeup: Practical and Theoretical Approaches* (257-278). New York: Peter Lang.

Ulanoff, L. (2019, January 17). Why Gen Z loves closed captioning [Web log post]. Retrieved from https://onezero.medium.com/why-gen-z-loves-closed-captioning-ec4e44b8d02f

United Nations (n/d). Convention on the Rights of Persons with Disabilities. Retrieved from https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html

Utray, F., Pereira, A. M. & Orero, P. (2009). The present and future of audio description and subtitling for the deaf and hard of hearing in Spain. *Meta*, *54*(2), 248-263. https://doi.org/10.7202/037679ar

Utray Delgado, F., Ruiz Mezcua, B., & Moreiro, J. A. (2010). Maximum font size for subtitles in standard definition digital television. Tests for a font magnifying application. In A. Matamala, & P. Orero (Eds.). *Listening to subtitles: Subtitles for the deaf and hard of hearing* (59-68). Frankfurt: Peter Lang.

Vervecken, A. (2012). Surtitling for the stage and directors's attitudes. Room for change. In A. Remael, P. Orero, & M. Carroll (Eds.). *Audiovisual Translation and Media Accessibility at the Crossroads. Media for All 3* (229-247). Amsterdam and New York: Rodopi.

Vilaró, A. (2011) Does 3D stereopsis change the way we watch movies? An eye-tracking study. Paper presented at 4th International Conference Media for All (London, 28 June-1 July 2011).

Virvou, M., & Katsionis, G. (2008). On the usability and likeability of virtual reality games for education: the case of VR-ENGAGE. *Computers & Education*, *50*(1), 154-178. https://doi.org/10.1016/j.compedu.2006.04.004

VR Intelligence (2017). *Virtual Reality Industry Survey*. Retrieved from http://vr-intelligence.com/vrx/docs/VRX-2017-Survey.pdf

Walczak, K., Cellary, W., & White, M. (2006). Virtual museum exhibitions. *Computer*, *39*(3), 93-95. http://doi.org/10.1109/MC.2006.108

Wilkinson, S. (1998). Focus group methodology: a review. *International Journal of Social Research Methodology*, *1*(3), 181-203. http://doi.org/10.1080/13645579.1998.10846874

Wissmath, B., Weibel, D., & Groner, R. (2009). Dubbing or subtitling? Effects on spatial presence, transportation, flow, and enjoyment. *Journal of Media Psychology*, *21*(3), 114-125. http://doi.org/10.1027/1864-1105.21.3.114

Witczak, O., & Jaworski, R. (2018). CAT tools usability test with eye-tracking and key-logging: where Translation Studies meets natural language processing. *Między Oryginałem a Przekładem*, *24*(41), 49-74. https://doi.org/10.12797/MOaP.24.2018.41.03

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence Teleoperators & Virtual Environments*, *7*(3), 225-240. http://doi.org/10.1162/105474698565686

Wohlsen, M. (2015). Google Cardboard’s New York Times experiment just hooked a generation on VR. *Wired Magazine*. Retrieved from https://www.wired.com/2015/11/google-cardboards-new-york-times-experiment-just-hooked-a-generation-on-vr/

Wojciechowski, R., Walczak, K., White, M., & Cellary, W. (2004). Building virtual and augmented reality museum exhibitions. *Proceedings of the Ninth International Conference on 3D Web* technology *(Web3D '04)*, 135-144. New York: ACM. http://dx.doi.org/10.1145/985040.985060

World Health Organization (n/d). *Deafness and hearing loss*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

Annexes

Annex 1 – Additional articles

1.1. Article 1: Agulló, B., Matamala, A., & 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

1.2. Article 2: 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

1.1. Article 1: Agulló, B., Matamala, A., & Orero, P. (2018). From disabilities to capabilities: testing subtitles in immersive environments with end users. *HIKMA, 17*, 195-220.

From disabilities to capabilities: testing subtitles in immersive environments with end users

De discapacidades a capacidades: testando subtítulos en medios inmersivos con usuarios

BELÉN AGULLÓ, ANNA MATAMALA, PILAR ORERO
belen.agullo@uab.cat
Universidat Autònoma de Barcelona

Fecha de recepción: 11 de septiembre de 2018
Fecha de aceptación: 29 de octubre de 2018

Abstract: User testing in Media Accessibility has often profiled users based on their disabilities. Subtitles for the deaf and hard of hearing, for instance, have been generally tested with their expected target audience, which is deaf and hard-of-hearing users. This article argues that selecting users based on sensory disabilities may not be the best strategy to obtain relevant results, as other capabilities—for instance, technological capabilities—may have a greater impact on the results. Moreover, the article argues that access services should not be exclusively for persons with disabilities but also for other audiences. If accessibility is mainstreamed, and ideally integrated in the creation and production processes, testing should expand beyond an exclusive approach based on accessibility to a more general approach based on usability where users with diverse capabilities are considered.

To illustrate this point and propose a new approach to user testing in Media Accessibility, moving from a disability to a capability model, specific examples from the European Union funded project ImAc (Immersive Accessibility) are shown in a chronological order. Then, the article presents the initial testing, targeting persons with disabilities, and describes the poor data results leading to a new approach. A new testing focus is proposed, and the methodological shift is justified. After that, the second test in which the new approach is implemented is described, using the same stimuli but users with different levels of knowledge regarding new technologies. The article finishes with conclusions and final remarks in which the door is opened to move from an accessibility approach to testing to a usability approach.

Keywords: subtitles, subtitling for the deaf and hard of hearing, immersive content, Media Accessibility, user profiling.

Resumen: En las pruebas con usuarios en los estudios de accesibilidad en los medios, generalmente se define el perfil de los usuarios en relación con su discapacidad: los subtítulos para sordos, por ejemplo, se suelen probar con usuarios sordos o con problemas de audición. En este artículo, se defiende que seleccionar a los usuarios según sus discapacidades sensoriales puede no resultar la mejor estrategia para obtener resultados relevantes, ya que otras capacidades (capacidades tecnológicas) pueden suponer un impacto mayor. Además, se argumenta que los servicios de accesibilidad no deberían estar destinados exclusivamente a las personas con discapacidades, sino que el público general también puede beneficiarse. Si los servicios de accesibilidad están cada vez más presentes en la vida de los usuarios e, idealmente, se integran en el proceso de creación y producción, las pruebas con usuarios deberían cambiar hacia un enfoque general basado en la usabilidad en lugar de exclusivamente la accesibilidad, en el que se consideren usuarios con diferentes capacidades.

Para ilustrar este marco conceptual y proponer un nuevo enfoque a la hora de abordar pruebas con usuarios en el área de accesibilidad en los medios audiovisuales, en el que se evoluciona de un modelo basado en las discapacidades a un modelo basado en las capacidades, se mostrarán ejemplos específicos en orden cronológico del proyecto ImAc (Immersive Accesibility) financiado por la Unión Europea. En la sección 2, se presentan las pruebas iniciales, que se desarrollaron con usuarios con discapacidades, y se describen los resultados insuficientes que se tradujeron en la adopción de un nuevo enfoque. En la sección 3, se propone un nuevo modelo para las pruebas con usuarios y se justifica el cambio metodológico. La sección 4 describe la segunda prueba en la que se puso en práctica este enfoque, usando los mismos estímulos, pero con usuarios con diferentes capacidades tecnológicas. En las conclusiones, se deja una puerta abierta que propone la evolución de un enfoque de pruebas con usuarios basado en la accesibilidad a un enfoque basado en la usabilidad.

Palabras clave: subtitulación, subtítulos para sordos, contenido inmersivo, accesibilidad en los medios, definición de perfil de usuarios.

INTRODUCTION[1]

Media Accessibility (MA) has been recently labelled by Greco (2016) as a field of research on its own merits. The various modalities, or access services, related to MA have been traditionally studied by different fields in order to understand them in all their complexity. Subtitling, dubbing, voice-over, audio description, audio subtitling and sign language interpreting have been approached from diverging perspectives, but two fundamental elements have always been at the centre of pioneering research: the focus on technology and on end users. Technology is basic, because it determines the service, its production, distribution and reception, and also has direct implications to quality (Bernard, Chia & Mills, 2001; Utray, Ruiz & Moreiro, 2010). Understanding the effect of various modalities on users is also fundamental, as access services are aimed at fulfilling audience needs. Audience reception has often determined the scope and approach of the research (Orero, 2008; Di Giovanni & Gambier, 2018).

Subtitling, the most popular access service, has been researched from the field of psychology (D'Ydewalle, Pollet & van Rensbergen, 1987; D'Ydewalle, Praet, Verfaillie & van Rensbergen, 1991; D'Ydewalle & Gielen, 1992; D'Ydewalle & Pavakanun, 1997; D'Ydewalle & van de Poel, 1999; De Bruycker & D'Ydewalle, 2003; Perego, Del Missier, Porta & Mosconi, 2010) to understand reading patterns and define quality in legibility and readability. From Audiovisual Translation (AVT) Studies, research has been focusing on how to produce quality content for the access services (Neves, 2007; Arnáiz-Uzquiza, 2008; Romero-Fresco, 2009; Bartoll & Martínez-Tejerina, 2010; Pereira, 2010; Romero-Fresco, 2010; Szarkowska, Krejtz, Kłyszejko & Wieczorek, 2011; Romero-Fresco, 2015; Szarkowska, Krejtz, Pilipczuk, Dutka & Kruger, 2016). In the field of engineering, attention has been paid to technical aspects of subtitles (Lambooij, Murdoch, Ijsselsteijn & Heynderickx, 2013; Hu, Kautz, Yu & Wang, 2015; Hughes, Armstrong, Jones & Crabb, 2015; Brown, 2017; Brown & Patterson, 2017; Brown et al., 2017).

Regardless the specific field and methodological research approach, technology and end user reception are always two central elements. Traditionally, however, MA services have been tested on persons with disabilities, regardless their technological capabilities (Orero & Tor-Carroggio,

[1] This article is related to the research carried out in the European funded projects ImAC (GA: 761974). The authors are members of TransMedia Catalonia, an SGR research group funded by "Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya" (2017SGR113). This article is part of Belén Agulló'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 the Universitat Autònoma de Barcelona.

2018). This user profiling approach responds to the United Nations Convention on the Rights of Persons with Disabilities (CRPD) requirement “nothing about us without us”. This article aims to challenge this exclusive approach, as it may not be always the only strategy possible to get useful results. In fact, it is argued that mainstreaming accessibility and involving user profiles beyond persons with disabilities may be more relevant in certain situations. The lack of capabilities linked to disability may have less impact than the lack of capabilities linked to other aspects such as technology. Additionally, access services are not only used by persons with disabilities. The approach in this research is in the realm of Universal Design and Design for All. Access services benefit the whole community from an inclusive perspective, hence the proposed move from accessibility to usability for diverse audiences. To prove this point, and to open up scientific debate on the need for a different approach to user profiling in MA research, the experience gained through the ImAc (Immersive Accessibility) project is put forward and discussed.

ImAc[2] is a Horizon 2020 project funded by the European Commission. It aims to research how access services (subtitling, audio description, audio subtitles, and sign language interpreting) can be integrated with immersive media, specifically 360-degree videos. 360º contents are a type of virtual reality experience. Users can both hear and look around synthetic worlds in 360° often with stereoscopic views (Mateer, 2017). The project research methodology was designed following a user-centric methodology (Orero & Matamala, 2016). End users were involved from the beginning of the project in defining system requirements for the different access services. Their input has been gathered through a series of qualitative tools, namely focus groups and interviews, in an iterative process. In the first stage (see Figure 1), general preliminary feedback was gathered from end users through a series of focus groups. Two types of users were identified as end users: professional users (i.e. professionals creating the access services or dealing with technology) and home users (i.e. persons with disabilities using the services who had some technological expertise) (Matamala et al., 2018). Focus groups were the methodological tool chosen for two reasons: they guarantee a close interaction, and there was no prototype available for testing. Therefore, open questions were put forward in order to encourage a lively discussion among the participants to trigger various possible scenarios and user interactions.

The second stage (see Figure 1) followed also in the form of focus groups and one-to-one qualitative interviews, depending on the service. To this aim, specific examples were developed for subtitles, audio description and

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

sign language interpreting. These were shown to participants, who were asked to provide feedback before implementation and further extensive testing.

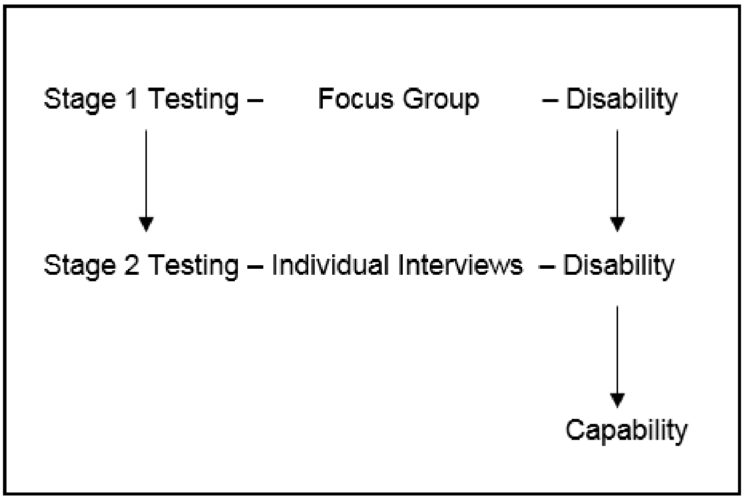


Figure 1. Testing workflow

This article reports on subtitling, because each accessibility service has a specific behaviour in immersive environments. The article discusses the methodological challenges found in this second round of user testing through interviews. Traditional profiling of persons with disabilities was initially used, leading to the deaf and hard-of-hearing community. Results, however, were unsatisfactory and the user profiling method was questioned. Consequently, the medical model of disability is challenged as the testing background for MA (Orero & Tor-Carroggio, 2018). This article aims to put forward a new approach to user profiling based on the methodological lessons learnt and shows the usefulness of the results provided by a group profiled following this new approach.

This article is developed in a chronological order following the two stages defined above in Figure 1. This choice was made to understand the process that led to a new profiling solution and presents the stages with its challenges. Section 2 of the article will present the methodological approach taken in the first tests and how the results led to rethink testing focus, moving from disabilities to capabilities. Section 3 will propose a capability-based approach when profiling end users. Section 4 will describe how this new approach was implemented using the same stimuli but different user profiles.

In Section 5 test results are presented to prove the usefulness of the new approach.

1. THE BACKGROUND: TESTING USERS WITH DISABILITIES

In the first stage of the project, immersive environments were presented to users. No prototype was available, and users identified their needs concerning subtitles for the deaf and hard of hearing (SDH) in immersive environments. This was achieved through two focus groups in which users engaged in a lively discussion guided by a facilitator. From the many suggestions two aspects were prioritised (as a precondition) for further user interaction in the second stage:

- the comfortable field of view for reading subtitles, and
- the speaker location identification.

Speaker identification and location—who is talking and where—in a 360-degree environment are paramount. The user needs to be given this basic information to navigate in the immersive world. It is taken for granted that speech carries the largest semantic load and having access to the speaker also offers secondary semantic markers such as emotions, or information regarding the speaker. Hence locating where is the person speaking who in turn is subtitled has been identified as a pre-condition (Brown et al. 2018). Similarly, the definition of the comfortable field of view to consume subtitles for 360-degree media format was considered a priority. This is because there are no standardised comfort values. Some broadcasters and content creators such as The New York Times or the BBC have developed 360-degree videos with subtitles, but still their solution is not openly shared, tested, or standardised.

For the two conditions, different alternatives were produced based on user suggestions given during the focus group sessions in the first stage. It was considered that prior to extensive experimental testing, it would be interesting to narrow down the number of possibilities through an early rating of the different alternatives. This second intervention was performed through individual interviews.

1.1. Stimuli

Two three-minute videos were recorded in German using a 360º camera. The video takes place in a radio station where two people engage in dialogue: the interviewer and interviewee. Speakers were located in such a way that users could not see both speakers in one field of view at the same time. This setup was designed to elicit user search for speakers in the 360º content. Subtitles were generated in Catalan following ISO 20071-23 recommendations regarding the number of lines (maximum of 3) and character limitation, as well as the background box and the use of Spanish colour coding to identify the different speakers (AENOR, 2003).

Figure 2. Different colours are used to differentiate the speakers

To test the comfortable field of view, the video was split into six parts. Each part was 30 seconds long and one of the six viewing field comfort levels was applied (30, 40, 50, 60, 70 and 80 per cent), departing from the value 16:9 as the natural field of view (NFV) ratio. Subtitle font size was adjusted according to the size of the field of view.

Figure 3. Different field of views and font sizes

The edges of each field of view were indicated with three different positions of the subtitles. Giving that reading is performed left to right in the writing system used for the subtitles, subtitles were prepared to be positioned: 1) bottom/centre, 2) bottom/left aligned, and 3) top/centre.

Figure 4. Different positions to delimit the different levels of the field of view

To test preferences regarding different methods for speaker location identification, the video was split in three 30-second clips. The three different options to identify the location of the speaker were indicated by: 1) an arrow, 2) a compass, and 3) sided text. For the arrow and the compass, the subtitles were presented at the bottom and centred.

Figure 5. Different speaker identification solutions

The arrow and compass icons were positioned next to the subtitle and were aligned horizontally. For the sided text, the subtitles were positioned in the centre when the speaker was in the field of view. If the speaker was outside the field of view, subtitles were aligned to the right or left edge (depending on the location of the speaker).

The stimuli above were integrated in a series of still images to guarantee the same procedure was followed by all participants. The order of presentation was not randomised, since this was planned as qualitative one-to-one interviews rather than experimental testing.

Each clip had an introduction in which the user was advised on what would be shown: a still image with the text "Level X – starts now", in which the level was indicated, was used for comfortable field of view videos. A still image with the text "Video – Guide to speaker with arrow/compass/subtitles positioned right and left – starts now" was used for the second part.

Clips were played one after another without interruption. After each part, another still image was used to display an evaluation question. For the stimuli concerning the comfortable field of views, the question was "How comfortable was viewing the subtitles? Do not take into consideration the device used but the viewing experience. Please rate it on a 5 to 1 scale." For the stimuli concerning the speaker location identification, the question was: "How useful were the arrow/compass/subtitles positioned right and left to guide to you to the speaker? Please rate it on a 5 to 1 scale." Participants were expected to produce a reply orally. All questions were asked in Catalan, as it was the language of the participants.

1.2. Procedure

The test was carried out as follows. First, each participant was welcomed. The facilitator briefly explained the context and the ImAc project and the aim of the interview. Then, a consent form was handed to the participants to be signed, in order to comply with ethical procedures. After filling in a pre-questionnaire with demographic data, participants were asked to watch each video and rate the different solutions, as explained above. An Oculus Rift HMD was used for this test. At the end of each part, and at the end of the test, they were also given the opportunity to make any comments, suggestions or recommendations regarding any aspects of subtitles in immersive media.

1.3. Participants

The expected participants were deaf and hard-of-hearing users, and it was recommended that five or six people be recruited. No further instructions on their profile was given. Involving end users with disabilities proved to be very challenging, and only two users finally turned up for the Barcelona test. This low turn-up challenged the validity of the test. Moreover, users were not familiar with virtual reality, nor were they technologically proficient. The users were Catalan speakers and with ages ranging from 50 to 60. Their feedback was biased, because they seemed to be distracted by the novelty of the interaction and did not pay attention to the main goal of the tests. The users were not comfortable with the technology and refused to move and explore the possibilities of virtual reality. One even claimed to be afraid of moving and standing up. This lack of engagement with the technology was a hindrance for the running of the test.

Other concerns were raised as a result of this test. The pre-questionnaire administered to participants also proved to be insufficient to account for all the user profiles. Although extensive work was put into designing the questionnaires for ImAc (Matamala et al., 2018) and adapting them to user needs, when participants were asked how they would define themselves, the options still followed a medical model and provided the following replies: blind person, low-vision person, deaf person, hearing-impaired person, blind-deaf person. Although this pre-questionnaire was piloted with users, when testing there were some users who did not identify themselves with any of the options.

These two methodological setbacks showed that the user profiling approach for the test might have been unsuitable. Addressing persons with disabilities with experience with immersive technologies may have been a solution. The reality is that immersive content is not accessible and, therefore, it is difficult to find experienced users. Additionally, the aim of the test was to assess usability for the field of view and the speaker location indicator, and being deaf and hard of hearing is not a requirement for carrying out such a usability test. Consequently, it was decided to change the user profiling strategy from disabilities to capabilities and to include hearing users. The next section discusses the rationale behind this methodological shift, in which the study moves away from an exclusive disability-based approach and adopts a new capability-based model.

2. Shifting the approach: from disabilities to capabilities

From early research in AVT and MA, the model chosen for user profiling was and still is the medical approach. Most studies have focused either on the collective sensory disabilities: sight and hearing. Most studies started, and they still do, with an overview defining the medical conditions of the expected or intended audience (Díaz-Cintas, Orero & Remael, 2007; Neves, 2008). In terms of metrics such as health-adjusted users-services and quality-adjusted services, there have been some studies but no debate (Romero-Fresco, 2015; Miquel-Iriarte, 2017). Shortcomings of the studies and doctoral theses point to two commonalities. The first is the heterogeneity of profiles and conditions within a group, for example the deaf and hard of hearing (Arnáiz-Uzquiza, 2012; Oncins Noguer, 2014; Tsaousi 2017; Miquel-Iriarte 2017). Choosing a testing group by one of their disabilities does not mean more homogeneous participation than choosing other demographic feature such as age (children or the aged with no disabilities), or groups of people with low levels of literacy. A person with a hearing disability who reads slowly may end up reading faster than someone who is illiterate or does not speak the language. Also, it may be the case that a person has both hearing and vision impairments, and reading becomes cumbersome and tiresome. This last example is a good illustration of the ageing population, who are not profiled in any of the sensory disability group, yet they form the highest population of users of MA services.

Moreover, while SDH is purposely designed for this single group defined by its medical pathology (Bartoll 2004, 2008, 2012), the fact is they are not the main users. Up to 85 per cent of social media video consumers watch it muted with the aid of subtitles (Patel, 2016). There is also a lack of coherence between broadcasters' claim to fulfilling deaf and hard-of-hearing requirements, while the subtitles on offer are not SDH as can be seen in the 2016 EBU report (Linder, 2016). The BBC has probably the best record, as a public broadcaster, for subtitle provision, and in 2008 claimed 100 per cent of its programmes, live or recorded, were subtitled[3]. The BBC blog suggests that their subtitles are produced for the deaf and hard-of-hearing community, yet they are more or less adapted transcriptions, with little, if any, compliance to SDH features (Neves, 2008; ISO 20071-23). A similar situation is found in the subtitles produced by broadcasters in Scandinavian countries, where subtitles have no added features, yet they are considered to cater for the deaf and hard of hearing. Video streaming platforms such as HBO, Amazon Prime Video and Netflix, champion of SDH production, still abound with irregular subtitling styles and are often no better than transcriptions. This is the case for popular series such as *Breaking Bad* where mixed styles and conventions were used across the episodes, and even within the same episode.

The lack of service terminology agreement (caption/subtitle/SDH) and a heterogeneous population constantly challenge and impact on scientific studies in MA. As this article shows, and it adds to the list of failed studies, profiling end users within a medical framework (Marks, 1997) to perform tests on human interaction capabilities does not make sense. In 2001, the UN World

[3] https://bbc.in/2zeBkw3 [retrieved 20/05/2018]

Health Organisation published the International Classification of Functioning, Disability and Health (ICF). The ICF was intended to complement its sister classification system, the International Classification of Diseases (ICD) (Brown & Lent, 2008). The ICF model sees disability as the result of a combination of individual, institutional and societal factors that define the environment surrounding a person with an impairment (Dubois & Trani, 2009). It is operationalised through the World Health Organization Disability Assessment Schedule II (WHODAS II) and it covers all types of disabilities, for various countries, languages and contexts, which makes it suitable for cross-cultural use. Dubois and Trani (2009) consider the ICF to be limited in its scope and use, as its primary purpose is classification. They believe the complexity of disability requires a wider and more comprehensive analytical view. Ellis (2016) raised also this issue underlying the difference between disability and impairment, offering examples where, under the same conditions – a noisy party –, the deaf and hard-of-hearing person stands more chance of good communication than a hearing person, because they can read lips, or in a dark room a blind person will be able to navigate better than a sighted person. Ellis presents the environment as the disabler, and not the physical condition. This concept was adopted by the UN agency International Telecommunication Union (ITU). In 2017, they released a report addressing access to telecommunication/Information and Communication Technologies (ICT) services by persons with disabilities and with specific needs that stated the following:

> Besides the more commonly used "medical model of disability", which consider disability "a physical, mental, or psychological condition that limits a person's activities", there is a more recent "social model of disability," which has emerged and is considered a more effective or empowering conceptual framework for promoting the full inclusion of persons with disabilities in society. Within the social model, a disability results when a person who (a) has difficulties reading and writing; (b) attempts to communicate, yet does not understand or speak the national or local language, and (c) has never before operated a phone or computer attempts to use one – with no success. In all cases, disability has occurred, because the person was not able to interact with his or her environment. (ITU, 2017: 2)

This implies that it is of little or no use to profile by disabilities according to a medical model in MA studies. The ITU is calling for a new social model approach that analyses different aspects of each individual that might have an influence on what researchers are testing.

However, the social model falls short when offering a framework to define and profile end users for tests, as the object of study is the performance of the person given a task under a determinate condition. It will be the competence to perform the task that should be analysed, and questionnaires should be defined accordingly. In other words, defining the users to perform the tasks should not be based on a medical condition, but the capability to perform the task. Selecting relevant capabilities or "functionings" to form an "evaluative space" is needed (Mitra, 2006). The approach should continue by drafting a set of "functionings" (or capabilities), a method to measure them, and a threshold below which a person is considered to have a deprivation. This can be applied to people with similar personal characteristics, commodities, and environment (Orero & Tor-Carroggio, 2018). This has already been found relevant in previous studies such as Romero-Fresco (2015) who pointed out that reading subtitles was related to a person's educational background rather than to their hearing impairment. This is the starting point when revising the user-centred research on MA and implementing a new capabilities model.

3. Implementing the new capabilities profiling

Given that tests with end users profiled using a disability-based medical model were not as successful as expected, it was decided to carry out tests using the same stimuli and methodology but changing user profiling and focusing on technology capabilities rather than on sensory disabilities. The main reason behind this choice is that immersive content consumption requires advanced knowledge of technology, specifically immersive technologies that are not mainstream. Immersive content is currently not accessible and, therefore, most users are not familiar with the technology. It was consequently considered that users were to be recruited depending on their capabilities regarding technology rather than on their sensory disability. Moreover, it was thought that subtitles would not only benefit persons who cannot hear the original but also persons who do not understand the language (German in this case). In this regard, rather than testing for accessibility, testing was focused on usability. Accessibility had already been taken into account for prototyping, and access solutions had been implemented based on the feedback from end users with disabilities. The priority was to find users with different levels of technology knowledge to evaluate user interaction with subtitles in immersive media, regardless of their hearing abilities, in order to suggest the best strategies for all users and mainstream access services.

3.1. Methodology adjustments and participants

Three different age ranges were considered when recruiting participants following the Prensky (2001) classification: from digital native to digital immigrant. It was expected that this would have an impact on their interaction and familiarity with immersive technology: children/teenagers (from 12 to 18 years old), young adults (from 25 to 30 years old) and adults (from 52 to 60 years old). User-profiling through a pre-questionnaire addressed two main capabilities: technological proficiency (can the user interact with the technology?) and subtitles consumption (can the user read subtitles?). It was assumed that children/teenager and young adult groups would be more capable with technology and subtitles than the adult group, which was expected to be more technology disabled, or digital immigrant. This assumption was later confirmed with the demographic pre-questionnaires.

To inform the different profiles, the demographic questionnaire designed at the beginning of the project was modified, only asking questions that were capability-relevant. Some adjustments were needed:

- Demographic questions related to sex, main language and disability were removed, since they were not considered to have an impact on the results for this usability test.
- Demographic questions related to age and level of studies were retained, since they were considered to have an impact on the results for this usability test.
- Other questions related to technology were also maintained:
  - What technology do you use on a daily basis?
  - Do you own any device used to access virtual reality content?
- New questions related to technology/subtitles habits and knowledge were included, such as:
  - How often do you watch virtual reality content (such as 360-degree videos)?
  - If you have never or only occasionally used virtual reality content such as 360-degree videos, please indicate why.
  - Please state your level of agreement with the following statement: “I am interested in virtual reality content (such as 360-degree videos).”
  - Do you like watching the following types of content on television or online?
  - When subtitling is available, do you activate it for the following type of content?
  - If it is available and you do not activate it, please select the reasons why.
  - How many hours a day do you watch subtitled content?
  - What do you use subtitles for?

New questions were aimed at profiling end users' capabilities related to reading subtitles or addressing the technology at stake. According to Orero et al. (2018), questions about TV viewing habits should always be included in AVT research questionnaires. This could be applied to other technologies apart from TV, such as immersive media.

Six users participated in the one-to-one interviews, two for each age group. Two participants (one young adult and one teenager) had previous experience with immersive contents. Two young adults and one teenager used subtitles frequently, and all users were familiar with subtitling practices. All users were engaged with the technology and showed interest in it. However, interaction with the technology differed among the profiled groups. Digital natives (teenagers and young adults) felt confident and behaved naturally when testing the different solutions. Digital immigrants (adults) seemed restrained by the technology and behaved more cautiously.

4. Results for subtitles in immersive media

The results from the test after adopting this new methodological approach are presented below, to show its relevance. Regarding comfortable field of view, level 5 (70 per cent of NVF) and level 6 (80 per cent of NVF) got the highest rating, followed by level 4 (60 per cent of NVF) and level 3 (50 per cent of NVF). Levels 1 and 2 were too small and difficult to read for all users.

Comfort	**Level 1**	**2**	**3**	**4**	**5**	**6**
P1	3	3	2	1	1	2
P2	2	2	3	3	4	4
P3	1	1	3	3	3	3
P4	1	2	3	3	4	1
P5	1	1	2	3	3	4

P6	1	1	2	3	4	5
Mean	1.5	1.7	2.5	2.7	3.2	3.2

Table 1. Ratings from users regarding comfortable field of view

Regarding subtitle placement, users generally preferred subtitles at the bottom of the field of view, because they claimed to be used to that position. Some also argued that reading the subtitles located at the top of the field of view was tiring and uncomfortable. Moreover, they did not like it when subtitles covered part of the image as they found it annoying and frustrating. As for the font size, for most users the biggest font was easier to read. A balance is required between font size and not covering the images, however. In any case, it seems that personalisation for font size may be the best solution.

Some users reported double vision when reading the subtitles. This was due to the implementation of the subtitles: they were too close to the eyes and need to be closer to the image to avoid depth issues. It was agreed with the development team that implementation needs to be improved in this aspect. Finally, one user also reported colour blindness issues. Personalisation for font colour therefore should be implemented.

As far as the speaker location identification is concerned, which was assessed through a second video, the following preferences for the different approaches applied: 1) arrow 2) compass and 3) sided text.

Guiding	**Position**	**Arrow**	**Compass**
P1	1	4	2
P2	1	5	3
P3	1	3	4
P4	1	3	3

P5	1	2	3
P6	1	4	1
Mean	1	3.5	2.7

Table 2. Ratings from users regarding methods for speaker location identification

Regarding the implementation of speaker location identification methods, some suggestions for improvement were offered. For example, one user noticed that the vertical axis (up and down) was missing from all options and should be implemented. Also, researchers realised that further tests on how to implement directions when two speakers (or more) are talking at the same time in different locations need to be carried out. This was confusing for the users. As far as the display of the indicators is concerned, most users preferred the arrow. They would prefer it to be clearer, though: bigger and in a colour different from the subtitle, to make it more noticeable.

Finally, two users claimed that all the three methods that were presented were unclear and confusing. One user suggested that, apart from the arrows, an indicator close to or above the speaker (for example, a red dot) could be implemented. This way, the users can be sure about who is talking once their eye has reached the speaker in the image. This was suggested by a user from the digital immigrants. The solution could be helpful for both digital immigrants and deaf users, for example, if the mouth of the speaker cannot be seen in the image.

5. Evaluating the new approach

The results presented in the previous section show that the input provided by participants in the test with the new approach was relevant, even if participants did not have hearing impairments. Age and technological capabilities proved to be a determining factor in the participants' feedback. Teenagers and millennials generally found the systems in place usable. As for the implementation of subtitles, they suggested some improvements in terms of aesthetical characteristics and smoothness of use of the system. They also required further enhancements to the presented solutions. Digital natives are given a solution and they automatically look for improvements and additional features. However, they are not worried about accessibility, because they can usually find workarounds. Adults, on the other hand, were not completely happy with the usability of the subtitles nor the speaker identification systems, which they found uncomfortable and sometimes confusing. They suggested possible solutions to make the subtitles more accessible to technology impaired population.

Thanks to this new approach, an implementation issue regarding sight (and not hearing) was detected. Two users reported seeing the subtitles duplicated. Further research shed some light on this issue. Convergence capacity and prismatic effect caused by lenses can provoke double vision. This can consequently impact on how the subtitles are seen in an immersive environment when wearing a head-mounted display and solutions need to be implemented.

Changing the focus from disabilities to capabilities in MA research when access services are heavily dependent on technology seems to be a fruitful approach. In this test, the most important questions to be addressed were: 1) can you read the subtitles? If so, which is the most comfortable solution?; and 2) can you identify where the speaker is? If so, which is the most usable and efficient solution? Also, user interaction was crucial, due to the novelty of immersive technologies. That was the reason for profiling different user groups depending on technological knowledge, to look for and find different user needs, prioritising the access to the content. Depending on the service and technology to be tested, different capabilities may need to be considered when selecting users. ImAc has moved from a medical model based on a sensory disability to a model based on capabilities, in this case related to technology and subtitle reading.

6. Conclusion

Some authors claim that access services should not be considered as an afterthought, but rather they should be part of the design from the beginning of the development process (Udo & Fels, 2010; Romero-Fresco, 2013). This can be applied to filmmaking, and it is also valid when it comes to making new media technologies accessible. This was the possibility that was raised during the ImAc project: implementing access services in 360-degree content before the technology and contents were fully mainstreamed, through the cooperation of experts from different fields.

Regarding SDH, feedback was initially gathered from end users in a focus group, to anticipate their needs before implementing the access service in the new medium. From the draft requirements and further technical discussions, two aspects were prioritised for a second round of user testing:

comfortable field of view for reading subtitles and different methods for identifying speaker location. Then, these two elements were tested by a reduced group of users to fine-tune the implementation specifications for SDH in immersive media. The results suggested that users preferred 70 per cent and 80 per cent of the natural field of view (16:9) and the largest font size. This might be due to the lack of habit in reading subtitles in immersive media, as bigger fonts usually are easier to read. However, users also raised their concern about the blocking effect of subtitles when consuming contents in immersive media. Sometimes, subtitles covered the image, and this was received as a negative effect. Implementation solutions need to focus on preventing blocking from occurring. Also, users preferred the arrow system for speaker location identification, although some improvements for usability were suggested.

Beyond the specific results obtained in the user testing which was limited due to the number of participants, the article has illustrated the challenges of designing and implementing access services following a user-centred methodology when access services are heavily dependent on technology. It has shown how a model based on disabilities may yield unsatisfactory results, especially when the technology and the content are not mainstreamed, and users are not familiar with them. It is precisely within this context that another approach to user profiling has been proposed as the central element in this article.

MA research has traditionally used the medical model to profile end users for their experimental research (Orero & Tor-Carroggio, 2018). A report issued by the UN agency International Telecommunication Union (ITU) (2017), however, calls for a new approach that substitutes the medical model for disabilities. In this new approach, different individual aspects such as literacy, language, and technology proficiency also impact on user interaction with technology. In this study, it became clear that the medical model needed to be adapted to meet the needs of emerging technologies, such as immersive media. The capability approach goes one step forward and explains that what needs to be considered is not the users' disabilities, but their capabilities given a specific task in a specific environment and with a specific technology (Orero & Tor-Carroggio, 2018). This approach, in which user profiling is adapted to match technology capabilities, has been successfully applied in this research.

In the same way, in a move from a model based on disabilities to a model based on capabilities, MA research may also benefit by moving from an exclusive accessibility-based approach to a more general usability approach in which the needs of different types of users are considered. SDH has the potential to benefit not only persons with hearing impairments but all users. If SDH is mainstreamed, usability tests may be more relevant than accessibility tests alone. As argued by Tullis and Albert (2013: 229), “accessibility is really just usability for a particular set of users”, namely “users with different types of disabilities” (ibid: 230). If there is a move from disabilities to capabilities, the focus should be on usability tests in which users with different capabilities, who are relevant to the object of research, are taken into account. Therefore, an interesting development of this study in future research would be to consider users with different key capabilities in relation to the actual content and technology tested, namely persons with different technological capabilities, persons with different hearing capabilities, and persons with different subtitling capabilities. This approach would also help overcoming current obstacles in MA research such as the low number of participants and the lack of statistical significance (Orero & Tor-Carroggio, 2018.

Changes in how MA is understood are necessary. Subtitles are a valid service to many users, not only persons with disabilities, so a broader user profiling should be adopted, moving from disabilities to capabilities and moving from accessibility to usability. In a constantly evolving world with new technologies being implemented at a fast pace, a flexible and adaptable workflow that favours the implementation of access services from the outset needs to be established. The collaboration between different stakeholders (broadcasters, engineers, media access scholars, end users) may offer the possibility of improving not just the quantity of accessible media, but also the quality, because access services would be approached from different perspectives, enriching the final outcome.

7. Bibliography

Asociación Española de Normalización y Certificación (AENOR) (2003). *Subtitulado para personas sordas y personas con discapacidad auditiva. Subtitulado a través del teletexto* (Norma UNE 153010). Retrieved from https://bit.ly/2z0fD3t.

Arnáiz-Uzquiza, V. (2008). La Objetividad en el Subtitulado. Justificación de los Parámetros Formales Mediante EyeTracking. In A. Pérez-Ugena, and R. Vizcaíno (Eds.), *ULISES: Hacia el desarrollo de tecnologías comunicativas para la igualdad de oportunidades. Retos y perspectivas para sordos signantes* (pp. 73-85). Madrid, España: Observatorio de las Realidades Sociales y de la Comunicación.

— (2012). *Subtitling for the deaf and the hard-of-hearing: Some parameters and their evaluation* (Doctoral thesis). Universitat Autònoma de Barcelona, Spain. Retrieved from: https://bit.ly/2PvBXvr

Bartoll, E. (2004). Parameters for the classification of subtitles. In P. Orero (Ed.), *Topics in Audiovisual Translation* (pp. 53-60). Amsterdam, Netherlands and Philadelphia, USA: John Benjamins.

— (2008). *Paràmetres per a una taxonomia de la subtitulació* (Doctoral thesis). Universitat Pompeu Fabra, Spain. Retrieved from https://bit.ly/2zXK1KQ

— (2012). *La subtitulació: Aspectes teòrics i pràctics*. Barcelona, Spain: Eumo.

— & Martínez-Tejerina, A. (2010). The positioning of subtitles for the deaf and hard of hearing. In A. Matamala, and P. Orero (Eds.), *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing* (pp. 69-86). Bern, Switzerland: Peter Lang.

Bernard, M., Chia, H. L., and Mills, M. (2001). The effect of font type and size on the legibility and reading time of online text by older adults. In *CHI EA'01 Extended Abstracts on Human Factors in Computer Systems* (pp. 175-176). New York, USA: ACM.

Brown, A. and Patterson, J. (2017). *Designing Subtitles for 360° Content*. In BBC Research & Development blog. Retrieved from https://bbc.in/2qPLJd9

— (2017). *User Testing Subtitles for 360° Content*. In BBC Research & Development blog. Retrieved from https://bbc.in/2RV43Nx

—, Turner, J., Patterson, J., Schmitz, A., Armstrong, M., and Glancy, M. (2017). Subtitles in 360-degree Video. In *TVX '17 ACM International Conference on Interactive Experiences for TV and Online Video* (pp. 3-8). New York, USA: ACM.

—. Turner, J., Patterson, J., Schmitz, A., Armstrong, M., and Glancy, M. (2018). *Exploring Subtitle Behaviour for 360° Video*. [White Paper]. Retrieved from https://bbc.in/2S0EdI3

— & Lent, R. (Eds.) (2008*). Handbook of Counseling Psychology*. New Jersey, USA: Wiley.

Díaz-Cintas, J., Orero, P., and Remael, A. (Eds.) (2007). *Media for all: subtitling for the deaf, audio description, and sign language*. Amsterdam, Netherlands: Rodopi.

Dubois, J. L., and Trani, J. F. (2009). Extending the capability paradigm to address the complexity of disability. *ALTER, European Journal of Disability Research, 3*, 192–218.

D'Ydewalle, G., Pollet, J., and van Rensbergen, J. (1987). Reading a message when the same message is available auditorily in another language: the case of subtitling. In J.K. O'Regan, and A. Lévy-Schoen (Eds.), *Eye Movements: From Physiology to Cognition* (pp. 313-321). North Holland, Netherlands: Elsevier Science Publishers.

—., Praet, C., Verfaillie, K., and van Rensbergen, J. (1991). Watching Subtitled Television: Automatic Reading Behaviour. *Communication Research*, *18*(5), 650-666.

—. & Gielen, I. (1992). Attention Allocation with Overlapping Sound, Image, and Text. In K. Rayner (Ed.), *Eye Movements and Visual Cognition* (pp. 415-527). New York, USA: Springer Verlag.

— & Pavakanun, U. (1997). Could enjoying a movie lead to language acquisition? In P. Winterhoff-Spurk and T.H. van der Voort (Eds.), *New horizons in media psychology* (pp. 145–55). Opladen, Germnay: Westdeutscher Verlag.

— & van de Poel, M. (1999). Incidental foreign language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research, 28*, 227–244.

De Bruycker, W.Y., and d'Ydewalle, G. (2003). Reading Native and Foreign Language Television Subtitles in Children and Adults. In J. Hyöna, R. Radach, and H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 671-684). Amsterdam, Netherlands: Elsevier Science Publishers.

Di Giovanni, E., and Gambier, Y. (Eds.) (2018). *Reception Studies and Audiovisual Translation*. Amsterdam, Netherlands: John Benjamins.

Ellis, G. (2016). Impairment and Disability: Challenging Concepts of "Normality". In A. Matamala and P. Orero (Eds.), *Researching Audio Description* (pp. 35-45). London, UK: Palgrave Macmillan.

Greco, G. M. (2016). On Accessibility as a Human Right, with an application to Media Accessibility. In A. Matamala, and P. Orero (Eds.), *Researching Audio Description*. London, UK: Palgrave Macmillan.

Hu, Y., Kautz, J., Yu, Y., and Wang, W. (2015). Speaker-Following Video Subtitles. *ACM Transactions on Multimedia Computing, Communications, and Applications, 11*(2).

Hughes, C. J., Armstrong, M., Jones, R., and Crabb, M. (2015). Responsive Design for Personalised Subtitles. In *W4A '15 Proceedings of the 12th Web for All Conference*. New York, USA: ACM.

International Organization for Standardization (forthcoming). *Information technology -- User interface component accessibility -- Part 23: Visual presentation of audio information (including captions and subtitles)* (Standard no. ISO/IEC FDIS 20071-23). Retrieved from https://bit.ly/2qOSJH1

International Telecommunication Union (ITU) (2017). *Question 7/1: Access to telecommunication/ICT services by persons with disabilities and with specific needs* [Report]. Retrieved from https://bit.ly/2Q1VyD3

Lambooij, M., Murdoch, M. J., Ijsselsteijn, W. A., and Heynderickx, I. (2013). The impact of video characteristics and subtitles on visual comfort of 3D TV. *Displays 34*(1), 8-16.

Linder, G. (2016). *ACCESS SERVICES PAN EUROPEAN SURVEY 2016* [Report]. Retrieved from https://bit.ly/2Drvb2L

Marks, D. (1997). Models of Disability. *Disability and Rehabilitation, 19*(3), 85-91.

Matamala, A., Orero, P., Rovira-Esteva, S., Casas-Tost, H., Morales, L. F., Soler-Vilageliu, O., Agulló, B., Fidyka, A., Segura, D. and Tor-Carroggio, I. (2018). User-centric approaches in Access Services Evaluation: Profiling the End User. In I. Schuurman, L. Sevens, V. Yaneva, J. O'Flaherty (Eds.), *Proceedings of the LREC 2018 Workshop. "Improving Social Inclusion using NLP: Tools, Methods and Resources"* (pp. 1-8). Miyazaki, Japan: ISI-NLP 2.

Mateer, J. 2017. Directing for Cinematic Virtual Reality: How the Traditional Film Director's Craft Applies to Immersive Environments and Notions of Presence. *Journal of Media Practice 18*(1), 14-25.

Miquel-Iriarte, M. (2017). *The reception of subtitling for the deaf and hard of hearing: viewers' hearing and communication profile & Subtitling speed of exposure* (Doctoral thesis). Universitat Autònoma de Barcelona, Spain. Retrieved from: https://bit.ly/2QPz5WU

Mitra, S. (2006). The Capability Approach and Disability. *Journal of Disability Policy Studies, 16*(4), 236-247.

Neves, J. (2007). There is Research and Research: Subtitling for the Deaf and Hard of Hearing (SDH). In C. Jiménez (Ed.), *Traducción y Accesibilidad. Subtitulación para sordos y audiodescripción para ciegos: Nuevas modalidades de Traducción Audiovisual*. Frankfurt, Germany: Peter Lang.

Neves, J. (2008). 10 fallacies about Subtitling for the d/Deaf and the hard of hearing. *The Journal of Specialised Translation, 10*, 128-143.

Oncins Noguer, E. (2014). *Accessibility for the scenic arts* (Doctoral thesis). Universitat Autònoma de Barcelona, Spain. Retrieved from https://bit.ly/2BcMKlt

Orero, P. (2008). Three different receptions of the same film. The Pear Stories applied to Audio Description. *European Journal of English Studies 12*(2), 179–193.

— & Matamala, A. (2016). User-centric audio description: a topsy-turvy research approach. In A. Manco, and A. Mancini (Eds.), *Scrittura brevi: segni, testi e contesti. Dalle iscrizione antiche ai tweet* (pp. 376-387). Naples, Italy: Università degli Studi di Napoli "L'Orientale”.

—, Doherty, S., Kruger, J.L., Matamala, A., Pedersen, J., Perego, E., Romero-Fresco, P., Rovira-Esteva, S., Soler-Vilageliu, O., and Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *The Journal of Specialised Translation, 30*, 105-126.

— & 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. 105-122). Switzerland: Springer International Publishing.

Patel, S. (2016). 85 percent of Facebook video is watched without sound. In *DigidayUK*. Retrieved from https://bit.ly/2tZlgdv

Perego, E., Del Missier, F., Porta, M., and Mosconi. M. (2010). The Cognitive Effectiveness of Subtitle Processing. *Media Psychology 13*(3), 243–272.

Pereira, A. (2010). Criteria for elaborating subtitles for the deaf and hard of hearing adults in Spain: Description of a case study. In A. Matamala, and P. Orero (Eds.), *Listening to subtitles* (pp. 87-102). Vienna, Austria: Peter Lang.

Prensky, M. (2001). Digital Natives, Digital Immigrants. *On the Horizon, 9* (5), 1–6.

Romero-Fresco, P. (2009). More haste less speed: edited vs. Verbatim respeaking. *Vigo International Journal of Applied Linguistics, 6*, 109–133.

—. (2010). D'Artagnan and the Seven Musketeers: SUBSORDIG travels to Europe. In A. Matamala, and P. Orero (Eds.), *Listening to subtitles* (pp. 175-189). Vienna, Austria: Peter Lang.

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

— (ed.) (2015). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe*. Bern, Switzerland: Peter Lang.

Szarkowska, A., Krejtz, I., Kłyszejko, Z., and Wieczorek, A. (2011). Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American Annals of the Deaf, 156*(4), 363-378.

—, Krejtz, I., Pilipczuk, Dutka, L., and Kruger, J.L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures, 17*(2), 183-204

Tsaousi, A. (2017). *Sound effect labelling in Subtitling for the D/deaf and Hard-of-Hearing: An experimental study on audience reception* (Doctoral Thesis). Universitat Autònoma de Barcelona, Spain. Retrieved from: https://bit.ly/2OQ4yGR

Tullis, T., and Albert, B. (2013). *Measuring user experience. Collecting, analyzing, and presenting usability metrics*. USA: Morgan Kaufmann.

Udo, J. P., and Fels, D. I. (2010). The rogue poster-children of universal design: closed captioning and audio description. *Journal of Engineering Design, 21*(2-3), 207-221.

Utray, F., Ruiz B., and Moreiro, J.A. (2010). Maximum font size for subtitles in Standard Definition Digital Television: Tests for a font magnifying application. In A. Matamala, and P. Orero (Eds.), *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing* (pp. 59-68). Bern, Switzerland: Peter Lang.

1.2. Article 2: 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.

Artificial Intelligence for Engineering Design, Analysis and Manufacturing

cambridge.org/aie

Research Article

Cite this article: 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* 1–13. https://doi.org/10.1017/S0890060419000362

Received: 30 November 2018
Revised: 8 March 2019
Accepted: 1 April 2019

Key words:
360° Videos; accessibility; human–computer interaction; subtitles; virtual reality

Author for correspondence:
Belén Agulló,
E-mail: belen.agullo.garcia@gmail.com

Making interaction with virtual reality accessible: rendering and guiding methods for subtitles

Belén Agulló[1], Mario Montagud[2,3] and Isaac Fraile[3]

[1]Universitat Autònoma de Barcelona, Spain; [2]University of Valencia, Spain and [3]i2Cat Foundation, Spain

Abstract

Accessibility in immersive media is a relevant research topic, still in its infancy. This article explores the appropriateness of two rendering modes (fixed-positioned and always-visible) and two guiding methods (arrows and auto-positioning) for subtitles in 360° video. All considered conditions have been implemented and integrated in an end-to-end platform (from production to consumption) for their validation and evaluation. A pilot study with end users has been conducted with the goals of determining the preferred options by users, the options that result in a higher presence, and of gathering extra valuable feedback from the end users. The obtained results reflect that, for the considered 360° content types, always-visible subtitles were more preferred by end users and received better results in the presence questionnaire than the fixed-positioned subtitles. Regarding guiding methods, participants preferred arrows over auto-positioning because arrows were considered more intuitive and easier to follow and reported better results in the presence questionnaire.

Introduction

There is a growing interest in virtual reality (VR) and the possibilities to develop immersive content, such as 360° videos. Viewers can watch 360° clips with head-mounted displays (HMDs) or directly from a flat screen on a smartphone or computer. In these videos, the viewers have the freedom to look around and explore the virtual scenarios that are presented to them. YouTube, Facebook, Jaunt VR, and The New York Times VR are some of the companies that are developing immersive experiences for their audience via online platforms. According to a report issued by the European Broadcasting Union (EBU, 2017), 49% of its members are exploring and devoting efforts into the production and development of immersive content and services, respectively. EBU members believe that immersive content presents a clear potential for broadcasters and content creators because it offers the opportunity to provide more interactive and engaging storytelling. For content creators and filmmakers, one of the main challenges when developing immersive content is the lack of control over the main focus of the video. Therefore, intelligent and effective strategies to present the content, attract and keep audience's attention and assist users' need to be explored and adopted. Nonetheless, interactive and immersive content creation development are still at an early stage and research is ongoing (Dooley, 2017; Mateer, 2017; Rothe *et al.*, 2017; Sheikh *et al.*, 2017). The main lines of research are focused on storytelling; for example, how to attract viewers' attention or the creation of a new screen grammar for 360° content.

Apart from open challenges in terms of high-resolution content, interaction, and storytelling formats for immersive media, a key issue needs to be taken into account: accessibility. It is not acceptable to consider accessibility as an afterthought, but it instead must be addressed in the specification and deployment of end-to-end immersive systems and services. Such an objective involves overcoming existing limitations in current technologies and systems to enable truly inclusive, immersive, and personalized experiences, adapted to the needs and/ or preferences of the users. Studies on access to audio-visual content can be found in the field of media accessibility (Remael *et al.*, 2014; Greco, 2016). The main access services under research have been subtitling for the deaf and hard-of-hearing, audio description, or sign language interpreting. Most studies have been carried out in the context of traditional audio-visual media, such as TV or cinema (Perego *et al.*, 2015; Romero-Fresco, 2015). In these cases, accessibility has generally been considered as an afterthought. However, some researchers have raised their voices in favor of including accessibility in the creation process (Romero-Fresco, 2013).

Although immersive technologies and content are on the rise, research studies on, and thus solutions for, accessibility in immersive media are limited so far. This hinders the interaction of part of the population with VR experiences. Proper technological solutions, interfaces, and recommendations need to be sought to ensure a proper narrative, interpretation of content and usability, regardless of the capacities of the users, their age, language, and/or other specific impairments. This will contribute to a global e-inclusion, offering equal opportunities for access to the whole consumers' spectrum, while ensuring compliance with regulatory guidelines (e.g., human rights obligations).

Many efforts must be devoted to providing efficient solutions and meaningful insights to, among others, the following research questions in this field:

- What are the requirements to enable truly accessible and inclusive immersive services?
- How current (immersive) technologies and systems can be augmented to seamlessly integrate and support accessibility services?
- What kind of assistive technologies can contribute to better accessibility in immersive media?
- Which presentation modes for accessibility content are better suited for specific content types?
- What personalization features should be provided to meet users' needs and preferences?
- What benefits are provided (e.g., in terms of usability, content comprehension, level of immersion, and engagement)? How to properly evaluate and define them?

By comparing with traditional audio-visual content, the integration of access services (i.e., subtitles, sign language interpreting, and audio description) faces two main challenges. First, there is more information to process, and users can feel overwhelmed. Second, the presentation is no longer purely time-based, but it involves a spatial dimension, determined by both the user's field of view (FoV) and the direction where the main actions are taking place.

This also applies to subtitles, which is one of the most mainstreamed access services, being provided by major TV channels, like BBC (Armstrong *et al.*, 2015), and Video on Demand platforms, such as Netflix, HBO, or Amazon Video. Subtitles are not only beneficial for viewers with hearing impairments but also for users with visual impairments if their presentation format can be customized, for non-native speakers, to support the comprehension of content, and in noisy/public environments where the audio cannot be listened or cannot be turned on. For example, up to 85% of Facebook videos are watched muted with the aid of subtitles (Patel, 2016). Beyond contributing to overcome audio-visual barriers, the applicability of subtitles enters the realm of other forms of social integration, can have an impact on education and on therapy, and can contribute to increase the engagement and Quality of Experience (QoE).

This article focuses on two essential issues in this research area: how to present subtitles in 360° videos without breaking immersion and how to guide the users for a more effective and a non-intrusive interaction and storytelling comprehension. The research tasks are being devoted after having conducted user-centric activities to gather requirements from which the proposed solutions have been derived (Agulló and Matamala, 2019; Agulló *et al.*, 2018). Two strategies are proposed and assessed for subtitle rendering modes: (1) always-visible – the subtitles are anchored to the FoV, always in the same bottom center position, regardless of where the user is looking at within the 360° and (2) fixed-positioned – the subtitles are anchored to the 360° video, being rendered in three fixed positions, evenly spaced every 120° around the 360° sphere. Likewise, two strategies are proposed and assessed for guiding methods (in order to guide the viewer to the speaker), when making use of the always-visible rendering mode: (1) arrows – a visual element (arrow) is displayed next to the subtitle to indicate the viewers where they need to look at to find the target speaker and (2) auto-positioning – an intelligent strategy that consists of automatically adjusting the FoV to match the position of the targeted speaker(s)/main action(s), by smoothly and automatically rotating the camera, as in Lin *et al.* (2017). Both strategies have been developed and tested in a pilot study. Their integration in an end-to-end platform, paying special attention to the content consumption part, the followed evaluation methodology, and the obtained results regarding the impact on immersion and the participants' preferences are reported in this article.

The rest of the article has been structured as follows: In the "Related work" section, the state of the art in this field is reviewed. In the "End-to-end platform for immersive accessibility" section, an overview of the developed end-to-end platform for the integration of accessibility services in immersive media is provided. This platform has served as the framework for conducting the pilot study. Next, the evaluation setup, methodology and obtained results are reported. Finally, the results and their scope are discussed, and some ideas for future work are provided in the "Conclusions and future work" section.

Related work

VR as a form of entertainment, especially in the form of 360° content or cinematic VR (Mateer, 2017), has attracted the interest of the research community and industry from different perspectives. There are several studies on narrative in VR (Aylett and Louchart, 2003; Dooley, 2017; Gödde *et al.*, 2018), mostly focused on better understanding the complexities of this new medium. Other studies are tackling the specific topic of focus and attracting attention in VR (Mateer, 2017; Rothe *et al.*, 2017; Sheikh *et al.*, 2017). In addition, some researchers have carried out studies on the impact of cinematic VR on immersion (De la Peña *et al.*, 2010; Cummings and Bailenson, 2016; Jones, 2017) and engagement (Wang *et al.*, 2018).

However, research on subtitling in immersive content is limited. There are few exceptions. The study carried out by the BBC (Brown *et al.*, 2017) was the first considering this topic and proposing some solutions. The main challenges identified by the BBC research team when developing subtitles for immersive content are as follows (Brown *et al.*, 2017):

- there is no area in the scene that is guaranteed to be visible to the viewer, so it is not possible to control what will appear behind the subtitle;
- immersion is important in this medium, so subtitles should not disrupt the experience;
- if subtitles are located outside of the FoV, then the effort to find them should be minimum; and
- including subtitles should not worsen the VR sickness effect.

Based on these premises, the BBC developed four solutions for subtitle rendering (Brown *et al.*, 2018):

a) Evenly spaced: subtitles are equally spaced with a separation of 120° in a fixed position below the eye line;
b) Follow head immediately: subtitles follow the viewer as he/she looks around, displayed always in front of him/her;

c) Follow with lag: subtitles appear directly in front of the viewer, and they remain there until the viewers look somewhere else; then, the subtitles rotate smoothly to the new position in front of the viewer; and
d) Appear in front, then fixed: subtitles appearing in front of users, and then fixed until they disappear (in this case, the subtitles do not follow the viewer if they look around).

They tested the different rendering modes with several clips (different durations: from 1 to 2 and a half minutes), and they concluded that "follow head immediately" (in our study, always-visible) was the most suitable, according to users' feedback. The reasons were that the implementation was easy to understand and subtitles easy to locate. Also, it gave viewers the freedom to navigate around the 360° environment without missing the subtitles. However, users complained about the blocking effect, that is, subtitles were blocking important parts of the image and were considered obstructive.

Following the above results, Rothe *et al.* (2018) also carried out a user study comparing two rendering modes: static subtitles (similar to always-visible in the present study) and dynamic subtitles (subtitles fixed in a dynamic position in the 360° sphere). They also tested speaker identification methods based on each mode and included name tags for each speaker. Participants did not state a clear preference for any of the methods. However, the results regarding key aspects of the VR experience (presence, sickness, and workload) favored the dynamic subtitles (Rothe *et al.*, 2018). In both studies, there is no clear solution and further testing is encouraged.

It is important to highlight that in each study a different terminology has been used. There is no consensus at this point on how to refer to the different types of rendering modes because this is an ongoing research and there is a lack of standardization. BBC terminology used "evenly spaced" and "follow head immediately" subtitle. Rothe *et al.* (2018) used "dynamic subtitles" and "static subtitles". In this study, a new terminology was tried to be defined in order to provide an intuitive solution to understand these concepts because the solutions for the three studies are slightly different. Therefore, the following terms that have been used in standardization forums (ISO, W3C, and MPEG) were suggested: "fixed-positioned" subtitles for those that are burnt-in in different fixed positions in the 360° sphere and "always-visible" subtitles for those that are anchored to the FoV and, therefore, always-visible for the viewer at a bottom centered position.

To shed some light on these open issues, we decided to test these two methods with longer and different content. We also decided to measure presence with the igroup presence questionnaire (IPQ)[1] to compare the impact of each method on viewers' presence if any. As explained in the methodology, IPQ is suitable for this type of content, and the measurements provided are accurate for our purpose. In other studies, presence questionnaire (Witmer and Singer, 1998) was used, such as in Rothe *et al.* (2018). This questionnaire includes a range of questions about interaction and control in the virtual world, which is not suitable for a 360° video with a passive observer. In the BBC study, only one question was asked regarding immersion "I felt immersed in the scene, like I was there" (Brown *et al.*, 2018). This study contributes to provide more information about the impact of the different subtitle rendering and presentation modes on presence.

To the best of our knowledge, no guiding methods for subtitles have been tested so far. This feature is especially important if subtitles are aimed at viewers with hearing impairments, or when the audio cannot be listened (e.g., noisy or public environments). When the audio cue is missing, support on how to locate the speakers and the main actions in the 360° scene is necessary. There are some studies, though, that tested different guiding methods for assisting focus in 360° videos, which are somehow related to, or have an impact on, guiding methods for subtitling. Some studies have tested different types of transitions and their impact on immersion and motion sickness. The preliminary results from the study by Men *et al.* (2017) concluded that the transition techniques being tested (Simple Cut Transition, Super Fast Transition, Fade Transition, and Vortex Transition) do not cause much sickness, contrary to what could be expected. The study carried out by Moghadam and Ragan (2017) concluded that each transition technique tested (Teleportation – involves an instant change in current FoV or rotation that is not perceived by the viewer; Animated Interpolation – smooth FoV transition from one state to another, which can be seen by the viewer; and Pulsed Interpolation – the pulsed view is faded in and out to different intermediate points from one state to another) had a different impact on the levels of presence and different techniques should be used depending on the desired effect. Lin *et al.* (2017) conducted an extensive study comparing two techniques to guide users to the focus of the 360° video: Auto Pilot – a method that takes the viewer directly to the intended target and Visual Guidance – a visual indicator that signals where the users should direct their view. The goal of this study was to establish which technique was better suited for the viewing experience when focus assistance is necessary. They concluded that both guiding methods were preferred by participants than no guiding method at all for focus assistance. They also argued that the specific content scenario and environment have an impact on which techniques are preferred by users. These insights are relevant for and support the importance of the research conducted in this work.

In the present study, the first goal was to gather participants' feedback (preferences and impact on presence) about two subtitle rendering modes (always-visible and fixed-positioned). In this regard, the limitations pointed out in previous studies (Brown *et al.*, 2018) were tried to be overcome, by using longer content (clips are longer than 2 min) and different genres. It was decided to use travel documentaries where the main goal was to have a look at the landscapes and listen to the narrator (voice-over), and thus become a suitable genre to test these research aspects. The following reasons support the decision on choosing this content genre: participants would have the freedom to look around without the main focus; no narrative complexities were introduced to avoid confusion; and there were no speakers on screen. Due to the fact that there were no speakers (only the narrator in some scenes), the variable (rendering modes) could be isolated, without introducing any guiding methods, tested in the second part of the experiment. The second goal of the study was to gather participants' feedback (preferences and presence) about two guiding methods (arrows and auto-positioning) to determine their acceptability by end users, in terms of presence, suitability, or preferences.

End-to-end platform for immersive accessibility

This research work has been conducted within the umbrella of the EU H2020 Immersive Accessibility (ImAc) project (October

[1]http://www.igroup.org/pq/ipq/index.php.

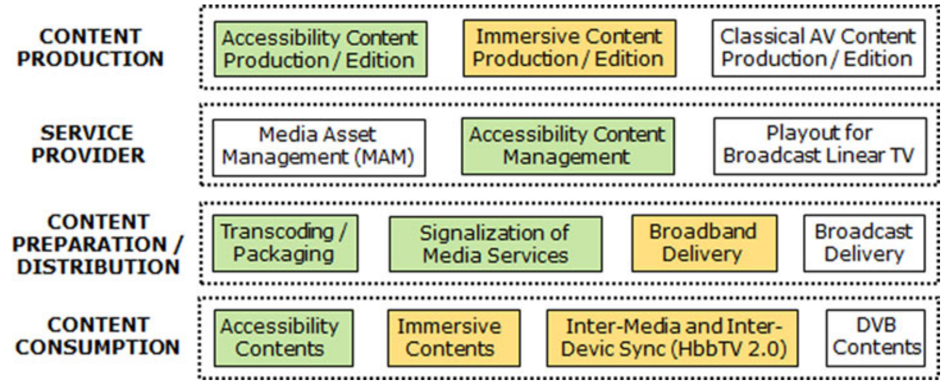


Fig. 1. Main layers or parts of the end-to-end platform.

2017–March 2020, http://www.imac-project.eu/). By following a user-centric methodology, ImAc is exploring how accessibility services (subtitling, audio description, and sign language interpreting) can be efficiently integrated within immersive media (360° video and spatial audio) while enabling different interaction modalities and personalization features. To achieve the targeted goals, ImAc is developing an end-to-end platform comprised of different parts where production, edition, management, preparation, delivery, and consumption of (immersive and accessibility) content take place. Figure 1 provides a high-level overview of the logical layers or main parts of the ImAc platform, which adhere to current-day media broadcast and delivery workflows. In this figure, green color is used to identify components being developed within the ImAc project, orange color is used to identify components that are relevant for ImAc but that have been developed in other related projects, and white color is used for components that exist in typical broadcast workflows but that are either not part of or not essential for ImAc. Next, an overview of each one of the platform parts is provided to better understand the context – and potential impact – of this work.

Content production

The content production part of the platform includes a set of (web-based) tools for the production and edition of access services (including subtitles, audio description, and sign language interpreting), and their integration with immersive media content. The subtitles production/edition tool enables the creation of subtitles for 360° videos. Unlike existing editors that mainly allow the production of subtitles frames with specific timing attributes (i.e., start and end times), the ImAc editor provides a set of additional features targeted at contributing to better accessibility (and engagement):[2]

- It allows setting different styling effects (e.g., colors and font) for different speakers.
- It allows indicating spatial attributes to set the region of the 360° area to which the subtitle frames refer. The spatial information consists of the latitude and longitude angles (although only the latitude ones are considered in this work). This is relevant, as the associated action(s)/speaker(s) can be placed in different parts of the 360° area and can even dynamically move. However, it is possible to indicate that no spatial information is linked to specific subtitle frames (e.g., for off-camera commentaries).
- It allows specifying two options for subtitles rendering (in flat format). The first option consists of using the 360° sphere as the rendering reference (see Fig. 2). This is called fixed-positioned, as subtitles are attached (i.e., statically placed) in a fixed region of the video sphere. Using this mode, subtitles will not be visible if the user's FoV is outside the subtitle's region. To overcome this, subtitles can be presented evenly spaced every 120° in the 360° sphere, ensuring at least one of them will be visible at any time, regardless of the current FoV. The second option consists of using the current FoV as a rendering reference (see Fig. 3). This is called always-visible, as subtitles are attached to the FoV, and thus positioned in the center at any moment, regardless of where the user is looking at.
- When using always-visible subtitles, it allows specifying different guiding methods to assist the users in finding the action(s)/speaker(s) associated with the subtitles in the 360° area. A first option consists of adding arrows to the left/right of the subtitle frames, indicating the direction toward the associated audiovisual elements in the 360° area (see Fig. 4). When this position is inside the user's FoV, the arrows are hidden. A second option consists of automatically adjusting the FoV based on the position of the associated action(s)/speaker(s). This auto-positioning mechanism is applied to every subtitle frame with spatial information if explicitly indicated in the editor.

All these rendering and presentation features are signalized as metadata extensions to the Internet Media Subtitles and Captions (IMSC) subtitles format, being used in ImAc. IMSC is a subset of the Timed Text Markup Language (TTML) for distribution of subtitles, which is drawing the attention of, and being adopted by, many standardization bodies.

Service provider

This part of the platform includes components for Media Asset Management (MAM), linking of additional content to main TV programs, and scheduling playout. In the context of ImAc, it additionally includes the Accessibility Content Manager (ACM), which is the component where the immersive content is uploaded, the creation of accessibility content is managed, and the preparation of content for their delivery is triggered.

Content preparation and distribution

This part of the platform includes components for preparing the content for their appropriate distribution via various technologies.

[2]Further features are planned to be incorporated in the near future, such as the use of visual icons augmenting the textual information.

Fig. 2. Fixed-positioned: subtitles attached to the video sphere.

Fig. 3. Always-visible: subtitles attached to the camera or FoV.

These components are in charge of encoding the content in multiple qualities (to adapt to the target consumption devices and available bandwidth), segmenting the content for an efficient quality adaption and re-transmission (e.g., in case of packet loss), signaling their availability, and describing them. The project focuses on the delivery of the content via broadband content delivery networks (CDNs), by making use of dynamic adaptive streaming over HTTP (DASH) as the media delivery technology. However, it is also envisioned to make use of DASH in the coordination of digital video broadcasting (DVB) services, as supported by the worldwide adopted hybrid broadcast-broadband TV (HbbTV) standard. This will enable augmenting traditional TV services with more interactive and personalized multiscreen experiences, enriching the traditional TV content with extra immersive and accessibility content presented on companion devices, like smartphones or even HMDs.

In this context, ImAc is exploring the specification of standard-compliant extensions to media formats and technologies [e.g., within the framework of Moving Picture Experts Group (MPEG)] to accommodate the envisioned immersive accessibility services and features.

Content consumption

The ImAc player is a core component of the ImAc platform, as it is the interface through which end users will consume the available immersive and accessibility content in an interactive and personalized manner. The design and implementation of the player

Fig. 4. Subtitles with arrows as a guiding mechanism.

face many challenges due to several facts, such as the nature and combination of media content to be consumed, the heterogeneity in terms of access networks and consumer devices to be employed, and the diverse needs and/or preferences of the target end users.

The player has been developed by exclusively relying on standard(-complaint) web-based technologies and components. This will guarantee cross-network, cross-platform, and cross-browser support, and eliminate the need for any installation or software updates. The use of web-based components also facilitates the embedding of the player within the web services of broadcasters and/or service providers, ensuring interoperability and scalability. Figure 5 illustrates the main layers and modules and libraries that make up the player, together with the relationships and interactions between them. All these components are mainly targeted at enabling the presentation of content, to enable different interaction features, and to dynamically set the available personalization options.

Three main layers are in charge of the presentation of content in the player. These include:

- The Immersive Layer: it is responsible for the presentation of both traditional and immersive audio-visual formats. For immersive media, it includes 360° videos and spatial audio (Ambisonics).
- The Accessibility Layer: it is responsible for the presentation of accessibility content considered in the project, namely audio and text subtitles; audio description; and sign language video.
- The Assistive Layer: it includes relevant features to assist the users for a more effective usage of the player. Some examples are voice control (recognition and synthesis) and augmentation/zooming capabilities.

Likewise, the Media Synchronization Layer is in charge of ensuring a synchronized consumption of content, both within each device (i.e., local inter-media synchronization) and across devices in a multiscreen scenario (i.e., inter-device synchronization).

In addition, two main modules in the ImAc player can be highlighted:

- The User Interface (UI): it is the module through which users enable the presentation of content, interact with the player and set the available personalization features. Indeed, two UIs have been designed and implemented: (1) a traditional UI but adapted to 360° environments (see Fig. 6) and (2) an enhanced accessibility (also known as low-sighted) UI, which occupies most part of the screen (see Fig. 7).
- The Session Manager: it is the module responsible for interpreting and selecting the list of available assets from the content provider, keeping an updated status about the content being presented and the active devices in multiscreen scenarios, and keeping track of the available personalization options together with the current settings.

Personalized presentation of accessibility services

The player provides support for a personalized presentation of access services, including subtitles, audio description, and sign language video (see UIs in Figs. 6, 7). Most interestingly for this work, the player allows dynamically setting the following personalization features for presentation of subtitles:

- language selection;
- three sizes for the subtitle font (large, medium, and small);
- position (top and bottom);
- three sizes for the safe area or the comfortable FoV where to place graphical elements on the screen. Although the screen size and resolution of the device in use is automatically detected, users can have different preferences regarding this aspect;
- background (semi-transparent box for the subtitles frame, outline);
- normal versus easy-to-read subtitles (i.e., more simple and shorter subtitles); and
- guiding methods: (1) none; (2) arrows indicate where the associated speaker is; and (3) auto-positioning: the FoV is

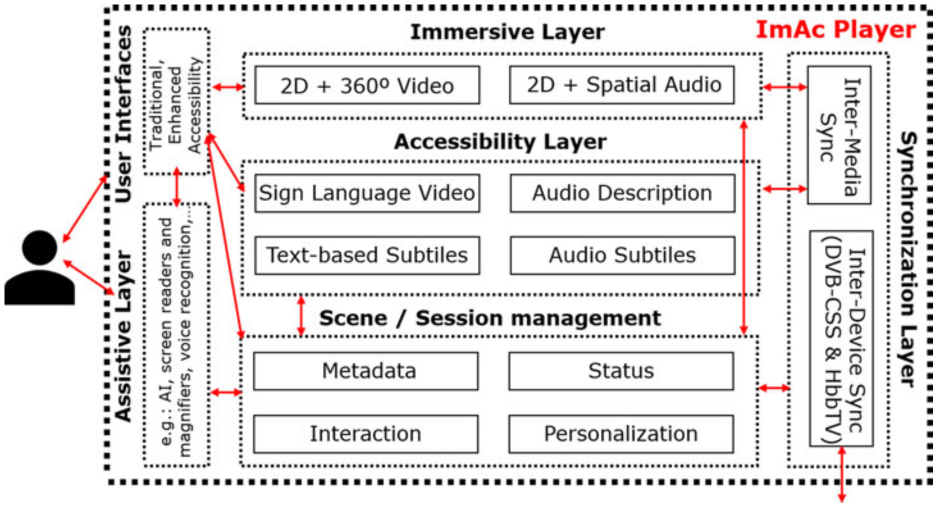


Fig. 5. Layers and modules making up the ImAc player.

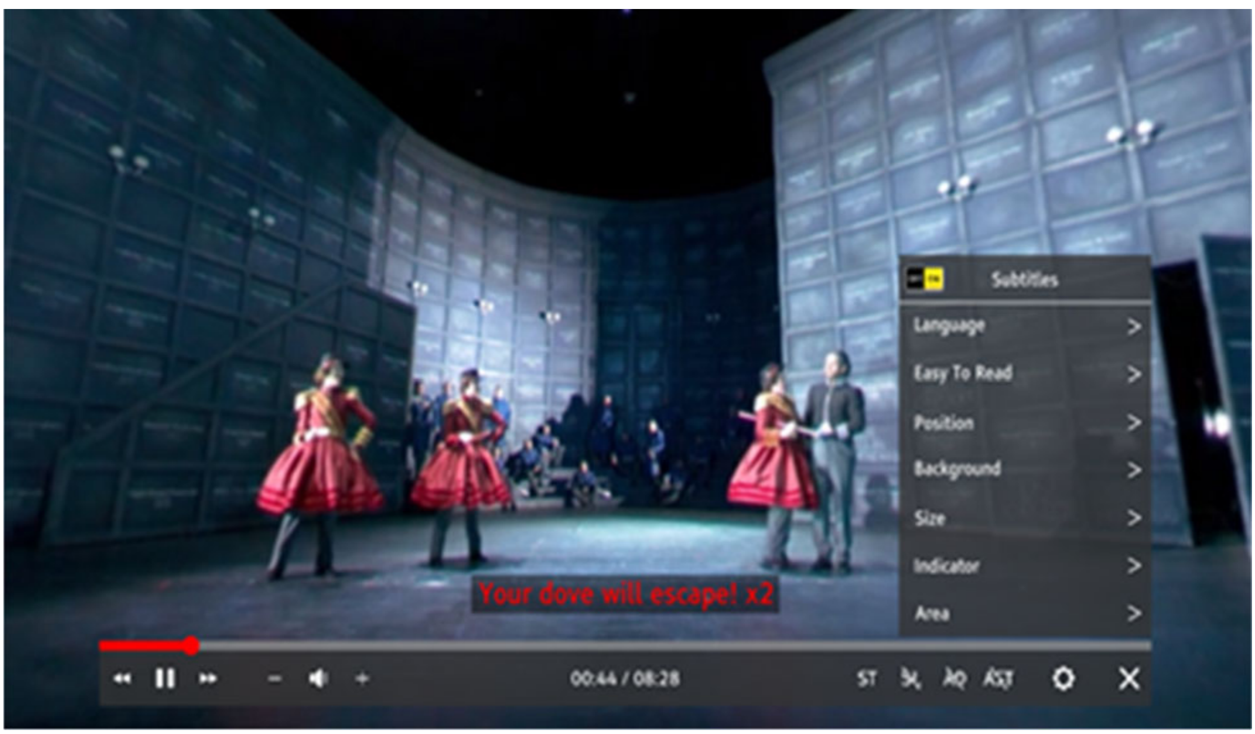


Fig. 6. Screenshot of the traditional UI of the player.

Fig. 7. Screenshots of the enhanced accessibility (or low-sighted) UI of the player.

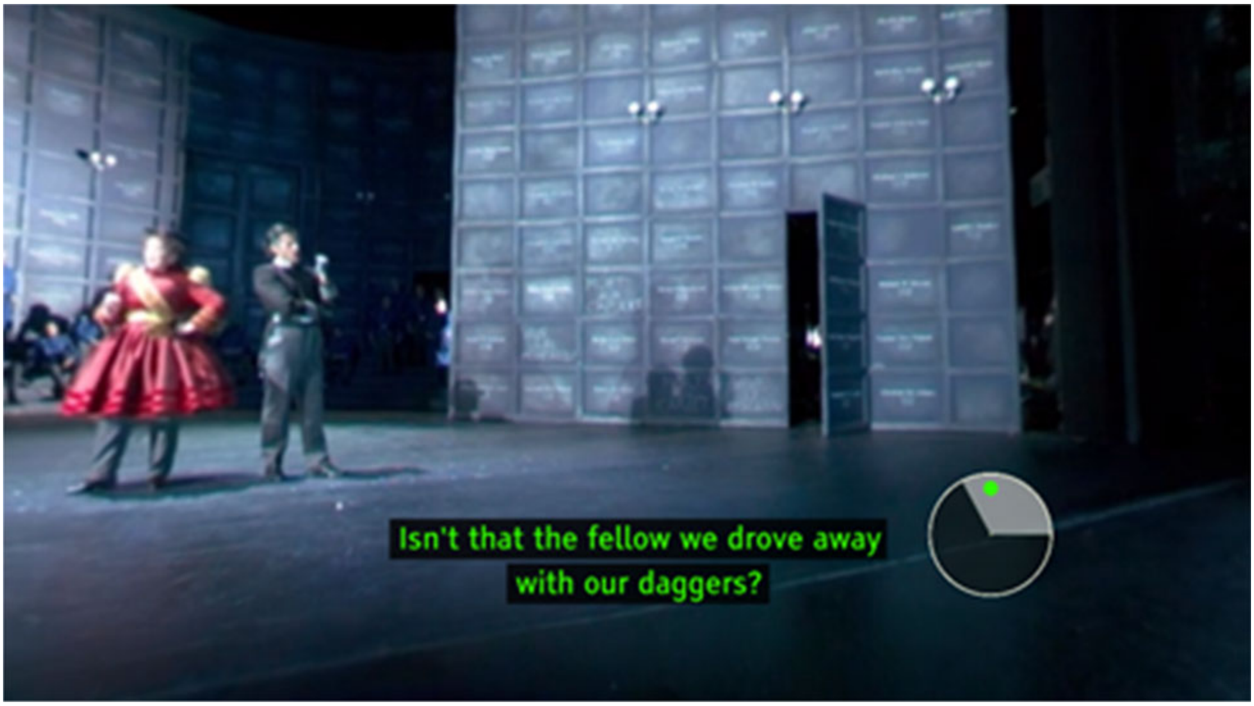


Fig. 8. Use of a dynamic radar as a guiding method.

automatically adjusted based on the location of the speaker. An additional method is available, which consists of using a dynamic radar to indicate where the associated speaker or main action is (see Fig. 8). However, this method has not been tested in this work, as pretests have indicated a preference toward the arrows.

Apart from the user-level personalization features, the rendering mode for subtitles and different styling effects (color and font) for each speaker can be set during the production/edition phase (at the content production part).

Evaluation

An experiment to test two different aspects of subtitles in 360° content (rendering and guiding methods) was conducted. The goal of this experiment was to clarify which options are preferred by users, as well as which ones result in higher levels of presence. This section describes the selected and created stimuli for conducting the tests, the evaluation scenario and setup, the followed evaluation methodology and then presents the obtained results.

Evaluation stimuli

An acclimation clip was introduced at the beginning of the test so that participants could become familiar with the HMD and the type of content, assuming that most participants did not have an extensive experience with the use of HMD and VR experiences. This was later confirmed by the replies to the demographic questionnaire. All clips included sound (voice-over in English) because it was considered that sound is an important part of the immersive experience and presence was being measured as part of the test. Subtitles for the deaf and hard-of-hearing were produced in Spanish, a language spoken and understood by all participants.

For the first condition (rendering modes), two videos from the series The Holy Land created by Ryot[3] jointly with Jaunt VR were used. Creators gave their permission to use these videos in the study. Specifically, the episodes 4 (duration of 4 min and 13 s)[4] and 5 (duration of 4 min and 58 s)[5] were chosen. The clips are travel documentaries depicting Israel and surrounding territories. Different locations and landscapes are featured. In the clips, there is only one speaker and most of the script is voice-over (narrator), except for some scenes where the hostess can be seen. The videos were considered suitable for testing the subtitle rendering modes because viewers could concentrate on reading subtitles and watching the scenes without the added effort of having to look for the speakers or any other narrative complexities.

For the second condition (guiding methods), the clip An American Classic: Guelaguetza[6], also created by Ryot, was used. In this case, the video was split into two parts, in order to have two comparable clips. The total duration of the clip is 7 min and 58 s (first part from 00:00 to 03:46 and the second part from 03:46 to 07:16 – credits start). This short documentary narrates the story of a family from Oaxaca that decided to migrate to Los Angeles and opened a restaurant there. In the video, the two generations of owners (mother and daughter) explain their experiences and what the restaurant and their food mean to them. The clip combines scenes with different locations and a voice-off narration with scenes were Bricia (daughter) and Maria (mother) appear explaining their experiences. This video was suitable for the test because it includes different people in

[3]https://www.jauntvr.com/lobby/ryot.
[4]https://www.jauntvr.com/title/b4f85188a2.
[5]https://www.jauntvr.com/title/fb1051a266.
[6]https://www.youtube.com/watch?v=zneKYGQgabk.

different locations. Therefore, participants had to look for the speakers. Moreover, the speakers in the video were easy to find (they are mostly located in the same area, standing or sitting), which was also desirable for the test to avoid confusion among viewers, especially for those ones not being familiar with any of the guiding methods or with VR technology in some cases.

Evaluation setup

The evaluations were conducted in a local scenario with of a PC with an Apache web server (no high computational resources are required) to host the player resources and the media assets (360° video and subtitles), a conventional 802.11b Wi-Fi network and a standalone VR Oculus GO (32GB) as a consumption device. The Oculus GO accessed the player via its Wi-Fi connection and by typing the target URL pointing to the server resources. Note that the web server and clients could have been placed in remote locations and that other types of consumption devices, and other HMDs, could have been used.

The 360° videos were converted into the DASH format, being encoded in multiple qualities (with bit rates ranging from 8 to 2 Mbps) and segmented in chunks of a duration of 3 s. This allows an efficient quality switching adaptation, based on the network and consumption devices conditions. The subtitle files were delivered independently to the video segments, but they were signalized as part of the video metadata files. An overview of the evaluation scenario and setup can be seen in Figure 9.

Evaluation methodology

A within-subject design was used to test the different subtitle presentation conditions. Each participant watched four clips (plus the acclimation video), being each of them presented with a different variable (fixed-positioned, always-visible, arrows, and auto-positioning). The four clips and four conditions were randomized using a Latin square (see Table 1), to avoid the order of presentation affecting the results. The Holy Land clips were independent and, therefore, could be watched in a random order without affecting the narrative. The clip An American Classic: Guelaguetza, however, was always shown in chronological order, otherwise the participants would have not been able to understand the story.

The experiment was organized in one session divided into two parts: Part 1 – rendering modes and Part 2 – guiding methods. The experiment was focused on assessing users' preferences and presence. One of the main goals of immersive content, such as 360° videos, is to create an immersive experience. Therefore, it was paramount to design subtitles that would enhance the experience making it more accessible rather than disrupting it. Likewise, an additional goal of the test was to gather feedback from the users. This would allow deriving potential requirements for improving the provided functionalities or even incorporating additional ones, thus following the user-centric methodology being used in ImAc. To gather this feedback, questionnaires were used.

For presence, a translation into Spanish of the IPQ questionnaire was used. After a review of different presence questionnaires, such as Slater-Usoh-Steed presence questionnaire (Slater and Usoh, 1993), presence questionnaire (Witmer and Singer, 1998), or ICT-SOPI (Lessiter *et al.*, 2001), IPQ was chosen for different reasons. First, it includes questions from different questionnaires and it specifically differentiates between presence, spatial presence, involvement and realness. The questionnaire has been validated in different virtual environments (users of VR or CAVE-like systems, desktop VR, and players of 3D games). Also, unlike other questionnaires, such as the presence questionnaire by Witmer and Singer (1998), the questions in IPQ do not involve interaction with the virtual world. This was important because the 360° clips that were chosen for the test are not interactive.

For preferences, an ad-hoc questionnaire in Spanish for this test was created for each part (rendering and guiding methods). The questionnaires included closed questions to assess which system users preferred and questions related to subtitles' blocking or distracting effects. Also, open questions were used to gather feedback about the reasons to choose one method over the other, and 7-point Likert-scale questions were added to determine how easy it was to find or read subtitles, as well as to find the speaker in the video.

After watching each clip, participants were asked to fill in the IPQ questionnaire so that the level of presence could be later compared between the two options. The impact of the different subtitle strategies on presence, if any, could then be measured and reported. After each part, participants were also asked to fill in the preference questionnaires so that they could report on their experience with both options for rendering and guiding methods.

Participants

Eight participants took part in the test (three female and five male), with ages ranging from 26 to 59 (mean = 35.5; standard deviation = 13.18). Two participants were deaf. Our aim was to include different profiles of subtitle users to gather relevant feedback in this preliminary study. To that end, users from different ages and hearing abilities were included. As explained before, subtitles are not only beneficial for the deaf audience but also for hearing audience with different needs (non-native speaker and noisy environments). This is due to the fact of the wide applicability of subtitles, as discussed in the "Introduction" section. Despite the fact that not a high number of users participated in the study, a bigger sample will be used in future work in order to support the significance of the obtained results, even when considering different users' profiles. The followed methodology has been chosen for achieving continuity of results.

Evaluation results

The results from the different questionnaires are reported in this subsection.

Demographic information

Some more demographic information about participants was gathered. Five participants had a university education, two had professional training, and one had primary education. Two participants were familiar with VR content (one participant stated to use VR once a week and another participant once a month). Five participants were interested in VR content and three were neutral. Three participants owned VR equipment: one had cardboard, another had a PlayStation VR, and the last one had a Google Cardboard and a PlayStation VR. Two participants claimed that they never use subtitles, four participants claimed to use subtitles sometimes (depending on the content, the language and the context – noisy room, other people watching the content at the same time, etc.) and two participants always used

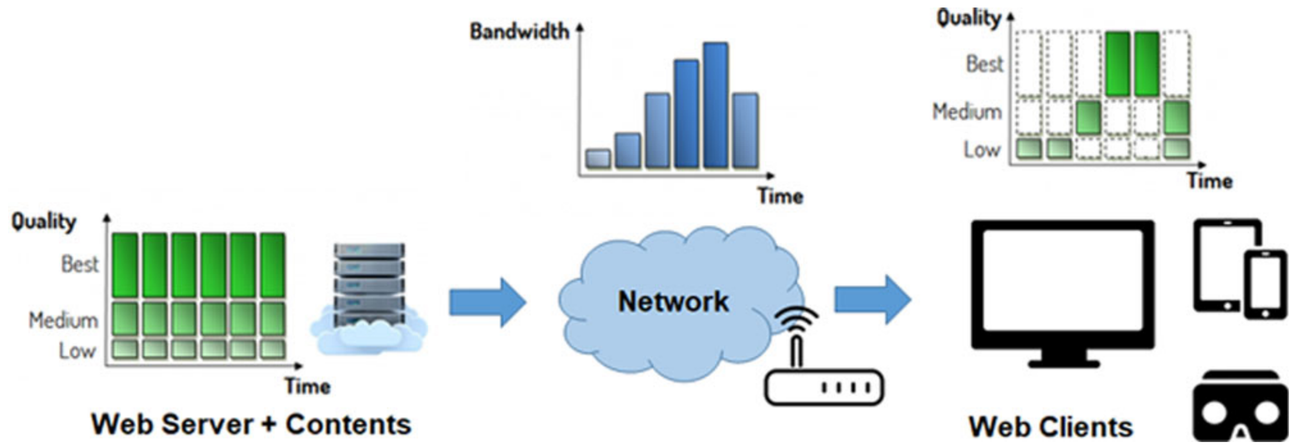


Fig. 9. Overview of the evaluation scenario and setup.

Table 1. Latin square used in the tests

Participant	Rendering mode	Rendering mode	Guiding mechanism	Guiding mechanism
PPilot01	A1 Holy Land 1-fixed-positioned	B2 Holy Land 2-always-visible	C1 Guelaguetza 1-arrow	D2 Guelaguetza 2-auto-positioning
PPilot02	A2 Holy Land 1-always-visible	B1 Holy Land 2-fixed-positioned	C2 Guelaguetza 1-auto-positioning	D1 Guelaguetza 2-arrow
PPilot03	B1 Holy Land 2-fixed-positioned	A2 Holy Land 1-always-visible	C1 Guelaguetza 1-arrow	D2 Guelaguetza 2-auto-positioning
PPilot04	B2 Holy Land 2-always-visible	A1 Holy Land 1-fixed-positioned	C2 Guelaguetza 1-auto-positioning	D1 Guelaguetza 2-arrow
PPilot05	A1 Holy Land 1-fixed-positioned	B2 Holy Land 2-always-visible	C1 Guelaguetza 1-arrow	D2 Guelaguetza 2-auto-positioning
PPilot06	A2 Holy Land 1-always-visible	B1 Holy Land 2-fixed-positioned	C2 Guelaguetza 1-auto-positioning	D1 Guelaguetza 2-arrow
PPilot07	B1 Holy Land 2-fixed-positioned	A2 Holy Land 1-always-visible	C1 Guelaguetza 1-arrow	D2 Guelaguetza 2-auto-positioning
PPilot08	B2 Holy Land 2-always-visible	A1 Holy Land 1-fixed-positioned	C2 Guelaguetza 1-auto-positioning	D1 Guelaguetza 2-arrow

subtitles. When asked about the reasons to use subtitles, one participant said to learn languages, four said that they used them because subtitles helped them to understand the content, one participant claimed that subtitles are the only way to access the dialogs and two said that they never use subtitles.

IPQ and preferences

Participants' self-assessed experiences were analyzed based on two types of questionnaires: IPQ to measure and compare the levels of presence and ad-hoc questionnaires to gather feedback about participants' preferences regarding the considered subtitle presentation modes. Preferences results have been analyzed using a Wilcoxon Signed Rank test. The results for the IPQ test aimed at detecting differences in levels of presence between always-visible and fixed-positioned subtitles and between subtitles with arrows and auto-positioning, and the existence of significant differences between the tested conditions has been also analyzed using a Wilcoxon Signed Rank test (with a threshold value of 0.05). The IPQ is divided into four main blocks: presence, spatial presence, involvement and realness, and the results are reported below based on that classification.

Always-visible versus fixed-positioned

All participants preferred the always-visible subtitle rendering mode. According to the participants, the main reasons for having chosen this option is that with the always-visible subtitles they had more freedom to look around without missing the subtitle content and the video scenes. When asked about how easy it was to find the always-visible subtitles based on a 7-point Likert scale (7 being the easiest and 1 being the most difficult), six participants (75%) replied 7, one (12.5%) replied 6, and one (12.5%) replied 5. When asked the same question about fixed-positioned subtitles, three participants (37.5%) replied 2, two (25%) replied 3, two (25%) replied 4, and one (12.5%) replied 5. Then, according to these results, always-visible subtitles (mean = 5.78) were considered easier to find than fixed-positioned subtitles (mean = 3.12). This difference is statistically significant ($Z = -2.555$, $p = 0.011$, ties = 0). When asked about how easy it was to read always-visible subtitles based on a 7-point Likert scale (7 being the easiest and 1 being the most difficult), three participants (37.5%) replied 6, two (25%) replied 2, one (12.5%) replied 7, one (12.5%) replied 5, and one (12.5%) replied 3. When asked the same question about fixed-positioned subtitles (7 being the easiest and 1 being the most difficult), two (25%) replied 6, two (25%) replied 5, two (25%) replied 2, one (12.5%) replied 7, and one (12.5%) replied 3. Therefore, according to these results, always-visible subtitles (mean = 4.62) were considered slightly easier to read than fixed-positioned subtitles (mean = 4.5). However, this difference is not statistically significant ($Z = -0.086$, $p = 0.031$, ties = 1). When participants were asked whether subtitles were obstructing important parts of the image, five participants (62.5%) replied "no" and three (37.5%) replied "yes" for always-visible subtitles, and seven participants (87.5%) replied "no" and one (12.5%) replied "yes" for fixed-positioned subtitles.

The comparison of results from IPQ between the always-visible and fixed-positioned are as follows: For the presence scale, the test indicated that the difference between results is not statistically significant ($Z = -1.000$, $p = 0.317$, ties = 7). For the spatial presence scale, the test indicated that the difference between results is not statistically significant ($Z = -1.103$, $p = 0.270$, ties = 1). However, for the realness scale ($Z = -2.060$, $p = 0.039$, ties = 3) and the involvement scale ($Z = -2.384$, $p = 0.017$, ties = 1), the test reported that the difference between results is statistically significant. This means that the fixed-positioned subtitles had a negative impact on the involvement of participants and their perception of realness. According to their comments in the open questions, this could be because they felt less free to explore the 360° scene and claimed to have missed parts of the subtitles content. Moreover, as reported above, participants found more difficult to find subtitles in this mode. Therefore, this extra effort could have caused a negative impact on involvement and realness.

Arrows versus auto-positioning

Seven participants (87.5%) preferred the arrows over the auto-positioning method. Participants who favored the arrows argued that this guiding method is more intuitive and comfortable. Three participants suggested that the arrow guiding mechanism should also include indications for the vertical axis (up, down), not only for the horizontal one (left, right). The participant who preferred the auto-positioning considered that it was more comfortable because there was no need to move or look for the speaker. One participant also argued that she would like to have a focus assistance technique not only for speakers but also for the main action in the videos. For example, if a specific event is happening in a part of the video (even if no one is speaking), she considered that it would be useful to have an indicator to avoid getting lost. When asked about how easy it was to find the speaker with the arrow guiding method, based on a 7-point Likert scale (7 being the easiest and 1 being the most difficult), three participants (37.5%) replied 6, two (25%) replied 7, two (25%) replied 4, and one (12.5%) replied 5. When asked the same question about the auto-positioning (7 being the easiest and 1 being the most difficult), three participants (37.5%) replied 7, three (37.5%) replied 1, one (12.5%) replied 6, and one (12.5%) replied 3. The different results in the latter could be because some participants reported feeling dizzy and disoriented with the auto-positioning system and others did not have the same experience. According to the results, arrows (mean = 5.62) were considered more effective to find the speaker than auto-positioning (mean = 4.12). However, the difference is not statistically significant ($Z = -1.476$, $p = 0.140$, ties = 2). When asked whether the guiding methods distracted participants from the story, seven participants (87.5%) replied "no" and one (12.5%) replied "yes" for the arrows, and five participants (62.5%) replied "yes" and three (37.5%) replied "no" for the auto-positioning.

The comparison of results from IPQ between arrows and auto-positioning methods are as follows: For the spatial presence scale, the test indicated that the difference between results is not statistically significant ($Z = -0.256$, $p = 0.798$, ties = 1). For the involvement scale, the test indicated that the difference between results is not statistically significant ($Z = -0.412$, $p = 0.680$, ties = 3). For the realness scale, the test indicated that the difference between results is not statistically significant ($Z = -0.850$, $p = 0.395$, ties = 2). However, for the presence scale, the test reported that the difference between results is statistically significant ($Z = -2.000$, $p = 0.046$, ties = 4). This means that the auto-positioning method had a negative impact on the presence in the virtual world. According to comments in the open questions, some participants stated that auto-positioning caused dizziness, and had an impact on immersion, and resulted in confusion for some users.

Conclusions and future work

This article has investigated the suitability of different rendering modes and guiding methods for subtitles in 360° videos. The considered options have been integrated in an end-to-end platform being developed in the ImAc project. An overview is such a platform has been provided to better understand the context of this work and its potential impact. According to the obtained results, it can be concluded that always-visible subtitles are more appropriate than fixed-positioned subtitles. These findings are in line with the ones from the study carried out by BBC (Brown *et al.*, 2018), but we have tried to overcome some limitations of that work (such as the duration of the content). Even if the content is longer, always-visible subtitles seem to be the most suitable of the rendering modes explored so far. Moreover, in our case, participants did not complain about the blocking effect of the subtitles, as it happened in the BBC study. This could be due to the fact that we did not use a background box in the subtitles, and therefore, they were less intrusive. As explained in the "End-to-end platform for immersive accessibility" section, the use of a background box or an outline can be dynamically set in the developed 360° player. Also, the results from the IPQ have shed some light on the potential impact of rendering modes on presence levels reported by participants. Fixed-positioned subtitles might have a negative impact on presence, while always-visible subtitles seemed to be more adequate in that sense.

Regarding the two analyzed guiding methods, it can be concluded that the use of arrows is more intuitive and effective than auto-positioning. Even if previous studies argued that auto-positioning methods are accepted by users (Lin *et al.*, 2017), in our study it can be concluded that auto-positioning can provoke dizziness (as reported by participants) and might have a negative impact on presence, at least for the considered content types.

The scope of this preliminary study was to test several subtitle modes with a limited number of participants. Including diverse profiles was sought to clarify the different needs of subtitle users. The selected content might have had an impact on preferences and presence results that is not directly related to the different subtitle modes. For the rendering modes options, two travel documentaries were used. In this type of content, the aim is to look around and, then, it is desirable to have the freedom to move. However, if the video features a conversation between two people in a bar, perhaps the fixed-positioned solution would be more accepted. A similar content (two people talking and sitting next to each other) was tested in the study by Rothe *et al.* (2018) and the results favored fixed-positioned subtitles. Also, some participants argued that the videos were not first-person and, therefore, were less immersive. Others thought that the quality, scales and type of scenes also had a negative impact on immersion. For the guiding methods, we used a content where the speakers were mainly in a fixed position and did not rapidly move. Perhaps, if the content includes speakers moving fast, an improved auto-positioning system could assist viewers keeping the focus of the video. These hypotheses are worth testing in future studies. Likewise, a wider sample of participants, with different profiles, will be considered to test these conditions, maybe with some variants, in the near future.

Different ideas for future work are planned. Regarding rendering modes, it could be interesting to compare the appropriateness/effectiveness of always-visible and fixed-positioned subtitles depending on the type of content (static scenes vs. action-based scenes), by analyzing whether the type of content has a direct impact on the viewers' preferences and levels of presence. Combining the two rendering modes in a content with different types of scenes (static and action-based) and measure the reaction of participants is also an option worth exploring. Regarding guiding methods, auto-positioning strategies could be refined to reduce the VR sickness effect and test it with other types of content (action-based). Likewise, the use of a dynamic and intuitive radar (as introduced in the "End-to-end platform for immersive accessibility" section) could be explored in future tests. In addition, the Assistive Layer of the player (see Fig. 5) will be further developed to integrate Artificial Intelligence and signal processing techniques, as well as automatic adaptation strategies, with the final goal of maximizing the perceived QoE and accessibility.

Finally, the feedback from the participants will be considered to explore the suitability of refining the adopted solutions and/or adopting extra alternatives (e.g., including guiding methods not only for speakers but also for main actions).

Acknowledgments. The authors thank ImAc partners, for their technical contributions in preparing the end-to-end platform, from production to consumption of immersive and accessibility content; Irene Tor-Carroggio for her help analyzing the results in this research; Ryot for the content; and all users for participating in this research.

Financial support. This work has been funded by the European Union's Horizon 2020 program, under agreement no. 761974 (ImAc project). Belén Agulló is a member of TransMedia Catalonia, an SGR research group funded by "Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya" (2017SGR113). This article is part of Belén Agulló'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 the Universitat Autònoma de Barcelona. The work by Mario Montagud has been partially funded by the Spanish Ministry of Economy and Competitiveness (MINECO), in the project with reference IJCI-2017-34611.

References

Agulló B and Matamala A (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* **32**, 217–235

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

Armstrong M, Brown A, Crabb M, Hughes C and Sandford J (2015) Understanding the diverse needs of subtitle users in a rapidly evolving media landscape. *IBC 2015*, Amsterdam, The Netherlands, September 2015.

Aylett R and Louchart S (2003) Towards a narrative theory of virtual reality. *Virtual Reality* **7**, 2–9.

Brown A, Turner J, Patterson J, Schmitz A, Armstrong M and Glancy M (2017) Subtitles in 360° video. Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video. ACM: New York.

Brown A, Turner J, Patterson J, Schmitz A, Armstrong M and Glancy M (2018) Exploring subtitle behaviour for 360° video. *White Paper WHP 330*. BBC: Manchester.

Cummings JJ and Bailenson JN (2016) How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology* **19**, 272–309.

De la Peña N, Weil P, Llobera J, Giannopoulos E, Pomés A, Spanlang B, Friedman D, Sanchez-Vives MV and Slater M (2010) Immersive journalism: immersive virtual reality for the first-person experience of news. *Presence: Teleoperators and Virtual Environments* **19**, 291–301.

Dooley K (2017) Storytelling with virtual reality in 360-degrees: a new screen grammar. *Studies in Australasian Cinema* **11**, 161–171.

European Broadcasting Union (EBU) (2017) *Virtual reality: how are public broadcasters using it?* Available at: https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it (consulted on 28 November 2018).

Gödde M, Gabler F, Siegmund D and Braun A (2018) Cinematic narration in VR – Rethinking Film conventions for 360 degrees. In Chen J and Fragomeni G (eds), *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry, VAMR 2018, Vol. 10910. Lecture Notes in Computer Science*, Cham: Springer.

Greco GM (2016) On accessibility as a human right, with an application to media accessibility. In Matamala A and Orero P (eds), *Researching Audio Description*. London: Palgrave Macmillan.

Jones S (2017) Disrupting the narrative: immersive journalism in virtual reality. *Journal of Media Practice* **18**, 171–185.

Lessiter J, Freeman J, Keogh E and Davidoff J (2001) A cross-media presence questionnaire: the ITC-sense of presence inventory. *Presence: Teleoperators, and Virtual Environments* **10**, 282–297.

Lin Y, Chang Y, Hu H, Cheng H, Huang C and Sun M (2017) Tell me where to look: investigating ways for assisting focus in 360° video. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. New York: ACM, pp. 2535–2545.

Mateer J (2017) Directing for cinematic virtual reality: how the traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice* **18**, 14–25.

Men L, Bryan-Kinns N, Hassard AS and Ma Z (2017) The impact of transitions on user experience in virtual reality. *2017 IEEE Virtual Reality (VR)*, Los Angeles, pp. 285–286.

Moghadam KR and Ragan ED (2017) Towards understanding scene transition techniques in immersive 360 movies and cinematic experiences. *2017 IEEE Virtual Reality (VR)*, Los Angeles, pp. 375–376.

Patel S (2016) 85 percent of Facebook video is watched without sound. In *DigidayUK*. Retrieved from: https://digiday.com/media/silent-world-facebook-video/.

Perego E, Del Missier F and Bottiroli S (2015) Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects. *Perspectives: Studies in Translation Theory and Practice* **23**, 1–21.

Remael A, Orero P and Carroll M (eds) (2014) *Audiovisual Translation and Media Accessibility at the Crossroads*. Amsterdam/New York: Rodopi.

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

Romero-Fresco P (ed.) (2015) *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe*. Bern: Peter Lang.

Rothe S, Heinrich H and Mathias A (2017) Diegetic cues for guiding the viewer in cinematic virtual reality. *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. New York: ACM.

Rothe S, Kim T and Hussmann H (2018) Dynamic subtitles in cinematic virtual reality. *Proceedings of the 15th European Interactive TV Conference (ACM TVX 2018)*. New York: ACM.

Sheikh A, Brown A, Watson Z and Evans M (2017) Directing attention in 360-degree video. *IBC 2016*, Amsterdam, 9–13 September.

Slater M and Usoh M (1993) Representations systems, perceptual position, and presence in immersive virtual environments. *Presence* **2**, 221–233.

Wang G, Gu W and Suh A (2018) The effects of 360-degree VR videos on audience engagement: evidence from the New York Times. In Nah FH and Xiao B (eds), *HCI in Business, Government, and Organizations, Vol. 10923. HCIBGO 2018. Lecture Notes in Computer Science*. Cham: Springer.

Witmer BG and Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. *Presence Teleoperators & Virtual Environments* **7**, 225–240.

Belén Agulló is a predoctoral researcher in the Department of Translation, Interpreting and Eastern Asian Studies at the Universitat Autònoma de Barcelona, where she works on the Horizon 2020-funded project Immersive Accessibility (ImAc). Her PhD focus is subtitling for the deaf and hard-of-hearing in immersive media. Her research interests include game localization, audiovisual translation, and media accessibility.

Dr. Mario Montagud is a senior researcher at i2CAT Foundation (Barcelona, Spain) and a part-time professor at the University of Valencia (Spain). He received a PhD degree in Telecommunications (Cum Laude Distinction) at the Polytechnic University of Valencia (UPV, Spain) in 2015. His topics of interest include Computer Networks, Interactive and Immersive Media, Synchronization and QoE. Mario is (co-) author of over 70 scientific and teaching publications and has contributed to standardization within the Internet Engineering Task Force (IETF). He is a member of the Editorial Board of international journals and has been a member of the Organization Committee of the many international workshops and conferences. He is also lead editor of "MediaSync: Handbook on Multimedia Synchronization" (Springer, 2018) and Communication Ambassador of ACM SIGCHI (Special Interest Group in Human–Computer Interaction). He is currently involved in three EU H2020 projects, being WP leader in two of them: VR-Together, ImAc, and 5G-Picture.

Mr. Isaac Fraile is a network engineer by the Polytechnic University of Catalonia (UPC, Spain) and is currently studying an MSc degree in Multimedia Applications at the Universitat Oberta de Catalunya (UOC, Spain). He works as a Project Engineer in the Media Unit of i2CAT Foundation. His work is focused on the design and implementation of immersive media platforms, paying special attention to delivery and synchronization techniques, and to web-based components. Previously, he has worked for a year and a half for the CCMA, the public regional broadcaster in Catalonia. He has participated in three EU projects: TV-Ring, ImmersiaTV, and ImAc.

Annex 2 – Published article within this dissertation

2.1. Article 1: 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

2.1. Article 1: 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.

Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group

Belén Agulló, Universitat Autònoma de Barcelona
Anna Matamala, Universitat Autònoma de Barcelona

ABSTRACT

Immersive media such as virtual reality or 360º content is increasingly present in our society. However, immersive content is not always accessible to all, and research is needed on how to cater for the needs of different kinds of users. This article will review the current situation of immersive technologies and their applications in media. Also, research studies carried out so far concerning subtitling and SDH in immersive media will be discussed, as well as current implementation of subtitles in immersive content, such as VR video games or 360º videos. Finally, the results from a focus group carried out in Spain with deaf and hard-of-hearing users, as well as professional subtitlers, on how to subtitle 360º video content will be presented. The results from the focus group shed light on how to address the new implications brought by immersive media in regard to subtitling. Some of the main areas discussed in the results are: subtitle position, contrast considerations for a correct reading, and how to indicate the location of the speakers, among others. Also, results show that users are willing to accept the implementation of new features in SDH in immersive content, such as icons for non-speech information or improvements to current standards.

KEYWORDS

Subtitles, subtitling for the deaf and hard-of-hearing, SDH, immersive content, virtual reality, 360º content, focus group.

1. Introduction

Immersive media are increasingly present in our society and considerable research efforts are put in developing better immersive technologies (Jayaraj *et al.* 2017). According to a report by VR Intelligence (2017), nearly half of the surveyed VR companies (46%) reported a strong or very strong growth in sales. However, some reports agree that the main reasons preventing users to buy VR headsets are economic, because prices are too high, and the lack of available content (VR Intelligence 2017; Sketchfab 2017).

Even if data point to expected growth, immersive content is not always accessible to all, and research is needed on how to cater for the needs of diverse users. Audiovisual Translation (AVT) and more specifically Media Accessibility (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 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 other. Still, most research has dealt with traditional media such as TV or cinema (Perego *et al.* 2015; Romero-Fresco (ed.) 2015), museums (Jiménez Hurtado *et al.* 2012; Szarkowska *et al*. 2016; Neves 2018) or live events (Orero and Matamala 2007; Udo and Fels 2011). In these environments, as in many others such as the localisation and game industry, accessibility has generally been considered an afterthought, despite many voices asking for the inclusion of accessibility in the creation process (Romero-Fresco 2013). To date, little research on accessibility in immersive media has been carried out and immersive technologies are on the rise, but they are still not fully implemented. This scenario was seen as an opportunity to start researching access services while immersive media were being developed, and the ImAc project was set up.

ImAc is a European project funded by the European Commission that aims to research how access services (subtitling, AD, audio subtitles, SL) can be integrated with immersive media. The project aims to move away from the constraints of existing technologies into an environment where consumers can fully customise their experience. The key action in ImAc is to ensure that immersive experiences address the needs of different kinds of users. ImAc follows a user-centred methodology (Matamala *et al*. 2018), so the first step in the project has been to ask users about their expectations. Two focus groups have been carried out in two countries (Germany and Spain) with different user profiles to gather feedback, define user scenarios and establish user requirements regarding SDH in 360º video content. This article aims to discuss the results of the focus group carried out at the Catalan Media Corporation (CCMA) in Spain.

The article begins with an overview of immersive technologies and immersive content in media, in order to contextualise our research. It then explains the limited research that has been carried out so far concerning subtitling in immersive media. Section 5 describes the methodology for the focus group and its implementation, and Section 6 reviews the results.

2. Immersive content: an overview

Immersive content allows users to feel as if they were physically transported into a different location. There are different solutions that can provide such experience. Fulldomes are one of those solutions. These are based on panoramic 360º videos projected on a dome structure, such as those that can be seen in planetariums, museums or flight simulators, for example.

Stereoscopic 3D technology represents another type of immersive technology that has had a relative presence in cinemas and homes during the past decade. Specifically, stereoscopy or 3D imaging:

> refers to a technique to create or enhance the illusion of depth in an image by presenting two offset images separately to the left and right eye of the viewer. The two 2D images are then combined in the brain to give the perception of 3D depth. The visual cortex of the brain fuses the two images into the perception of a three-dimensional scene or composition. (Agulló and Orero 2017: 92)

However, 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). Depending on the display design, the way of accessing stereoscopic 3D images differs (anaglyph glasses, head-mounted displays, active/passive glasses, etc.). This lack of standardisation, the intrusive nature of 3D and some uncomfortable side effects (headache or eyestrain) might prevent stereoscopy to become the main display for audiovisual (AV) products (Belton 2012).

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. Sherman and Craig (2003) define four key factors in VR (virtual world, immersion, sensory feedback and interactivity) which result in the following definition:

> **virtual reality** [is] a medium composed of interactive computer simulations that sense the participant's position and actions and replace or augment the feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world). (Sherman and Craig 2003: 13)

VR has also been referred to as a new medium, like the telephone or television, by authors such as Steuer (1992), who considers VR as a set of technical components, for example computers, head-mounted displays, sensors, among others. Most recently, VR has also been referred to as a set of computer-generated images that reproduce a reality and allow users to interact with their surroundings with the appropriate equipment (BBC 2014). However, some authors (Biocca 1992; Steuer 1992) consider technical definitions of VR to be limited. They suggest defining VR in terms of human experience and introducing the concept of *presence*, which can be defined as “the experience of one’s physical environment; it refers not to one’s surroundings as they exist in the physical world, but to the perception of those surroundings as mediated by both automatic and controlled mental processes” (Steuer 1992: 75).

This definition can be applied to different types of immersive content which, contrary to VR, does not need to involve an interactive response and an advanced equipment. Therefore, 360º or omnidirectional videos can also be defined as immersive content: they reproduce highly-realistic images recorded with special camera sets that represent a reality in which the users are observants and cannot interact (BBC 2014). Moreover, images can be combined with audio technologies such as immersive audio aimed to increase immersion. The spatial sounds are created with specially designed microphones that simulate the way in which human beings receive aural stimuli through the auditory system (BBC 2012).

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

> 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.

In this definition, the authors explain their concept of "virtuality continuum" in which MR is the hypernym that includes the more specific term AR. Azuma (1997) further defines AR as a medium that combines the real world with virtual objects that appear superimposed in the real world. 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).

VR and 360º content is accessed by different types of equipment. It can be directly viewed on a flat screen (for example, a computer, smartphone or a TV set) with a remote, touch pad or mouse to change the direction of the field of view (FoV), or it can be accessed with a head-mounted display (HMD), which can be either a tethered or a mobile device. Tethered HMD, such as PlayStation VR, Oculus Rift or Vive, incorporate high definition screens and are connected to high-performance computers or last generation consoles (Deloitte 2016). They are generally used for gaming purposes and the quality of the VR experience is higher. On the other hand, mobile HMD, such as Samsung Gear VR or Google Cardboard, are dependent on smartphone technology (such as accelerometer or gyroscope).

3. Immersive content in media

Broadcasters and video content developers are starting to experiment with immersive content. According to a report on VR by the European Broadcasting Union (EBU 2017a), 49% of its members are starting to explore or further develop immersive content. Most EBU members think that the potential of immersive content is clear, because it offers new opportunities to tell stories from a different perspective. However, factors such as technical limitations, lack of knowledge and uncertainty about the return on investment are holding some of them back (EBU 2017a). The current preferred format is 360º video over VR or AR/MR and the trend in terms of duration is 5-10 minutes. Most stories told in 360º/VR are "history or news and current affairs products, as 360º/VR allows the user to gain better understanding of the story being told" (EBU 2017a: 9). Sports and music events are clear candidates for 360º/VR. In the report, attention is also directed at the challenges posed in terms of storytelling, since the plots are non-linear and the level of user interaction is not determined.

Immersive journalism is featured as a key concept in the EBU report. This relatively new concept entails "the production of news in a form in which people can gain first-person experiences of events or situation described in news stories" (De la Peña *et al.* 2010: 291). According to De la Peña *et al.*, VR is a perfect media that could help journalists in eliciting deeper emotions in the audience. Major broadcasters and newspapers such as the BBC, *The New York Times*, *The Washington Post* and *ABC News* have already started working on this. To name just two examples, in 2015 *The New York Times* decided to reward their subscribers with a pair of Google

Cardboard glasses, as a strategy to promote their own immersive content (Wohlsen 2015). The company even launched its own smartphone application (NYT VR). More recently, as reported by Jones (2017), *ABC News* recorded a 360º short film about North Korea with a relatively positive outcome. Immersive journalistic experiences have been proven to elicit emotions and increase audience engagement (Jones 2017), showing greater potential for journalism and broadcasting.

As a genre, fiction appears to be a rather underexplored area in relation to immersive content. However, there are some indicators which show that VR has had a moderate penetration in the entertainment industry. For example, the appearance of immersive content in major film and TV industry events is one of them. In 2016, the VR film *Henry* was awarded an Emmy for Outstanding Original Interactive Program (Oculus Team 2016). In 2017, the immersive short film *Pearl* was nominated for the Oscars, under the category of Best Animated Short Film (Hall 2017). Pixar has also developed a VR and interactive experience to market their animated film *Coco* (Lee 2017). However, most fictional immersive experiences are computer-generated products and not images recorded with 360º cameras. According to the BBC, "truly interactive VR video is in its infancy and can be expensive to create, but total or partial animation or CGI can be used very effectively and efficiently, while other production techniques may yet emerge or become more accessible over time" (Conroy 2017). Therefore, it could be inferred that immersive technologies are still not sufficiently developed to be implemented in creating successful fictional films or movies. Some of the reasons could be the hindrances posed by technology, which is delivering a quality that is still not considered suitable for the audience (EBU 2017a); also, the lack of knowledge in delivering well-written immersive stories. It is in this context of development that the integration of access services in the production line should be researched, adopting a user-centred approach. The users' voice needs to be heard before the technology is fully implemented.

4. Subtitling and SDH in immersive media

AVT and MA research on immersive content is at an early stage. Nevertheless, we can find studies that have addressed the challenges of creating and consuming subtitles in stereoscopic 3D content (Vilaró 2011; Lambooij *et al.* 2013). Some of the main issues when integrating subtitles in 3D imaging is that superimposing 2D subtitles on a 3D image can generate effects such as ghosting, which hinders the readability of the subtitles and can cause headaches and eyestrain (Agulló and Orero 2017). However, the implementation of some techniques such as the positioning of the subtitle close to the screen plane or the use of lighting, shades and colours to reduce contrast between the screen and the subtitle (González-Zúñiga *et al.* 2013) could contribute to minimising the impact of such issues.

Reception studies on access services in VR and 360º content are almost non-existent, with research on subtitling by the BBC being an exception. Although not focused specifically on SDH, audiences with hearing loss might be potential users of intralinguistic subtitles. The BBC research designed four subtitle scenarios for 360º content (Brown *et al.* 2018: 3-6): (a) subtitles equally spaced by 120º in a fixed position below the eye line; (b) subtitles following head immediately always in front of the user; (c) subtitles following head with lag in front of the user; and (d) subtitles appearing in front of users, and then fixed until they disappear. 24 participants, frequent users of TV subtitles, took part in the study. They randomly viewed the four variables on an Oculus Rift HMD, and replied to a questionnaire (Brown *et al.* 2018: 7-9). Results show that the preferred solution was (b), in line with a similar behaviour to subtitles in 2D. Their conclusion was that sometimes the simplest solution is the best (Brown *et al.* 2018: 29-33), but it remains to be seen whether results would be the same with longer content.

4.1. Revisiting subtitling parameters in immersive content

User reception studies in subtitling and in SDH have allowed the definition of a set of preferred parameters for users (Jensema *et al.* 1996; Romero-Fresco 2009; Matamala and Orero (eds) 2010) in a myriad of aspects such as: the number of characters and lines per subtitle; subtitle editing; font type and size; boxes, borders and shadows; justification and spacing; paralinguistic information, and subtitle speed (Neves 2005; Arnáiz-Uzquiza 2012; Romero-Fresco (ed.) 2015). Immersive environments, however, pose specific challenges that need to be considered.

Subtitle positioning, which is a widely researched (Bartoll and Martínez-Tejerina 2010) and standardised parameter for 2D products, is one of the main issues when designing subtitles for immersive content, since user behaviour in the immersive environment is unpredictable (Arrés *forthcoming*). While safe areas for subtitling are already defined and guidelines are provided for content consumed in TV or flat screens (EBU 2017b), recommendations for safe areas in immersive devices such as HMD are still lacking. The FoV for users in VR environments is wider than the FoV in 2D products on a flat screen. But to the best of our knowledge, eye-tracking studies, showing where the users direct their attention when reading subtitles in VR or 360º content, are lacking. Therefore, the safe area for subtitles in immersive environments needs to be defined very carefully and tested later on.

The interactivity and the freedom of movement that are inherent to immersive products also impact other subtitling parameters such as character identification, because there is the possibility that a character speaks but is located outside the FoV of the user. In 2D SDH, character identification is usually solved by using different colours, name tags or speaker-dependent placement of subtitles or a combination of these (Neves 2005; Arnáiz-Uzquiza 2012). However, immersive content introduces a new dimension: direction, which is particularly relevant in immersive environments.

Another SDH parameter that may change in subtitles for immersive content is the display of non-speech information, such as music or sounds. Immersive technologies offer new opportunities for implementing new features in SDH. In previous studies, the implementation of graphic elements such as icons to display non-speech information has been suggested as a way to optimise reception (Civera and Orero 2010). Other authors have also tested the reception of emotions via emoticons (Arnáiz-Uzquiza 2015) and other creative approaches (Sala Robert 2016). Although these are not extended practices for SDH, the technical advances provided by immersive technologies could open the possibility of introducing new elements that might counterbalance other VR limitations. However, it remains to be seen whether alternative approaches such as using icons may help reduce that kind of discomfort.

4.2. Some examples from current practices

Some randomly selected current experiences in subtitling immersive environments can give us food for thought as to the opportunities and challenges subtitles in immersive AV products may pose. The Spanish television series *El Ministerio del Tiempo* (The Ministry of Time) launched an immersive experience in the form of an interactive episode. The episode "El tiempo en tus manos" (The time is in your hands) is one of the first fictional TV episodes launched in an immersive format. The interlingual subtitles (Spanish into English) in this short episode are positioned slightly below the centre of the screen, following the movement of the head and floating through the screen as the user's head moves. The transition of the subtitles, when the users move their heads, presents a slight delay in reaction time. Therefore, when the movement is abrupt, subtitles are not positioned in the centre of the screen, but float in the direction of the movement of the head. They only settle into a fixed position in the centre of the image when the user's head is still. The font type is a white sans serif font without a background box. The justification is centred. The segmentation rules are not followed, and some linguistic issues are found, such as missing information.

Another example is the clip *The Displaced* created by *The New York Times*, in which children who have been driven away from their homes explain their current situation as refugees. In this video, subtitles are burnt in in three different fixed positions in the 360º video, so when the users move their head to explore the scenario, they will always find the subtitles somewhere in their FoV (Brown *et al.* 2017). The font is white sans serif, smaller than the previous example, which hinders readability. Moreover, these subtitles do not include a background box and the contrast is very low, meaning that sometimes the text is very difficult to read.

Video games provide other examples of subtitles in immersive environments, although subtitling practices in video games do not always follow the same rules as in other AV content (Mangiron 2013). In the game *Eve Valkyrie*, for PlayStation VR, intra- and interlingual subtitles are located in a fixed position in the centre of the screen. Therefore, if the users turn their head towards a different part of the scene, they will not be able to read the subtitle. This strategy might result into less freedom of movement, but it might also help avoiding distractions from the main action when the narrative requires the user's attention. The font of the subtitles is sans serif and yellow. Subtitles contain more than 2 lines in many cases, and do not follow segmentation rules.

Summer Lesson and *London Heist*, both games for PlayStation VR, use a similar strategy to implement subtitles. In this case, intra- and interlingual subtitles are always displayed in front of the user, at the bottom of the FoV and centred, which is less intrusive for the scene. They both use sans serif fonts in white. The subtitles in *London Heist* include a black background box to facilitate reading. This strategy could be appropriate for immersive environments, because the user has freedom of movement and it would be very complex for the professional subtitler to foresee where the video background could interfere with the reading and, therefore, change the position of the subtitle, as is the case in current subtitling practices for 2D content. Finally, another strategy for implementing subtitles in VR games appears in *Battle Zone*, for PlayStation VR. In this game, subtitles are integrated in the scene, as they would appear in a head-up display. In this example, the subtitles are not obtrusive in the scene because they appear as if they were part of the environment, integrated in a futuristic spaceship.

The previous examples have shown how some critical issues such as subtitling positioning have been implemented in a selection of subtitled immersive content. Others, such as character identification or non-speech information display, have not been addressed at all. However, it is paramount to gather user feedback in order to generate subtitles that can be easily implemented and accepted by end users. Focus groups such as the one presented in the next section can contribute to such an end.

5. Gathering user feedback: focus group methodology

Focus groups were considered appropriate for identifying user needs at the beginning of the ImAc project, before access services in immersive media were actually implemented. A shared methodology was developed for the five focus groups (three on AD and two on SDH) which took place in four different countries, and ethical clearance was obtained. The preparation stage for the focus groups involved two main steps.

To the best of our knowledge, access services such as subtitles or audio description in immersive media were not fully implemented at the time of conducting the focus groups. Therefore, in order to identify the most relevant questions to be posed to participants, it was necessary to define user types and scenarios. Two main user profiles were defined: those creating the services, i.e. professional users (for instance, IT, graphic designers, subtitlers, audio describers, and SL interpreters), and those consuming the services, i.e. home users (for instance, deaf, hard-of-hearing, blind, low vision users, the elderly). At this stage it was decided that the focus would be mainly on those consuming the services, gathering additional data from a few professionals and opening the door to future research with other professional profiles such as content creators. It was also decided that home users would be advanced, meaning they would have some knowledge or special interest on the technologies being developed.

The focus group presented in this paper was organised by the Catalan Media Corporation (CCMA) in collaboration with Universitat Autònoma de Barcelona and was held on 28 November 2017 at the CCMA premises. The main aim of the focus group was to obtain feedback regarding expectations, recommendations and desires from professional and home users when consuming and editing SDH in 360º videos, as well as SL access services. The focus of this article is on subtitling, so only the results related to SDH will be reported. The results of this focus group reflect the needs of the Spanish SDH audience. Another focus group regarding SDH was carried out in Berlin (Germany) by another project partner for which the results have not been published. The contents of the focus groups were different and so was the target audience (different language and subtitling habits). Therefore, the results were not fully comparable, and it was decided not to include them in the present study.

5.1. Participants

There were 14 participants (6 males, 8 females): 10 advanced home users (6 signers, 4 oralists) and 4 professional users (2 subtitlers, 1 technical expert, 1 representative from a user association). Age range was 21-40 (3), 41-60 (7), and +60 (4). Three participants had secondary education studies, four participants had further education studies, six had university studies and one person did not reply to this question. Three of them reported having a device to access VR content (VCR, glasses and PC, respectively). Mobile phones were the technology most frequently used by the participants on a daily basis (14), followed by TV (14), laptop (12), PC (10) and tablet (8). The advanced home users were deaf (8) and hearing-impaired people (2), most having the disability from birth (4) or when they were between 0-4 years (5) or 41-60 years (1). The preferred devices for watching online video content was PC (7) and laptop (7), followed by smartphone (5), tablet (3) and TV (3).

Even though the recommended group size is up to 10 (Bryman 2004: 507), the profiles of deaf and hard-of-hearing users are diverse (Báez Montero and Fernández Soneira 2010: 26) and it was therefore considered that including the maximum number of home users would provide a more accurate demographic sample. The diversity in the results of the focus group confirmed that this approach was appropriate. Moreover, following standard recommendations on focus groups, it was deliberately decided to over-recruit in order to allow for no-shows (Wilkinson 1999: 188) and a higher number of short suggestions (Morgan 1998: 75).

5.2. Procedure

The focus group included five stages. First, participants were welcomed by the facilitator who briefly explained the aim of the ImAc project. The focus group took place in a meeting room equipped with a table and chairs for the participants and a computer and a large TV screen to show the examples to be discussed. A SL interpreter was present, as well as two researchers who took notes and summarised the conclusions in real time. Secondly, the aim of the focus group was explained to the participants, and they were asked to sign informed consent sheets. The third step was filling in a short questionnaire on demographic information. Finally, the group discussion began. To trigger the discussion, the facilitator gave a short introduction to VR and 360º content and explained how 360º content can be accessed, showing VR glasses to the participants. He explained that 360º content can also be accessed on a flat TV screen using a mouse to move around the 360º scene. As a specific example, an excerpt of the TV show *Polònia* was shown to participants on a flat TV screen. Different types of subtitles were presented to give users some ideas about how SDH could be implemented in immersive content and to stimulate their imagination: subtitles located in a fixed position, subtitles located close to the speaking character, and subtitles located each 120º in the 360º view. The facilitator also posed questions about how users would like to interact with this type of access services and what features a future platform giving access to these services should have. Together with these stimuli, the facilitator also used a list of guiding questions grouped under major topics to generate participants' reactions, taking special care to allow participants to raise aspects that they considered relevant even if not included in the list. A balance between an open-ended and a structured approach was sought, and the result was a lively discussion in which interesting suggestions were made.

As the focus group took place, one researcher was drafting a list of conclusions. Reading these conclusions and agreeing on them was the last step of the focus group, which lasted 90 minutes. At the end of the session, participants were thanked for their participation and they were told about the next steps in the project.

6. Focus group results

Data analysis followed a qualitative approach, due to the number of participants and the methodological tool chosen. As explained above, two researchers took notes on a shared document and summarised the conclusions in real time. After the focus group, the notes were thoroughly revised and tagged. This procedure allowed to identify three main areas in which users voiced their views: (1) feedback from advanced home users concerning the services; (2) feedback from advanced home users concerning the interaction with a future platform giving access to the services; and (3) feedback from professional users concerning content creation. The analysis also allowed to define aspects in which there was consensus among users and aspects in which opinions diverged, as described next.

6.1. Advanced home users: services

In general, home users considered that subtitles in immersive media should be based on approved subtitling rules (AENOR 2003) and, if necessary, improvements might be implemented to adapt existing rules to the new needs posed by immersive environments.

Regarding the position of the subtitles, users suggested that subtitles should always appear in a fixed position in relation to the users' FoV. They also agreed that subtitles should always appear at the bottom — except in specific cases, such as football matches. There was a brief discussion about the possibility of customising the position of the subtitles. It was even suggested that the placement of subtitles in real time should be changed while watching the 360º video. However, users finally disregarded this option, since they all agreed that subtitles at the bottom of the FoV was the most comfortable solution. It remains to be seen whether research will actually confirm this is the best solution.

Most participants were concerned about the fact that sometimes the subtitles could not be read because of the background image. They stated that it is important, therefore, to have the possibility to choose subtitles with a black background box to facilitate the reading. Also, some participants expressed their worry about the fact that subtitles in immersive media could be disruptive if they appear in a close-up or some other scenarios where the subtitle is obstructing the image. They said that subtitle editors should pay special attention to avoid disrupting the immersive experience.

For character identification, users stated that it is necessary to maintain colour coding to identify characters, as this is already done in SDH for 2D content.

Concerning the display of non-speech information (sounds, music, paralinguistic information, etc.), different options were proposed. In general, users requested that basic subtitling elements that have been previously approved in the regulations (for example, how to indicate music) should be retained. However, they accepted that new technologies may bring new possibilities. Some users preferred to receive non-speech information in the form of text in brackets, as is now the case in most subtitled TV programmes. Other users, considering the new technology in use, preferred to receive this information as icons. In that sense, users suggested the possibility of using a closed list of icons. For example, a lightning icon to indicate the sound of a storm. Regarding the position of non-speech information, users did not reach a consensus. Some stated that they preferred them to be located at the top, others at the bottom close to the subtitle (dialogue), and others would like to move them to a different location. In general, participants did not like the idea of having non-speech information at the top, as it is stated in the current Spanish UNE rule for SDH, because they do not have time to read both the subtitle (at the bottom) and the non-speech information (at the top). They suggested to change this in immersive environments and place the non-speech information in form of icons or text between brackets close to the subtitle area, stating that this would be easier to process. Also, some hard-of-hearing participants stated that they do not need non-speech information in the subtitles, and they would prefer to deactivate them if possible. In this sense, users would like to be able to customise the position of non-speech information.

Users also considered the challenges that the new dimension brought by immersive media (i.e. space and the need to indicate directions) would entail when it comes to SDH in immersive content. In that sense, users stated that it was difficult to know where to look to see the character speaking. They considered that the subtitle should indicate how you need to move your head (four directions), with icons (arrows), indicators between brackets (to the left, to the right) or some sort of mechanism. It was suggested that a compass or radar could be used to that end and that it should be always visible on the screen. Participants also agreed that the radar or compass should be close to the subtitle, otherwise it could be distracting.

As for the subtitle content, users insisted that it should include all the information, both on screen and off screen; in other words, dialogues taking place both within and outside the user's FoV. They suggested that this could be indicated with ON and OFF tags. They also stated that there are different needs among users and, consequently, subtitles must be adapted to different profiles. For example, there could be different levels of speed (faster/slower). However, users considered that summarised or simplified subtitles do not generally help deaf people, because this type of subtitle make it more difficult to follow the AV content. Nevertheless, they conceded that simplified subtitles may be useful for users with other types of needs and could be considered an alternative. It was clear that user profiles are diverse, and customisation should be a priority.

6.2. Advanced home users: platform interaction

Users were asked about the options and features that they would like to have in an immersive platform which would give access to virtual content with access services implemented. At the time of the focus group, no prototype was available. Therefore, thought-provoking questions were presented to participants based on hypothetical user scenarios.

Regarding interaction with an immersive interface, users positively valued the possibility of personalisation, i.e. having different layers that could be activated or not. For example, some participants preferred subtitles only for dialogues, others needed non-speech information indications and others wanted to have as much indications as possible, including directions. However, some elements were not considered in need of customisation, such as the position of the subtitles, which it was reported should always be at the bottom of the field of view, because they considered it would be easier to read. Moreover, both professional and home users considered that the user should customise this future platform the first time but then those parameters should be recorded by the interface for future use. Users also suggested that this customisation should be transferrable from one device to the other (importing profile, that is, the user profile could be imported) and they requested the possibility of creating more than one profile. They also considered the possibility of transferring a profile from the user’s device to another external device (for example, at a friend’s home).

Regarding interaction with access services, users positively valued the possibility of alternative interactions (for example, voice commands), although they did not find it necessary for their specific needs and indicated that implementation costs should be taken into account. However, they added that if this platform were to be developed for other profile types (for example, blind users), it could be an additional resource.

Regarding companion screens, participants liked the possibility of using the smartphone to interact with the platform as a touch-screen (like a “mouse”) and to customise their preferences. One user even suggested the possibility of including a finger-sensor that would allow users to see their own fingers on the virtual image. There were different opinions regarding the need of reproducing the same content on the smartphone screen, since the smartphone is often used as an element to access additional content. When accessing AV content together with other people, users did not want to consume subtitles on a different screen because this made them feel excluded.

6.3. Professional users: SDH creation

Professional users expressed their proposals regarding the production of SDH in immersive environments. They agreed that vertical positioning of the subtitles could be an interesting option for separating dialogue subtitles from non-speech information, although they considered that home users must be able to decide or set up where they prefer to locate the subtitle.

Regarding the production of subtitles, they stated that they preferred an on-screen display (player) showing one dynamic angle of the 360º view, so that they could choose which angle to see using cursors or mouse movements. Professional users considered that they should be able to test the results with both HMD and flat screen (for instance, a PC screen).

Regarding the subtitling tool, users indicated that they would need a subtitling editor similar to the existing ones for SDH, but it should add the

360º displaying and the possibility of adding emoticons and text messages to show sound actions that take place parallel to the dialogue subtitles. They also highlighted the need for the editor tool to offer original 360º immersive audio because it is important to identify where the sound/dialogues come from, as this information is requested by end users.

7. Conclusions

This article has described the role of immersive media in our society and has put forward the need to make them accessible to all users. The emphasis has been put on how subtitles can be integrated in immersive environments, with reference to the limited existing research and practice. Adopting a user-centred approach, the results of a focus group on SDH developed as part of the ImAc project have been presented. According to participants' feedback, SDH for 360º videos should: (1) be located in a fixed position and always visible in relation to the FoV and preferably at the bottom; (2) have a background box to avoid contrast issues with an unpredictable background; and (3) include a system to indicate directions when the speaker is outside the FoV, such as arrows, a compass or text between brackets. Also, results show that home users are willing to accept the implementation of new features in SDH in immersive content, such as icons for non-speech information, because of the new possibilities and dimensions brought by this medium. Moreover, customisation options appear to be a desirable feature among participants. Users also show their agreement and interest in continuing established practices and regulations for SDH, such as the Spanish subtitling standard UNE 153010 (AENOR 2003). However, they agree in introducing some changes to improve the current standards. For example, they would like to have the non-speech information or direction information closed to the subtitle area and not at the top as it currently is, in order to avoid distractions.

One of the limitations of the present study is that it only applies to the Spanish audience and should be replicated in other countries to confirm the validity and generalisation of the conclusions. Another focus group was carried out in Germany for the ImAc project,[1] with similar results (German participants would also like the subtitles always visible in the FoV and also suggested using an arrow to indicate the location of the speaker). However, the contents and examples were not the same, because the audiences spoke different languages and, therefore, the results are not comparable. It was not the intention of the project to compare the two focus groups, but rather gather a general feedback from end users that would set the basis to start developing a prototype for access services in 360º videos. The prototype will be later tested with a larger number of participants in the next stages of the project.

Therefore, the next step will be to transfer user feedback into user requirements and implement the features in immersive content in order to verify whether it is technically possible or whether there are limitations.

Once implemented, user testing will be necessary to verify or reject the validity of the proposed SDH models.

In conclusion, it is clear that research is needed in the field of MA for immersive media. For this purpose, the ImAc project will be a perfect laboratory environment for the development of a successful SDH model for immersive content. This is a significant step in the field of AVT and MA, since the design of accessibility will be taken into account before the technology is fully implemented in society.

Acknowledgements

ImAc has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 761974. The authors are members 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 Belén Agulló'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.

We would like to thank CCMA, and especially Francesc Mas and Jordi Mata, who were involved in the focus group management at the CCMA premises, and to all participants who took part in the focus group.

References

- **Agulló, Belén and Pilar Orero** (2017). "3D Movie Subtitling: Searching for the best viewing experience." *CoMe - Studi di Comunicazione e Mediazione linguistica e culturale* 2, 91-101.
- **Arnáiz-Uzquiza, Verónica** (2012). *Subtitling for the deaf and the hard-of-hearing: Some parameters and their evaluation*. PhD thesis. Universitat Autònoma de Barcelona. https://www.tdx.cat/handle/10803/117528 (consulted on 08.01.2019).
- **Arnáiz-Uzquiza, Verónica** (2015). "Eye tracking in Spain." Pablo Romero-Fresco (ed.) (2015). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe*. Bern: Peter Lang, 263-287.
- **Arrés, Eugenia** (forthcoming). "Traducir en entornos de realidad virtual: cómo garantizar la inmersión del usuario." To appear in *New Technologies Applied to Translation Studies: strategies, tools and resources*.
- **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*. Madrid: Asociación Española de Normalización y Certificación.
- **Azuma, Ronald** (1997). "A Survey of Augmented Reality." *Presence: Teleoperators and Virtual Environments* 6(4), 355-385.

- **Azuma, Ronald *et al.*** (2001). "Recent Advances in Augmented Reality." *IEEE Computer Graphics and Applications* 21(6), 34-47.
- **Báez Montero, Inmaculada Concepción and Ana María Fernández Soneira** (2010). "Spanish deaf people as recipients of closed captioning." Anna Matamala and Pilar Orero (eds) (2010). *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*. Frankfurt: Peter Lang, 25-44.
- **Bartoll, Eduard and Anjana Martínez-Tejerina** (2010). "The positioning of subtitles for the deaf and hard of hearing." Anna Matamala and Pilar Orero (eds) (2010). *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*. Frankfurt: Peter Lang, 69-86.
- **British Broadcasting Corporation (BBC)** (2012). "Binaural Sound: Immersive spatial audio for headphones." http://www.bbc.co.uk/rd/projects/binaural-broadcasting (consulted on 19.12.2017).
- **British Broadcasting Corporation (BBC)** (2014). "360 Video and Virtual Reality: Investigating and developing 360-degree video and VR for broadcast-related applications." http://www.bbc.co.uk/rd/projects/360-video-virtual-reality (consulted on 19.12.2017).
- **Belton, John** (2012) "Digital 3d cinema: Digital cinema's missing novelty phase." *Film History: An International Journal* 24(2), 187–195.
- **Biocca, Frank** (1992). "Virtual Reality Technology: A Tutorial." *Journal of Communication* 42(4), 23–72.
- **Brown, Andy *et al.*** (2017). "Subtitles in 360-degree Video." Judith Redi *et al.* (eds) (2017). *TVX '17 ACM International Conference on Interactive Experiences for TV and Online Video* (Hilversum, The Netherlands, 14-16 June 2017). New York: ACM, 3-8.
- **Brown, Andy *et al.*** (2018). "Exploring Subtitle Behaviour for 360° Video. White Paper WHP 330." https://www.bbc.co.uk/rd/publications/whitepaper330 (consulted on 13.02.2019).
- **Bryman, Alan** (2004). *Social Research Methods*. New York: Oxford University Press.
- **Civera, Clara and Pilar Orero** (2010). "Introducing icons in subtitles for deaf and hard of hearing: Optimising reception?" Anna Matamala and Pilar Orero (eds) (2010). *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*. Frankfurt: Peter Lang, 149-162.
- **Conroy, Andy** (2017). "The BBC and Virtual Reality." *BBC Research & Development*. http://www.bbc.co.uk/rd/blog/2016-06-the-bbc-and-virtual-reality (consulted on 21.12.2017).
- **De la Peña, Nonny *et al.*** (2010). "Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News." *Presence: Teleoperators and Virtual Environments* 19(4), 291-301.
- **Deloitte** (2016). *Technology, Media & Communications Predictions*. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Technology-Media-Telecommunications/gx-tmt-prediction-2016-full-report.pdf (consulted on 20.12.2017).
- **European Broadcasting Union (EBU)** (2017a). *Virtual Reality: How are public broadcasters using it?* https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it (consulted on 18.12.2017).
- **European Broadcasting Union (EBU)** (2017b). *Safe Areas for 16:9 Television Production*. https://tech.ebu.ch/docs/r/r095.pdf (consulted on 31.01.2018).
- **González-Zúñiga, Diekus, Carrabina, Jordi and Pilar Orero** (2013). "Evaluation of Depth Cues in 3D Subtitling." *Online Journal of Art and Design* 1(3), 16–29.

- **Greco, Gian Maria** (2016). "On Accessibility as a Human Right, with an Application to Media Accessibility." Anna Matamala and Pilar Orero (eds) (2016). *Researching Audio Description. New Approaches.* London: Palgrave Macmillan, 11-33.
- **Hall, Charli** (2017). *Watch VR's first Oscar-nominated short film in Polygon*. https://www.polygon.com/2017/1/24/14370892/virtual-reality-first-oscar-nominated-short-film-pearl (consulted 24.10.2017).
- **Holliman, Nicolas *et al*.** (2011). "Three-dimensional displays: a review and applications analysis." *IEEE Trans. Broadcast* 57(2), 362–371.
- **Jayaraj, Lionel, Wood, James and Marcia Gibson** (2017). "Improving the Immersion in Virtual Reality with Real-Time Avatar and Haptic Feedback in a Cricket Simulation." Wolfgang Broll *et al*. (eds) (2017). *Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality*. Los Alamitos, California: Conference Publishing Services, 310-314.
- **Jensema, Carl, McCann, Ralph and Scott Ramsey** (1996). "Closed-captioned television presentation speed and vocabulary." *American Annals of the Deaf* 141(4), 284-292.
- **Jiménez Hurtado, Catalina, Seibel, Claudia and Silvia Soler Gallego** (2012). "Museos para todos. La traducción e interpretación para entornos multimodales como herramientas de accesibilidad universal." *MonTI. Monografies de Traducció i Interpretació* 4, 349-383.
- **Jones, Sarah** (2017). "Disrupting the narrative: immersive journalism in virtual reality." *Journal of Media Practice* 18(2–3), 171–185.
- **Lambooij, Marc, Murdoch, Michael, Ijsselsteijn, Wijnand and Ingrid Heynderickx** (2013). "The impact of video characteristics and subtitles on visual comfort of 3D TV." *Displays* 34(1), 8-16.
- **Lee, Nicole** (2017). "Pixar's VR debut takes you inside the entrancing world of 'Coco'". *Engadget*. https://www.engadget.com/2017/11/15/pixar-coco-vr/ (consulted on 21.12.2017).
- **Mangiron, Carme** (2013). "Subtitling in game localisation: a descriptive study." *Perspectives: Studies in Translation Theory and Practice* 21(1), 42-56.
- **Matamala, Anna and Pilar Orero** (eds) (2010). *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*. Bern: Peter Lang.
- **Matamala, Anna et al.** (2018). "User-centric approaches in Access Services Evaluation: Profiling the End User." *Proceedings of LREC Workshop "Improving Social Inclusion using NLP: Tools, Methods and Resources", ISINLP2 2018*. https://ddd.uab.cat/record/189707 (consulted on 13.02.2019).
- **Milgram, Paul and Fumio Kishino** (1994). "A taxonomy of mixed reality visual displays." *IEICE Transactions on Information Systems* 77, 1321–1329.
- **Morgan, David L.** (1998). *Planning Focus Groups*. Thousand Oaks: Sage.
- **Neves, Joselia** (2005*). Audiovisual translation: Subtitling for the deaf and hard of hearing*. PhD Thesis. Roehampton University. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.1405&rep=rep1&type=pdf (consulted on 08.01.2019).
- **Neves, Joselia** (2018). "Cultures of Accessibility: Translation making cultural heritage in museums accessible to people of all abilities." Sue-Ann Harding and Ovidi Carbonell Cortes (eds). *The Routledge Handbook of Translation and Culture*. London/New York: Routledge, 415-430.

- **Oculus Team** (2016). "Oculus Film Short "Henry" Wins an Emmy!" *Oculus Blog*. https://www.oculus.com/blog/oculus-film-short-henry-wins-an-emmy/ (consulted on 24.10.2017).
- **Orero, Pilar and Anna Matamala** (2007). "Accessible opera: Overcoming linguistic and sensorial barriers." *Perspectives: Studies in Translation Theory and Practice* 15(4), 262–277.
- **Perego, Elisa, Del Missier, Fabio and Sara Bottiroli** (2015). "Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects." *Perspectives: Studies in Translation Theory and Practice* 23(1), 1-21.
- **Remael, Aline, Orero, Pilar and Mary Carroll** (eds) (2014). *Audiovisual Translation and Media Accessibility at the Crossroads*. Amsterdam/New York: Rodopi.
- **Romero-Fresco, Pablo** (2009). "More haste less seed: Edited vs. verbatim respoken subtitles." *VIAL (Vigo International Journal of Applied Linguistics)* 6, 109-133.
- **Romero-Fresco, Pablo** (2013). "Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking." *The Journal of the Specialised Translation* 20, 201-223.
- **Romero-Fresco, Pablo** (ed.) (2015). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europ*e. Bern: Peter Lang.
- **Sala Robert, Elia** (2016). *Creactive subtitles. Subtitling for all*. PhD Thesis. Universitat Pompeu Fabra. https://repositori.upf.edu/handle/10230/27737 (consulted on 08.01.2019).
- **Sherman, William and Alan Craig** (2003). *Understanding Virtual Reality: Interface, Application, and Design*. San Francisco: Morgan Kaufmann Publishers (An imprint of Elsevier Science).
- **Sketchfab** (2017). *VR Industry Trends: A comprehensive overview on the state of the VR Industry in Q2-2017*. https://sketchfab.com/trends/q2-2017 (consulted on 18.12.2017).
- **Steuer, Jonathan** (1992). "Defining Virtual Reality: Dimensions Determining Telepresence." *Journal of Communication* 42(4), 73-93.
- **Szarkowska, Agnieszka *et al.*** (2016). "Open Art: Designing Accessible Content in a Multimedia Guide App for Visitors with and without Sensory Impairments." Anna Matamala and Pilar Orero (eds). *Researching Audio Description. New Approaches*. London: Palgrave Macmillan, 301-320.
- **Udo, John Patrick and Deborah Fels** (2011). "From the describer's mouth: reflections on creating unconventional audio description for live theatre." Adriana Serban, Anna Matamala and Jean-Marc Lavaur (eds). *Audiovisual Translation in Close-Up: Practical and Theoretical Approaches*. Bern: Peter Lang, 257-278.
- **Vilaró, Anna** (2011) "Does 3D stereopsis change the way we watch movies? An eye-tracking study". Paper presented at *4th International Conference Media for All* (London, 28 June-1 July 2011).
- **VR Intelligence** (2017). *Virtual Reality Industry Survey*. http://vr-intelligence.com/vrx/docs/VRX-2017-Survey.pdf (consulted 19.12.2017).
- **Wilkinson, Sue** (1999). "Focus Groups Methodology: A Review." *International Journal of Social Research Methodology* 1(3), 181-203.
- **Wohlsen, Marcus** (2015). "Google Cardboard's New York Times Experiment Just Hooked a Generation on VR." *Wired Magazine*. https://www.wired.com/2015/11/google-cardboards-new-york-times-experiment-just-hooked-a-generation-on-vr/ (consulted on 21.12.2017).

Biographies

Belén Agulló is a predoctoral researcher in the Department of Translation, Interpreting and Eastern Asian Studies at Universitat Autònoma de Barcelona, where she is working on the Horizon 2020-funded project Immersive Accessibility (ImAc). Her PhD focus is subtitling for the deaf and hard-of-hearing in immersive media. Previously, she worked for more than 5 years in the game localisation industry. She teaches game localisation in different masters in Spain and France. Her research interests include audiovisual translation and media accessibility.

E-mail: belen.agullo@uab.cat

Anna Matamala, BA in Translation (UAB) and PhD in Applied Linguistics (UPF), is an associate professor at the Universitat Autònoma de Barcelona. Currently leading the TransMedia Catalonia group, she has participated and led projects in audiovisual translation and media accessibility. She has taken an active role in the organisation of scientific events (M4ALL, ARSAD), and has published in journals such as *Meta*, *Translator*, *Perspectives*, *Babel*, *Translation Studies*. She is currently involved in standardisation work.

E-mail: anna.matamala@uab.cat

Notes

[1] Results for the German focus group can be found in the public report: http://www.imac-project.eu/documentation/deliverables/.

Annex 3 – Ethical Committee Documentation

3.1. ImAc Project's ethical procedure approved and signed
3.2. Written informed consent forms and information sheets for participating in the experiments (Spanish)
3.3. Written informed consent forms for video, photographs or audio recording (Spanish)
3.4. Written consent form online (for subtitle editor online test) (English)

3.1. ImAc Project's ethical procedure approved and signed

Universitat Autònoma de Barcelona
Vicerectorat d'Investigació

Comisión de Ética en la Experimentación Animal y Humana (CEEAH)

Universitat Autònoma de Barcelona
08193 Bellaterra (Cerdanyola del Vallès)

La Comisión de Ética en la Experimentación Animal y Humana (CEEAH) de la Universitat Autònoma de Barcelona, reunida el día **29-03-2019**, acuerda informar favorablemente el proyecto con número de referéncia **CEEAH 4025R** y que tiene por título **"Immersive Accessibility"** presentado por **Pilar Orero Clavero**

Elaborado:	**Aprovado:**
Nombre: Nuria Perez Pastor Cargo: Secretària de la CEEA de la UAB Fecha: NURIA PEREZ PASTOR Firmado digitalmente por NURIA PEREZ PASTOR Nombre de reconocimiento (DN): c=ES, ou=Vegeu https://www.aoc.cat/CATCert/Regulacio, sn=PEREZ PASTOR, givenName=NURIA, serialNumber=35109638T, cn=NURIA PEREZ PASTOR Fecha: 2019.04.04 16:12:20 +02'00'	Nombre: José Luis Molina González Cargo: President de la CEEAH de la UAB Fecha: JOSE LUIS MOLINA GONZALEZ - DNI 36561625C Firmado digitalmente por JOSE LUIS MOLINA GONZALEZ - DNI 36561625C Fecha: 2019.04.03 16:14:55 +02'00'

3.2. Written informed consent forms and information sheets for participating in the experiments (Spanish)

El programa de investigación e innovación de la Unión Europea Horizon 2020 ha financiado este proyecto de acuerdo con el convenio de subvención número 761974.

INFORMACIÓN

Proyecto: ImAc (Immersive Accessibility, en castellano, Accesibilidad Inmersiva)

Investigador principal: Sergi Fernández (i2Cat)
Asesora sobre temas relacionados con ética: Pilar Orero

El objetivo de estos experimentos es recopilar información sobre cómo se pueden implementar los servicios de accesibilidad en medios inmersivos. Esto permitirá identificar las necesidades de los diferentes tipos de público e investigar cómo se puede mejorar la calidad de la experiencia de usuario y de los servicios ofrecidos.

Durante el “focus group” (grupo focal), se le pedirá que proporcione datos demográficos. A continuación, se le pedirá que visualice o escuche un estímulo determinado y que exprese su opinión sobre diversos aspectos. La sesión no se grabará. Se tomarán notas para recopilar los comentarios y la información discutida durante la sesión. Al final, estos comentarios e información se leerán en voz alta para que los apruebe. Estas notas se considerarán el resultado del grupo focal. Si tiene alguna duda sobre el procedimiento, puede realizar las preguntas que estime necesarias.

Si detecta cualquier tipo de molestia o se siente incómodo/a, puede abandonar el grupo en cualquier momento sin necesidad de justificarlo.

Proceda a la lectura del consentimiento informado.

Im Ac Immersive Accessibility	
El programa de investigación e innovación de la Unión Europea Horizon 2020 ha financiado este proyecto de acuerdo con el convenio de subvención número 761974.	

CONSENTIMIENTO (versión escrita)

Proyecto: ImAC (Immersive Accessibility, en castellano, Accesibilidad Inmersiva)

Su participación en los experimentos es totalmente voluntaria.

Puede abandonar el estudio en cualquier momento sin necesidad de justificarlo. En caso de hacerlo, no existirá ninguna repercusión o consecuencia negativa de ningún tipo.

Toda información que proporcione se utilizará en el proyecto de forma anónima.

ImAc es un proyecto europeo liderado por Sergi Fernández, de la empresa i2Cat. Pilar Orero es la asesora sobre temas éticos y es la responsable de los procedimientos éticos. Puede contactar con Pilar Orero enviando un correo a pilar.orero@uab.cat para pedir más información sobre el proyecto o los resultados del mismo.

El investigador que llevará a cabo este experimento es ((nombre y apellidos)).

Si quiere participar en el estudio, firme al final del documento mostrando su conformidad con las siguientes afirmaciones.

- He leído y entendido la información proporcionada para esta investigación o me han leído dicha información.
- He tenido la posibilidad de preguntar sobre la investigación.
- Doy mi consentimiento para formar parte de las sesiones de investigación.

____

Nombre del participante Fecha Firma

____

Nombre del investigador Fecha Firma

3.3. Written informed consent forms for video, photographs or audio recording (Spanish)

UNIVERSITAT AUTÒNOMA DE BARCELONA DOCUMENTO DE CESIÓN DE DERECHOS Y CONSENTIMIENTO SOBRE FOTOGRAFÍAS, VÍDEOS Y GRABACIONES DE SONIDO

Al firmar el presente documento de cesión de derechos y consentimiento sobre fotografías, vídeos y grabaciones de sonido, otorgo de manera irrevocable mi permiso a los responsables asignados de la UAB, a directores, representantes, empleados, sucesores, licenciatarios y cesionarios, para que utilicen fotografías, vídeos y/o grabaciones de sonido de mi persona para el siguiente proyecto: ImAc. Otorgo este permiso de manera totalmente voluntaria.

Mi consentimiento para el uso de fotografías, vídeos y grabaciones de sonido y para el uso de mi imagen, semblante, apariencia y voz es para siempre. No recibiré ninguna compensación por el uso de mi imagen, semblante, apariencia y voz ni ahora ni en el futuro. La universidad puede utilizar fotografías, vídeos y grabaciones de sonido que contengan mi imagen, semblante, apariencia y voz de cualquier modo o en cualquier medio, incluido el uso en páginas web. Las fotografías, vídeos y grabaciones de sonido se podrán usar en su totalidad o en parte, solas o junto con otras grabaciones. Las fotografías, vídeos y grabaciones de sonido se podrán usar con cualquier fin educativo, institucional, científico o informativo, pero jamás con fines comerciales de ningún tipo. La Universidad tiene el derecho a permitir que terceras partes fuera de la Universidad puedan copiar, editar, modificar, retocar, revisar y cambiar de cualquier manera las fotografías, vídeos y grabaciones de sonido a criterio de la Universidad para fines no comerciales. Todos los derechos, títulos e intereses sobre las fotografías, vídeos y grabaciones de sonido pertenecen única y exclusivamente a los responsables asignados de la UAB.

Por el presente, otorgo mi permiso a la Universidad para usar mi nombre, biografía y cualquier otra información personal, hechos o cualquier otro material en relación con tales usos de las fotografías, vídeos y grabaciones de sonido acorde con la ley española de protección de datos.

Comprendo y estoy de acuerdo con las condiciones estipuladas en el presente documento de cesión de derechos y consentimiento sobre fotografías, vídeos y grabaciones de sonido. Otorgo mi consentimiento de manera irrevocable y para siempre a los responsables asignados de la UAB, así como a los directores, representantes, empleados, sucesores, licenciatarios y concesionarios de la Universidad, para hacer uso de mi imagen, semblante, apariencia y voz registrados en las fotografías, vídeos y grabaciones de sonido arriba descritos. Por el presente, reconozco que estoy en pleno uso de mis facultades para entender el contenido de este documento de cesión de derechos y que no presento ninguna discapacidad cognitiva, ni me han sometido a coacción o presión indebida a la hora de firmar el presente documento.

____

Nombre impreso del participante

______________________________ ____________________

Firma del participante Fecha

Acrónimo del proyecto:	**IMAC**
Número del convenio de subvención:	761974
Título del proyecto:	*Immersive Accessibility (en castellano, Accesibilidad Inmersiva)*

3.4. Written consent form online (for subtitle editor online test) (English)

Consent form

Please read the information below and proceed to give your consent to participate in this study if you agree.

* Required

Information sheet

Project: ImAc (Immersive Accessibility)

Main researcher: Sergi Fernández (i2Cat)
Ethical adviser: Pilar Orero

The aim of the tests is to get feed-back on how access services can be implemented in immersive media. This will allow us to identify the needs of diverse audiences and research how the quality of experience and the quality of the service can be improved.

During the test, which can take various forms (experiment with questionnaire, focus groups, interviews, etc.), you will be asked to provide some demographic data. Then, you will be asked to watch an input, perform a task or give your opinion on various aspects. If needed, objective data will be recorded during the session. The researcher will give you more details of the specific test assigned to you and the data collection methods. Please ask as many questions as needed to clarify the procedure.

Virtual reality may produce some sort of discomfort such as virtual reality sickness when visualizing virtual reality contents, information will be provided and appropriate measures will be taken to guarantee the participants' safety and well-being. Immersive environments are not recommended for individuals with claustrophobia, heart conditions, back conditions, a history of seizures, epilepsy, and/or sensitivity to flashing lights. Also participants thought to be unstable or under the influence of drugs or alcohol will not be admitted.

In the case that some physiological or eye-tracking apparatus are used to gather data, you will not experience any discomfort, since the apparatus used are the latest generation and are not invasive.

Now please read the consent form.

Consent form

Project: ImAC (Immersive Accessibility)

Your participation in the tests is absolutely voluntary.

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.

In the case that some physiological or eye-tracking apparatus are used to gather data, you will not experience any discomfort, since the apparatus used are the latest generation and are not invasive.

Virtual reality may produce some sort of discomfort such as virtual reality sickness when visualizing virtual reality contents, information will be provided and appropriate measures will be taken to guarantee the participants' safety and well-being. Immersive environments are not recommended for individuals with claustrophobia, heart conditions, back conditions, a history of seizures, epilepsy, and/or sensitivity to flashing lights. Also participants thought to be unstable or under the influence of drugs or alcohol will not be admitted.

The information you provide will be used in the project but it will remain anonymous.

ImAc is a European project led by Sergi Fernández, from the company i2Cat. 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 BELÉN AGULLÓ. In case of doubts before or after the test, you can send an email to belen.agullo@uab.cat.

If you are willing to participate, please confirm the following statements by selecting the "yes" button at the end of this form. If you select "no" it means you do not give your consent.

- 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.

1. **Please, enter your name and surname:** *

____

2. **Do you give your consent to participate in this test?** *

   Mark only one oval.

   ◯ Yes

   ◯ No *Stop filling out this form.*

Annex 4 – Questionnaires

4.1. Demographic questionnaire – Focus group (Catalan)
4.2. Demographic questionnaire – For article "From disabilities to capabilities: testing subtitles in immersive environments with end users"
4.3. Demographic questionnaire – Web editor test
4.4. Postquestionnaire – Web editor test
4.5. Demographic questionnaire – For article "Making interaction with virtual reality accessible: rendering and guiding methods for subtitles"
4.6. Postquestionnaire (behavior) – For article "Making interaction with virtual reality accessible: rendering and guiding methods for subtitles"
4.7. Postquestionnaire (guiding) – For article "Making interaction with virtual reality accessible: rendering and guiding methods for subtitles"
4.8. IPQ Questionnaire (Spanish)
4.9. Demographic questionnaire – For article "Subtitles in virtual reality: a reception study"
4.10. Postquestionnaire (behavior) – For article "Subtitles in virtual reality: a reception study"
4.11. Postquestionnaire (guiding) – For article "Subtitles in virtual reality: a reception study"

4.1. Demographic questionnaire – Focus group (Catalan)

Immersive Accessibility		T1.2. Qüestionari general
El programa de recerca i innovació de la Unió Europea Horizon 2020 ha finançat aquest projecte d'acord amb el conveni de subvenció número 761974.		

Preguntes generals

1. Sexe

a) Dona
b) Home
c) Altre
d) Prefereixo no respondre

2. Edat _________________

3. Llengua principal _____________

4. Nivell d'estudis acabats

a) Sense estudis
b) Primària
c) Secundària
d) Estudis superiors
e) Universitat

5. Em defineixo com una persona...

a) cega
b) amb baixa visió
c) sorda
d) amb pèrdua d'audició
e) sordcega

6. Edat d'inici de la discapacitat

a) De naixement
b) 0-4 anys
c) 5-12 anys
d) 13-20 anys
e) 21-40 anys
f) 41-60 anys
g) Més de 60 anys

7. Quines tecnologies utilitza diàriament? Pot seleccionar més d'una opció.

a) TV
b) Ordinador de taula
c) Ordinador portàtil
d) Telèfon mòbil
e) Tauleta

8. Disposa d'algun dispositiu per accedir a contingut de realitat virtual?

a) Sí (Si és que sí, quin? __________)
b) No
c) No ho sé o no vull contestar

9. Quin és el seu dispositiu preferit per veure vídeos en línia (per exemple, a Youtube, Vimeo, Netflix, Amazon Prime, vídeos a la carta, etc.)?

a) Ordinador de taula
b) Ordinador portàtil
c) Telèfon mòbil
d) Tauleta
e) No accedeixo a continguts en línia
f) Altres (especifiqui quins:________)

10. (només per a persones cegues o amb baixa visió) Què fa servir per accedir als continguts en línia?

a) Magnificador (per exemple, ZoomText)
b) Lector de pantalla (per exemple, JAWS, VoiceOver, TalkBack)
c) Tots dos
d) Cap

12. (només per a persones cegues o amb baixa visió) Quins dels controls següents li agradaria fer servir amb el magnificador o lector de pantalla quan veu continguts en línia?

a) Explorar la biblioteca de continguts
b) Identificar contingut
c) Funcions como reproduir ("play"), pausa ("stop"), avançar, retrocedir
d) Activar i desactivar l'audiodescripció i els audiosubtítols

4.2. Demographic questionnaire – For article “From disabilities to capabilities: testing subtitles in immersive environments with end users”

Immersive Accessibility		**T1.2. General questionnaire**
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 761974		

1. **Edad: _______________**

2. **Nivel de estudios finalizados**
   a) Sin estudios
   b) Educación primaria
   c) Educación secundaria
   d) Educación superior
   e) Universidad

3. **¿Cuál de estos dispositivos usas a diario? Puedes elegir más de una opción.**
   a) Televisión
   b) Ordenador de sobremesa
   c) Ordenador portátil
   d) Teléfono móvil
   e) Tablet
   f) Gafas de realidad virtual
   g) Consola
   h) Otros:_____________

4. **¿Con qué frecuencia ves contenido de realidad virtual (por ejemplo, vídeos 360º)?**

	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					

5. Si nunca has visto contenido en realidad virtual como vídeos 360º o solo lo has hecho en alguna ocasión, indica el motivo. Puedes elegir más de una opción.

a) Porque no me interesa.
b) Porque no es accesible.
c) Porque no he tenido la oportunidad de usarlo.
d) Otros motivos. Por favor, explícalos: ___________________

6. Indica tu nivel de acuerdo con la siguiente afirmación: «Me interesa el contenido en realidad virtual (por ejemplo, vídeos 360º)».

a) Muy de acuerdo
b) De acuerdo
c) Ni de acuerdo ni en desacuerdo
d) En desacuerdo
e) Muy en desacuerdo

7. ¿Tienes algún dispositivo para acceder a contenido de realidad virtual?

a) Sí (Si es que sí, ¿cuál? ___________)
b) No
c) No lo sé o no quiero responder

8. ¿Te gusta ver los siguientes tipos de contenido en televisión o en línea?

	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					

9. Cuando hay subtítulos disponibles, ¿los activas para los siguientes tipos de contenido?

	Siempre	A veces	Alguna vez	Nunca
Noticias				
Ficción (series, películas)				
Programas de entrevistas				
Documentales				
Deportes				
Dibujos animados				

10. Si los subtítulos están disponibles y no los activas, indica los motivos.

a) Porque la interfaz no es accesible.
b) Porque no quiero subtítulos en todos los contenidos, solo en algunos contenidos.
c) Por otros motivos. Explica cuáles: ______________.

11. ¿Cuántas horas al día ves contenido subtitulado?

a) Ninguna
b) Menos de una hora
c) De una a dos horas
d) De dos a tres horas
e) De tres a cuatro horas
f) Cuatro horas o más

12. ¿Para qué usas los subtítulos?

a) Me ayudan a entender.
b) Son mi única manera de acceder al diálogo.
c) Los uso para aprender idiomas.
d) Otros motivos. Indícalos: ______________

13. Si tienes más comentarios, puedes incluirlos aquí:

4.3. Demographic questionnaire – Web editor test

Some questions about yourself

Please reply to these general questions about yourself.

* Required

1. **Please, enter you user code (provided in the e-mail):** *

2. **1. Sex:** *

   Mark only one oval.

   - Female
   - Male
   - Other
   - I prefer not to reply

3. **2. Age:** *

4. **3. Main language:** *

5. **4. Please, describe your current job:** *

6. **5. Have you ever subtitled a 360º video?** *

   Mark only one oval.

   - Yes
   - No

7. **6. For how long have you been working in the field of subtitling?** *

8. **7. How many hours of subtitling have you produced in your professional life?** *

Mark only one oval.

- Less than 50 hours
- 51-150 hours
- 151-300 hours
- More than 300 hours

9. **8. In what language or languages do you normally subtitle?** *

10. **9. What software do you normally use?** *

11. **10. Please indicate your level of studies.** *

Mark only one oval.

- Primary education
- Secondary education
- Further education
- University

12. **11. If you replied "Further education" or "University" in the previous question, please specify.**

13. **12. If you have received specific training on subtitling, please indicate it here.** *

14. **13. What devices do you use on a daily basis? Multiple replies are possible.** *

Check all that apply.

- [ ] TV
- [ ] PC
- [ ] Laptop
- [ ] Mobile phone
- [ ] Tablet
- [ ] Head Mounted Display
- [ ] Other: ____________________

15. **14. How often do you watch virtual reality content (for instance, 360º videos)?** *

Mark only one oval per row.

	Never	Occasionally	At least once a month	At least once a week	Every day
In smartphone	⬭	⬭	⬭	⬭	⬭
On a tablet	⬭	⬭	⬭	⬭	⬭
On a PC	⬭	⬭	⬭	⬭	⬭
In smartphone plugged to HMD	⬭	⬭	⬭	⬭	⬭
In HMD	⬭	⬭	⬭	⬭	⬭

16. **15. If you have never used virtual reality content such as 360º videos or only occasionally, please indicate why. Multiple answers are possible.** *

Check all that apply.

- [ ] Because I am not interested.
- [ ] Because it is not accessible.
- [ ] Because I have not had the chance to use it.
- [ ] Other: ____________________

17. **16. Please state your level of agreement with the following statement: “I am interested in virtual reality content (such as 360º videos).”** *

Mark only one oval.

- ⬭ I strongly agree
- ⬭ I agree
- ⬭ Neither agree nor disagree
- ⬭ Disagree
- ⬭ Strongly disagree

18. **17. Do you own any device to access virtual reality content?** *

Mark only one oval.

- ⬭ Yes
- ⬭ No
- ⬭ I don’t know or I don’t want to reply

19. **18. If you replied "yes" to the previous question, please specify which device(s).**

____

4.4. Postquestionnaire – Web editor test

Please, provide some feedback about the editor

Please score the following statements.

* Required

1. **Please, enter you user code (provided in the e-mail):** *

1. I think that I would like to use this system frequently *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

2. I found the system unnecessarily complex *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

3. I thought the system was easy to use *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

4. I think that I would need the support of a technical person to be able to use this system *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

5. I found the various functions in this system were well integrated *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

6. I thought there was too much inconsistency in this system *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

7. I would imagine that most people would learn to use this system very quickly *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

8. I found the system very cumbersome to use *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

9. I felt very confident using the system *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

10. I needed to learn a lot of things before I could get going with this system *

Mark only one oval.

	1	2	3	4	5	
strongly disagree	⬭	⬭	⬭	⬭	⬭	strongly agree

More questions...

Please, reply the open questions with your own words. The aim of these questions is to gather feedback to improve the subtitle editor.

11. What did you like most about the subtitle editor?

____

____

____

____

____

12. What did you like less about the subtitle editor?

13. What do you think could be improved, and how?

14. Did you miss any funcionality? If yes, can you tell us which?

15. Do you find the feature for setting the angle for the subtitle easy to use? Explain why.

16. Were the preview modes useful for you? Explain why.

17. Do you think it will take you longer to subtitle videos in 360º? Why?

18. Do you think 360º videos will impact your work as a subtitler?

19. Other comments:

THANK YOU!

We would like to thank you once again for your participation.

If you would like to follow up this research or clarify anything, please send an e-mail to Belén Agulló (belen.agullo@uab.cat).

For further information on ImAc project, you can visit: www.imac-project.eu.

4.5. Demographic questionnaire – For article “Making interaction with virtual reality accessible: rendering and guiding methods for subtitles”

Algunas preguntas sobre ti

Por favor, responde a las siguientes preguntas sobre ti.

* Required

1. **Código de participante:** *

2. **1. Sexo:** *

Mark only one oval.

- Mujer
- Hombre
- Otro
- Prefiero no contestar

3. **2. Edad:** *

4. **3. Lengua materna:** *

5. **4. Por favor, indica tu nivel de estudios.** *

Mark only only one oval.

- Sin estudios
- Educación primaria
- Educación secundaria
- Formación profesional
- Universidad

6. **5. Me defino como....** *

Mark only one oval.

- Persona con pérdida auditiva
- Persona oyente
- Other:

7. **6. Edad en la que empezó tu pérdida auditiva (si procede):** *
Mark only one oval.

- De nacimiento
- De 0 a 4 años
- De 5 a 12 años
- De 13 a 20 años
- De 21 a 40 años
- De 41 a 60 años
- Después de los 60
- No procede

8. **7. ¿Qué dispositivos utilizas a diario? Puedes seleccionar más de una respuesta.** *
Check all that apply.

- Televisión
- Ordenador de sobremesa
- Ordenador portátil
- Teléfono móvil
- Tablet
- Gafas de realidad virtual (tipo Oculus Rift/PlayStation VR o Google Cardboard/Samsung VR)
- Videoconsola
- Other: ____________________

9. **8. ¿Cada cuánto ves contenidos de realidad virtual (por ejemplo, vídeos de 360º)?** *
Mark only one oval per row.

	Nunca	Ocasionalmente	Al menos una vez al mes	Al menos una vez a la semana	Todos los días
En el teléfono móvil					
En la tablet					
En el ordenador					
En el teléfono móvil conectado a unas gafas de realidad virtual (tipo Google Cardboard o Samsung VR)					
En unas gafas de realidad virtual (tipo Oculus Rift o PlayStation VR)					

10. **9. Si nunca has visto contenido en realidad virtual como vídeos de 360º o solo lo has hecho en alguna ocasión, indica el motivo. Puedes escoger más de una respuesta.** *
Check all that apply.

- Porque no me interesa.
- Porque no es accesible.
- Porque no he tenido la oportunidad de hacerlo.
- Other: ____________________

11. **10. Por favor, indica tu nivel de acuerdo con la siguiente afirmación: "Me interesa el contenido de realidad virtual (como vídeos de 360º)".** *

Mark only one oval.

- Muy de acuerdo
- De acuerdo
- Ni de acuerdo ni en desacuerdo
- En desacuerdo
- Muy en desacuerdo

12. **11. ¿Tienes algún dispositivo para acceder contenido de realidad virtual (por ejemplo, Google Cardboard, Samsung VR, Oculus Rift, PlayStation VR, etc.)?** *

Mark only one oval.

- Sí
- No
- No lo sé o no quiero contestar

13. **12. Si has respondido "sí" en la pregunta anterior, especifica qué dispositivo(s):**

____

14. **13. ¿Qué tipo de contenidos te gusta ver en televisión o en línea?** *

Mark only one oval per row.

	Me gusta mucho	Me gusta	Ni me gusta ni me disgusta	No me gusta	No me gusta nada
Noticias					
Ficción (series, películas)					
Programas de debate					
Documentales					
Deportes					
Dibujos animados					

15. **14. Si hay subtítulos disponibles, ¿los activas para los siguientes contenidos?** *

Mark only one oval per row.

	Siempre	A veces	Pocas veces	nunca
Noticias				
Ficción (series, películas)				
Programas de debate				
Documentales				
Deportes				
Dibujos animados				

16. **15. Si los subtítulos están disponibles y no los activas, ¿por qué no? Selecciona los motivos.** *

Mark only one oval.

- Porque la interfaz no es accesible.
- Porque no quiero subtítulos en todos los contenidos, solo en ciertos tipos de contenido.
- Porque no los necesito.
- Los activo siempre.
- Other: ______________________

17. **16. ¿Cuántas horas al día pasas viendo contenido subtitulado?** *

Mark only one oval.

- Ninguna
- Menos de una hora
- Entre 1 y 2 horas
- Entre 2 y 3 horas
- Entre 3 y 4 horas
- 4 horas o más

18. **17. ¿Para qué utilizas los subtítulos?** *

Mark only one oval.

- Me ayudan a entender.
- Son mi única manera de acceder a los diálogos.
- Los uso para aprender idiomas.
- No los uso nunca.
- Other: ______________________

4.6. Postquestionnaire (behavior) – For article "Making interaction with virtual reality accessible: rendering and guiding methods for subtitles"

Por favor, comenta tu experiencia con los vídeos y los subtítulos

Por favor, responde a las preguntas con tus propias palabras.

* Required

1. **Código de participante:** *

2. **1. ¿Qué sistema prefieres para leer los subtítulos en vídeos de 360º?** *
   Mark only one oval.
   - Subtítulos siempre visibles
   - Subtítulos colocados en posiciones fijas

3. **2. Explica con tus propias palabras por qué prefieres la opción seleccionada en la pregunta anterior.** *

4. **3. ¿Qué crees que se podría mejorar al colocar los subtítulos en vídeos de 360º y cómo lo mejorarías?** *

5. **4a. ¿Te ha parecido fácil encontrar los subtítulos siempre visibles?** *
   Mark only one oval.

	1	2	3	4	5	6	7	
muy difícl								muy fácil

6. **4b. ¿Te ha parecido fácil leer los subtítulos siempre visibles?** *
   Mark only one oval.

	1	2	3	4	5	6	7	
muy difícl								muy fácil

7. **5a. ¿Te ha parecido fácil encontrar los subtítulos colocados en posiciones fijas?** *
Mark only one oval.

	1	2	3	4	5	6	7	
muy difícil	◯	◯	◯	◯	◯	◯	◯	muy fácil

8. **5b. ¿Te ha parecido fácil leer los subtítulos colocados en posiciones fijas?** *
Mark only one oval.

	1	2	3	4	5	6	7	
muy difícil	◯	◯	◯	◯	◯	◯	◯	muy fácil

9. **6. ¿Crees que los subtítulos siempre visibles obstruyen partes importantes de la imagen?** *
Mark only one oval.

◯ Sí
◯ No

10. **7. ¿Crees que los subtítulos colocados en posiciones fijas obstruyen partes importantes de la imagen?** *
Mark only one oval.

◯ Sí
◯ No

11. **8. Otros comentarios:** *

____
____
____
____
____

4.7. Postquestionnaire (guiding) – For article “Making interaction with virtual reality accessible: rendering and guiding methods for subtitles”

Por favor, comenta tu experiencia con los vídeos y los subtítulos

Por favor, responde a las preguntas con tus propias palabras.

* Required

1. **Código de participante: ***

   ____

2. **1. ¿Qué sistema prefieres para que te indiquen la localización del personaje que habla? ***

   Mark only one oval.

   - ◯ Flechas
   - ◯ Posicionamiento automático

3. **2. Explica con tus propias palabras por qué prefieres la opción seleccionada en la pregunta anterior. ***

   ____
   ____
   ____
   ____
   ____

4. **3. ¿Qué crees que se podría mejorar al indicar dónde está el personaje que habla en vídeos de 360º y cómo lo mejorarías? ***

   ____
   ____
   ____
   ____
   ____

5. **4. ¿Te ha parecido fácil encontrar a la persona que habla con las flechas? ***

   Mark only one oval.

	1	2	3	4	5	6	7	
muy difícl	◯	◯	◯	◯	◯	◯	◯	muy fácil

6. **5. ¿Te ha parecido fácil encontrar a la persona que habla con el sistema de posicionamiento automático? ***

   Mark only one oval.

	1	2	3	4	5	6	7	
muy difícil	◯	◯	◯	◯	◯	◯	◯	muy fácil

7. **6. ¿Te ha parecido que las flechas te distraían de lo que estaba ocurriendo en el vídeo?** *

Mark only one oval.

- Sí
- No

8. **7. ¿Te ha parecido que el sistema de posicionamiento automático te distraía de lo que estaba ocurriendo en el vídeo?** *

Mark only one oval.

- Sí
- No

9. **8. Otros comentarios:** *

4.8. IPQ Questionnaire (Spanish)

Por favor, comenta tu experiencia.

Por favor, puntúa las siguientes preguntas o afirmaciones.
* Required

Código de participante: *

1. En el mundo virtual del ordenador, he tenido la sensación de "estar dentro". *

Mark only one oval.

	1	2	3	4	5	6	7	
nada	◯	◯	◯	◯	◯	◯	◯	muchísimo

2. He sentido que el mundo virtual me rodeaba. *

Mark only one oval.

	1	2	3	4	5	6	7	
totalmente en desacuerdo	◯	◯	◯	◯	◯	◯	◯	totalmente de acuerdo

3. Parecía que estaba viendo fotografías. *

Mark only one oval.

	1	2	3	4	5	6	7	
totalmente en desacuerdo	◯	◯	◯	◯	◯	◯	◯	totalmente de acuerdo

4. No me he sentido presente en el espacio virtual. *

Mark only one oval.

	1	2	3	4	5	6	7	
no me he sentido presente	◯	◯	◯	◯	◯	◯	◯	me he sentido presente

5. He tenido la sensación de estar dentro del espacio virtual, en lugar de observar el espacio virtual desde fuera. *

Mark only one oval.

	1	2	3	4	5	6	7	
totalmente en desacuerdo	◯	◯	◯	◯	◯	◯	◯	totalmente de acuerdo

6. Me he sentido presente en el espacio virtual. *

Mark only one oval.

	1	2	3	4	5	6	7	
totalmente en desacuerdo	◯	◯	◯	◯	◯	◯	◯	totalmente de acuerdo

7. ¿Te dabas cuenta del mundo real que te rodeaba cuando navegabas por el mundo virtual (por ejemplo, sonidos, la temperatura de la habitación, otras personas, etc.)? *

Mark only one oval.

	1	2	3	4	5	6	7	
sí me daba cuenta	◯	◯	◯	◯	◯	◯	◯	no me daba cuenta

8. No me daba cuenta del entorno real que me rodeaba. *

Mark only one oval.

	1	2	3	4	5	6	7	
sí me daba cuenta	◯	◯	◯	◯	◯	◯	◯	no me daba cuenta

9. He seguido atento/atenta al mundo real que me rodeaba. *

Mark only one oval.

	1	2	3	4	5	6	7	
totalmente en desacuerdo	◯	◯	◯	◯	◯	◯	◯	totalmente de acuerdo

10. Estaba totalmente cautivado/a por el mundo virtual. *

Mark only one oval.

	1	2	3	4	5	6	7	
totalmente en desacuerdo	◯	◯	◯	◯	◯	◯	◯	totalmente de acuerdo

11. ¿Te ha parecido real el mundo virtual? *

Mark only one oval.

	1	2	3	4	5	6	7	
completamente real	◯	◯	◯	◯	◯	◯	◯	nada real

12. ¿Te parece que se puede comparar la experiencia del mundo virtual con la experiencia del mundo real? *

Mark only one oval.

	1	2	3	4	5	6	7	
no se puede comparar	◯	◯	◯	◯	◯	◯	◯	sí se puede comparar

13. ¿Te ha parecido real el mundo virtual? *

Mark only one oval.

	1	2	3	4	5	6	7	
igual de real que un mundo imaginado	◯	◯	◯	◯	◯	◯	◯	no se puede diferenciar del mundo real

14. El mundo virtual parece más realista que el mundo real. *

Mark only one oval.

	1	2	3	4	5	6	7	
totalmente en desacuerdo	◯	◯	◯	◯	◯	◯	◯	totalmente de acuerdo

4.9. Demographic questionnaire – For article “Subtitles in virtual reality: a reception study”

Algunas preguntas sobre ti

Por favor, responde a las siguientes preguntas sobre ti.

* Required

1. **Código de participante:** *

2. **1. Sexo:** *

Mark only one oval.

- Mujer
- Hombre
- Otro
- Prefiero no contestar

3. **2. Edad:** *

4. **3. Lengua materna:** *

5. **4. Por favor, indica tu nivel de estudios.** *

Mark only one oval.

- Sin estudios
- Educación primaria
- Educación secundaria
- Formación profesional
- Universidad

6. **5. Me defino como....** *

Mark only one oval.

- Persona sorda
- Persona oyente
- Other: ______________

7. **6. Edad en la que empezó tu pérdida auditiva (si eres una persona sorda):** *

Mark only one oval.

- De nacimiento
- De 0 a 4 años
- De 5 a 12 años
- De 13 a 20 años
- De 21 a 40 años
- De 41 a 60 años
- Después de los 60
- No soy una persona sorda

8. **7. ¿Qué aparatos utilizas todos los días? Puedes poner más de una respuesta.** *

Check all that apply.

- Televisión
- Ordenador de sobremesa
- Ordenador portátil
- Teléfono móvil
- Tablet
- Gafas de realidad virtual (tipo Oculus Rift/PlayStation VR o Google Cardboard/Samsung VR)
- Videoconsola
- Other: ______________________

9. **8. ¿Cuántas veces ves contenidos de realidad virtual (por ejemplo, vídeos de 360º)?** *

Mark only one oval per row.

	Nunca	A veces	Una vez al mes	Una vez a la semana	Todos los días
En el teléfono móvil					
En la tablet					
En el ordenador					
En el teléfono móvil conectado a unas gafas de realidad virtual (tipo Google Cardboard o Samsung VR)					
En unas gafas de realidad virtual (tipo Oculus Rift o PlayStation VR)					

10. **9. Si has visto pocas veces o nunca contenido de realidad virtual, di por qué. Puedes poner más de una respuesta.** *

Check all that apply.

- Porque no me interesa.
- Porque no es accesible.
- Porque no he tenido la oportunidad de ver contenido de realidad virtual.
- Sí he visto contenidos en realidad virtual.
- Other: ______________________

11. **10. Por favor, di si estás de acuerdo con la siguiente frase: “Me interesa el contenido de realidad virtual (como vídeos de 360º)”.** *

Mark only one oval.

- Muy de acuerdo
- De acuerdo
- Ni de acuerdo ni en desacuerdo
- En desacuerdo
- Muy en desacuerdo

12. **11. ¿Tienes algún aparato para ver contenido de realidad virtual (por ejemplo, Google Cardboard, Samsung VR, Oculus Rift, PlayStation VR, etc.)?** *

Mark only one oval.

- Sí
- No
- No lo sé o no quiero contestar

13. **12. Si has respondido “sí” en la pregunta anterior, especifica qué aparatos(s):**

____

14. **13. ¿Qué contenidos te gusta ver en televisión o en línea?** *

Mark only one oval per row.

	Me gusta mucho	Me gusta	Ni me gusta ni me disgusta	No me gusta	No me gusta nada
Noticias	◯	◯	◯	◯	◯
Ficción (series, películas)	◯	◯	◯	◯	◯
Programas de debate	◯	◯	◯	◯	◯
Documentales	◯	◯	◯	◯	◯
Deportes	◯	◯	◯	◯	◯
Dibujos animados	◯	◯	◯	◯	◯

15. **14. Si tienes subtítulos, ¿cuándo pones los subtítulos?** *

Mark only one oval per row.

	Siempre	A veces	Pocas veces	nunca
Noticias	◯	◯	◯	◯
Ficción (series, películas)	◯	◯	◯	◯
Programas de debate	◯	◯	◯	◯
Documentales	◯	◯	◯	◯
Deportes	◯	◯	◯	◯
Dibujos animados	◯	◯	◯	◯

16. **15. ¿Por qué no pones los subtítulos cuando los tienes? Di el motivo.** *
Mark only one oval.

- Porque la interfaz no es accesible.
- Porque solo quiero subtítulos en algunos programas.
- Porque no necesito subtítulos.
- Siempre activo los subtítulos.
- Other: ____________________

17. **16. ¿Cuántas horas al día ves programas con subtítulos?** *
Mark only one oval.

- Ninguna
- Menos de una hora
- Entre 1 y 2 horas
- Entre 2 y 3 horas
- Entre 3 y 4 horas
- 4 horas o más

18. **17. ¿Para qué utilizas los subtítulos? Puedes poner más de una respuesta.** *
Check all that apply.

- [ ] Me ayudan a entender.
- [ ] Son mi única manera de ver los diálogos.
- [ ] Uso subtítulos para aprender idiomas.
- [ ] No uso subtítulos nunca.
- [ ] Other: ____________________

4.10. Postquestionnaire (behavior) – For article "Subtitles in virtual reality: a reception study"

Por favor, comenta tu experiencia con los vídeos y los subtítulos

Por favor, responde a las preguntas con tus propias palabras.

* Required

1. **Código de participante:** *

2. **1. ¿Qué sistema prefieres para leer los subtítulos en vídeos de 360º?** *
   Mark only one oval.
   - Subtítulos siempre visibles
   - Subtítulos colocados en tres posiciones fijas

3. **2. Explica con tus propias palabras por qué prefieres la opción seleccionada en la pregunta anterior.** *

4. **3. ¿Qué crees que se podría mejorar para colocar los subtítulos en vídeos de 360º y cómo lo mejorarías?** *

5. **4a. ¿Te ha parecido fácil encontrar los subtítulos siempre visibles?** *
   Mark only one oval.

	1	2	3	4	5	6	7	
muy difícil								muy fácil

6. **4b. ¿Te ha parecido fácil leer los subtítulos siempre visibles?** *
   Mark only one oval.

	1	2	3	4	5	6	7	
muy difícl								muy fácil

7. **5a. ¿Te ha parecido fácil encontrar los subtítulos colocados en tres posiciones fijas?** *
Mark only one oval.

	1	2	3	4	5	6	7	
muy difícil	◯	◯	◯	◯	◯	◯	◯	muy fácil

8. **5b. ¿Te ha parecido fácil leer los subtítulos colocados en tres posiciones fijas?** *
Mark only one oval.

	1	2	3	4	5	6	7	
muy difícil	◯	◯	◯	◯	◯	◯	◯	muy fácil

9. **6. ¿Crees que los subtítulos siempre visibles tapan partes importantes de la imagen?** *
Mark only one oval.

	1	2	3	4	5	6	7	
tapan mucho	◯	◯	◯	◯	◯	◯	◯	no tapan nada

10. **7. ¿Crees que los subtítulos colocados en tres posiciones fijas tapan partes importantes de la imagen?** *
Mark only one oval.

	1	2	3	4	5	6	7	
tapan mucho	◯	◯	◯	◯	◯	◯	◯	no tapan nada

11. **8. ¿Crees que los subtítulos siempre visibles te distraen del vídeo?** *
Mark only one oval.

	1	2	3	4	5	6	7	
sí me distraen mucho	◯	◯	◯	◯	◯	◯	◯	no me distraen nada

12. **9. ¿Crees que los subtítulos colocados en tres posiciones fijas te distraen del vídeo?** *
Mark only one oval.

	1	2	3	4	5	6	7	
sí me distraen mucho	◯	◯	◯	◯	◯	◯	◯	no me distraen nada

13. **8. Otros comentarios:** *

____

____

____

____

____

4.11. Postquestionnaire (guiding) – For article “Subtitles in virtual reality: a reception study”

Por favor, comenta tu experiencia con los vídeos y los subtítulos

Por favor, responde a las preguntas con tus propias palabras.

* Required

1. **Código de participante:** *

2. **1. ¿Qué sistema prefieres para que te indiquen dónde está la persona que habla?** *
   Mark only one oval.
   - Flechas
   - Radar

3. **2. Explica con tus propias palabras por qué prefieres la opción seleccionada en la pregunta anterior.** *

4. **3. ¿Qué crees que se podría mejorar para indicar dónde está el personaje que habla en vídeos de 360º y cómo lo mejorarías?** *

5. **4. ¿Te ha parecido fácil encontrar a la persona que habla con las flechas?** *
   Mark only one oval.

	1	2	3	4	5	6	7	
muy difícil								muy fácil

6. **5. ¿Te ha parecido fácil encontrar a la persona que habla con el radar?** *
   Mark only one oval.

	1	2	3	4	5	6	7	
muy difícil								muy fácil

7. **6. ¿Te ha parecido que las flechas te distraían de las cosas que estaban pasando en el vídeo?** *

Mark only one oval.

	1	2	3	4	5	6	7	
me distraían mucho	⬭	⬭	⬭	⬭	⬭	⬭	⬭	no me distraían nada

8. **7. ¿Te ha parecido que el radar te distraía de las cosas que estaban pasando en el vídeo?** *

Mark only one oval.

	1	2	3	4	5	6	7	
me distraía mucho	⬭	⬭	⬭	⬭	⬭	⬭	⬭	no me distraía nada

9. **8. Otros comentarios:** *

____

____

____

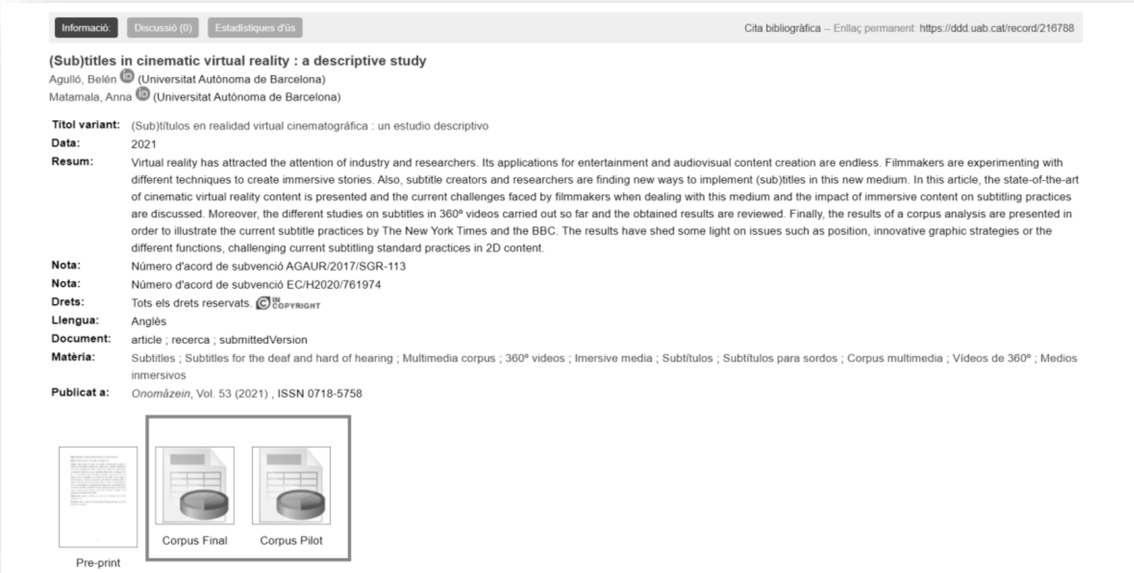
____

____

Annex 5 – Digital and multimedia resources

5.1. Multimedia corpus

a) Excel sheet – Pilot version: https://ddd.uab.cat/record/216788
b) Excel sheet – Final version: https://ddd.uab.cat/record/216788



5.2. Videos for the reception study

Available in the ImAc server: http://84.88.32.46/UAB-test4/