Universitat Autònoma de Barcelona

360° content made accessible:
A user-centered study on audio description

Anita Fidyka

UNIVERSITAT AUTÒNOMA DE BARCELONA

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

Doctorat en Traducció i Estudis Interculturals

360° content made accessible:

A user-centered study on audio description

PhD thesis presented by:
Anita Fidyka

Supervised by:
Dr. Anna Matamala

2021

Moim Rodzicom
Ewie i Rafałowi

Acknowledgements

There are many people whom I would like to thank warmly for their support.

Most of all, I would like to express my sincere gratitude to my supervisor, dr. Anna Matamala, for her knowledge, excellent academic guidance, patience and kind words of encouragement in the last three years.

I am also very grateful to dr. Pilar Orero for letting me become a part of Transmedia Catalonia Research Group and part of the ImAc project team, which allowed me to broaden my knowledge on accessibility and develop as a researcher.

I owe my warmest thanks to B1B2B3 Associació Discapacitat Visual Catalunya for their support in finding participants for my reception studies. In particular, to Manel Martí, Anna, Salud, Dani and Felipe who welcomed me warmly at their association.

My most sincere thanks to everyone with sight loss who participated in my studies, for their patience, their positive energy and engagement. I would also like to thank dr. Blanca Arias Badia and dr. Belén Agulló for their linguistic support during these tests.

I am also grateful to dr. Anna Jankowska, dr. Gian Maria Greco and my MRA colleagues for our inspiring research discussions and precious time shared in Barcelona.

On a personal note, I would like to thank my parents who always encouraged and supported my education and education of my siblings. I would also like to thank my siblings Magda and Krzysztof for being a source of inspiration for me.

Last but not least, I would like to thank my colleagues from Fundacja Siódmy Zmysł (The Seventh Sense Foundation) for their continuous support and friendship.

This research was supported by the European project Immersive Accessibility (ImAc) [grant agreement 761974]. The author 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). The author of this thesis is grateful to Ryot and Jaunt VR for letting ImAc partners use “Holy Land” for the studies conducted within the project.

Table of contents

Index of Tables

Index of Figures

Acronym Glossary

2D	Two-dimensional
3D	Three-dimensional
AD	Audio description
AVSMD	Audiovisual Services Media Directive
AVTS	Audiovisual Translation Studies
ARSAD	Advanced Research Seminar on Audio Description
AST	Audio subtitles
DOF	Degrees of freedom
EAA	European Accessibility Act
EBU	European Broadcasting Union
EEG	Electroencephalography
HMD	Head-mounted display
ICT-SOPI	ICT-Sense of Presence Inventory
ImAc	Immersive Accessibility project
IPQ	Igroup Presence Questionnaire
MA	Media Accessibility
SDH	Subtitles for the deaf and hard-of-hearing
VE	Virtual environments
VR	Virtual Reality

Chapter 1. Introduction

1. Introduction

> Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits.
>
> Article 27 (1) of the Universal Declaration of Human Rights

> 2. States Parties shall also take appropriate measures:
> g) To promote access for persons with disabilities to new information and communications technologies and systems, including the Internet.
> h) To promote the design, development, production and distribution of accessible information and communications technologies and systems at an early stage, so that these technologies and systems become accessible at minimum cost.
>
> 9th article of the UN's Convention on the Rights of Persons with Disabilities

> The right of persons with a disability and of the elderly to participate and be integrated in the social and cultural life of the Union is inextricably linked to the provision of accessible audiovisual media services. The means to achieve accessibility should include, but need not be limited to, sign language, subtitling, audio-description and easily understandable menu navigation.
>
> Consideration 46, Audiovisual Media Services Directive (2010/13/EU)

Due to sustained efforts over the years in research and legislation, there is a rapidly growing awareness towards accessibility. The way we think about accessibility has also changed over the past decades, moving away from the approaches that interpret it as a domain concerning only specific groups towards viewing accessibility as concerning all human beings.

Since the beginning of the information age, one of the areas towards which most attention has been given to is the accessibility of media products, services and environments. Especially since 2018, considerable progress has taken place, with the adoption of European Accessibility Act (EAA) and the publication of a revised version of the Audiovisual Services Media Directive (AVSMD), which is the final step in the legislative

process[1]. Article 7 of this directive establishes that TV channels, both public and commercial, as well as on-demand platforms have to make their services 'continuously and progressively more accessible to people with disabilities', which is an important step towards ensuring that the rights to access to information and culture are met for all.

One of the services employed to render audiovisual media products accessible is audio description (AD) – an additional audio track inserted between dialogues, songs, or other important sounds of an audiovisual production. Audio description can be described as painting pictures in words (Neves 2012), providing a verbal description of the visual layer of the production for those who do not have access to the visuals, or who can access them only partially, and as the result have to rely on the audio track in order to understand the plot, and enjoy it.

Most recently, rapid developments have been observed in the field of virtual environments. After many years of research and development, immersive media are emerging on the market (ECORYS 2017), and are already proposing a wide scope of new media formats (Allen and Tucker 2018). According the European Broadcasting Union (EBU) report (2017), 49% of its members develop or plan to develop immersive content. Some initial efforts have already been invested into researching how Virtual Reality (VR) can be made accessible technology-wise (Zhao et al. 2019). Nevertheless, exploratory research is needed to make immersive content accessible.

In the view of this need, this thesis explores the possible ways of implementing AD in content belonging to 360° videos with three degrees of freedom (3DOF), one of the novel media formats already available on such platforms as the New York Times and the ARTE television network.

This thesis has been developed within the framework of the Immersive Accessibility (ImAc) project, funded by the European Commission under grant agreement 761974. The

[1] Apart from the publication of those two directives, other important documents have been published since 2018, including the 'Design for All' standard (EN 17161:2019), a revised version of the 'Accessibility requirements for ICT products and services' standard (EN 301549:2018) and the implementing acts within the Web Accessibility Directive (Retrieved on 1 September 2020).

aim of this project was to research the possible ways of integrating access services[2] within immersive media, and to test the developed solutions with the actual users. As the result of the project, accessible solutions have also been developed in relation to subtitles for the deaf and hard-of-hearing (Agulló 2019; Agulló, Montagud, Fraile 2019), sign language and audio subtitles. The results of this project are publicly available[3].

1.1. Objectives and hypotheses

This research has two main objectives, which in turn have other specific, associated aims.

1. To define and evaluate different AD presentation modes in 360° videos.
2. To identify the needs of professional audio describers when describing 360° videos.

Presentation modes are defined as different types of AD to which participants with sight loss were exposed to during the pilot study and the main reception study.

Regarding the first main objective, the specific aims are following:

1. To define 360° videos within the taxonomy of immersive environments.
2. To compare standard 2D video content with 360° videos in terms of storytelling.
3. To obtain feedback from AD target users about their needs regarding AD in 360° videos and define different presentation modes.
4. To compare the levels of presence and preferences of AD target users when exposed to different presentation modes.

As it will be explained in Chapter 5, the presentation modes under analysis are Classic, Static and Dynamic in the first stage, and Classic, Radio and Extended in the second stage. In relation to these modes, two hypotheses are forwarded:

[2] The following access services were the focus of the project: audio description, audio subtitles, subtitles for the deaf and hard-of-hearing and sign language.

[3] https://www.imac-project.eu/

H1: Users will report higher presence levels when exposed to Static and Dynamic sound presentation mode, compared to Classic mode. Static and Dynamic will be the preferred options.

H2: Users will report higher presence levels when exposed to Radio and Extended presentation mode, compared to Classic mode. Radio and Extended will be the preferred options.

Regarding the second main objective, linked to identifying the needs of professional describers, the specific aims are:

1. To obtain feedback from professional audio describers about their needs regarding the production of AD in 360° videos.
2. To test an AD editor for describing 360° videos with professional describers.

To attain these aims, a methodological framework will need to be established, as described in section 1.4.

1.2. Thesis structure

This thesis has been written as a compendium of interrelated articles. It is composed of a chapter per article as well as three additional chapters. The final structure of the thesis is described in detail below.

Chapter 1 sets out the objectives and hypotheses for the research (section 1.1), explains the structure of the thesis (section 1.2), discusses the theoretical framework related to the subject developed in this PhD (section 1.3) and presents an introduction of the methodology (section 1.4). Section 1.4 is structured as follows:

— Methodological aspects of the organization of the focus groups (section 1.4.1).
— Methodological aspects of the two usability studies conducted with professional describers (sections 1.4.2 and 1.4.3).
— Methodological aspects of the experimental studies conducted with AD users (section 1.4.4)

Although it is not compulsory to include the section relating to theoretical framework in a thesis by compendium, this section is included to allow for a comprehensive presentation of the general framework within which this thesis is embedded. Similarly, even though not formally required in such a thesis, a section explaining the methodology applied to reach the objectives of this thesis has been included to offer greater clarity.

Chapter 2 includes Article 1: Fidyka, A., Matamala, A. (2018). Audio description in 360° videos: Results from focus groups in Barcelona and Kraków. *Translation Spaces*, 7(2), 285–303. This article presents a taxonomy of immersive environments, defining 360° videos. It also describes the methodology and results of two focus groups, conducted with audio describers and AD users. Its aim was the analysis of the needs of two groups of users related to audio description in 360° videos.

Chapter 3 is based on Article 2: Fidyka, A., Matamala, A. (2021). Retelling narrative in 360º videos: Implications for audio description. *Translation Studies*. This article presents the results of an extensive literature review conducted in order to compare standard narratives with 360° narrative videos and draw some preliminary insights for audio describers. It also presents the results of focus groups in relation to storytelling. The results show some possible approaches to AD for 360° content, including spatial sound and elements of interaction.

Chapter 4 corresponds to Article 3: Fidyka, A., Matamala, A. (2019). Production of access services in immersive content: Understanding the needs of audio describers. *Hikma*, 18(2), 277–300. This article presents the methodology and results of a usability study conducted with audio describers. The aim of this article was to gather feedback on how the functionality of an online tool for audio description of 360° content could be improved, and to conduct an analysis of the needs of professional describers.

Chapter 5 corresponds to Article 4: Fidyka, A., Matamala, A., Soler Vilageliu, O., Arias-Badia, B. (2021). Audio description in 360° content: Results from a reception study. *SKASE*. This article presents the pilot study and main reception study with AD target users, evaluating different AD presentation modes in 360° videos.

Chapter 6 includes a brief summary of the thesis, discussing the most relevant information, as required by all theses by compendium.

Chapter 7 is the concluding chapter and comprises a discussion of the results together with the final conclusions of the experimental studies conducted with AD users (section 7.1) and audio describers (section 7.2), contributions of this thesis, its limitations and the possible paths for future research.

After this, an updated bibliography is provided, unifying the different citation styles under one common format and offering an updated record of the references cited in the articles. Finally, the annexes are presented. They include:

— Annex 1: Articles as published.
— Annex 2: Ethical considerations.
— Annex 3: Documents related to the focus groups.
— Annex 4: Documents related to the usability studies with professional describers.
— Annex 5: Documents related to the reception studies with AD users.

Both paper and electronic annexes are provided. While paper annexes include the articles in their original formatting, ethics forms, questionnaires and other relevant documents, Annex 5.6 contains the link to the presentation modes used in the main study.

1.3. Theoretical framework

This section provides a theoretical framework for the studies conducted within this thesis. Section 1.3.1 contextualizes them within Audiovisual Translation Studies (AVTS), and more specifically within Media Accessibility (MA). Section 1.3.2 reports on the global approach to accessibility taken for all the studies reported within this thesis. The last two sections define two, key concepts used in this PhD: audio description (section 1.3.3) and presence (section 1.3.4). Other concepts, such as ‘virtual environments’, ‘Virtual Reality’ or ‘360° videos’ will not be defined here, to avoid repetition. They are instead defined in Chapter 2, corresponding to Article 1.

1.3.1. Audiovisual Translation and Media Accessibility

This thesis is developed within the framework of Audiovisual Translation Studies (AVTS), and more specifically within Media Accessibility (MA). The beginning of scientific interest

in the translation of audiovisual products can be traced back to the second half of the 20th century, when the first papers were published on the subjects of dubbing and subtitling. The burgeoning field developed further in subsequent years, until the first conference on screen translation was organized, in Stockholm in 1987.

Due to the expansion of media technologies in the 1990s, this branch of Translation Studies has been given much scientific attention, with courses being offered in various training institutions across the world (Orero 2007). Beside subtitles, voice-over and dubbing, a growing number of studies focused on audio description, subtitles for the deaf and hard-of-hearing and other access services. Since recently, they are also researched within Media Accessibility, a field defined as:

> dealing with theories, practices, technologies and instruments that provide access to media products, services and environments for people who cannot, or cannot properly, access content in its original form (Greco 2016).

Media Accessibility researches a variety of access services in order to cater for the individual needs of users of audiovisual products, including the needs of foreign-language speakers, the elderly, immigrants, persons with sight and hearing loss, and persons with cognitive disabilities. Although most scientific attention within the field is given to AD and SDH research, such access services as audio subtitling, respeaking, live subtitling, sign language, surtitling or easy-to-read language are also being investigated. As well as researching various access services, this field focuses on novel approaches to Media Accessibility, such as accessible filmmaking (Romero-Fresco 2013).

Taking into account the ever-changing landscape of Audiovisual Translation, the list of possible access services and approaches is not limited to those mentioned in the paragraph above. Moreover, it can be expected that with the emergence of new products, services and environments, this field of research will expand further, gradually moving from “a niche within Translation Studies to one of the liveliest and fast-growing subdomains of Accessibility Studies” (Greco 2019).

1.3.2. A universalist approach to (Media) Accessibility

Nothing about us without us
Motto of the European Blind Union

In recent years, a shift has taken place in the way accessibility is addressed; moving away from "particularist, maker-centred and reactive approaches" to approaches that are "universalist, user-centred and proactive" (Greco 2019). The studies conducted within this thesis are framed within the universalist approach, which interprets accessibility as concerning not only some specific groups at risk of social exclusion, e.g. persons with sight or hearing loss, but all human beings.

In addition to valuing the knowledge and experience of makers, this approach to accessibility also values the knowledge and experience of the actual users in designing accessible solutions, which stands in contrasts to the previous, maker-centered approach. As such, it allows for the adoption of user-centered research methodologies (ibid.).

It also calls for addressing accessibility at the earliest stages of the creation of the product, service or environment, through the active participation of users and accessibility experts in the design process. Within Media Accessibility, such a refusal to treat accessibility as an add-on can be already seen in Accessible Filmmaking, an approach that integrates access services already in the production process and favours cooperation with the creative team in order to bridge "the maker-user gap" (Greco 2018: 212).

To the best of my knowledge, the ImAc project, within which this thesis is framed was the first project within Media Accessibility in relation to Virtual Reality that adopted this holistic approach and:

— treated accessibility as a requisite by addressing accessibility concerns together with the development of immersive environments.
— cooperated with content creators (broadcasters).
— involved users at each stage of the research process.

1.3.3. Audio description

Audio description is "a verbal commentary providing visual information for those unable to perceive it themselves" (Fryer 2016: 1). It is used extensively in television and film in European countries, and its use is growing rapidly across the world. Apart from granting access to audiovisual media, it is also used, to a lesser extent, in theatres, museums, galleries, operas, among other live settings, and to describe the visual content on websites. Although it emerged initially in 1981 as a response to the needs of persons with sight loss, practice and research (AENOR 2005: 4; Greening 2011; Jankowska et al. 2017b) show that the inclusion of this access service benefits other groups, including elderly, language-learners, immigrants, people with decreased cognitive functions, or even sighted persons who want to gain a more in-depth understanding about the film or a piece of art by listening to an audio introduction or audio description. Of particular relevance to this respect is the opinion of Lou Giansante, the writer of audio tours and producer at Art Beyond Sight[4] who realized that not only persons with sight loss, but also sighted persons would listen to her descriptions, as it enriched their museum experience by giving them a more in-depth understanding about the given piece of art they were presented with.

Audio description is researched within Translation Studies as intersemiotic translation, a concept coined by Jakobson (2009: 41 [1959]). The original definition proposed by Jakobson refers to an interpretation of verbal signs by means of non-verbal signs. If reversed, this definition suits accurately the nature of audio description (Díaz-Cintas 2005). Within Translation Studies, audio description is researched in the framework of Audiovisual Translation Studies (Maszerowska, Matamala and Orero 2014), and more recently within Media Accessibility (Remael, Orero and Caroll 2012), as it provides access to audiovisual content.

In contrast to such modalities as dubbing, voice-over or subtitling, audio description rarely involves interlingual transfer, as its focus is not on the dialogue (Kruger 2012: 71). Drafting audio description is a creative process in which describers do not work on a "pre-existing

[4] http://www.artbeyondsight.org/mei/verbal-description-training/writing-verbal-description-for-audio-guides/

text that needs translating from one language to another" (Fryer 2016: 3), but prepare their scripts from the scratch.

Audio description is one of the most extensively researched modalities within Media Accessibility. As an academic field, it is informed by research from a variety of areas, including linguistics (Braun 2007), narrative studies (Kruger 2010, 2012; Vercauteren 2016), cognitive studies (Fresno 2014) and media psychology (Fryer and Freeman 2012, 2013, 2014; Fryer, Pring and Freeman 2013). Due to this broad scope of research subjects, it is one of the most interdisciplinary fields within Audiovisual Translation Studies (Baker and Saldanha 2019).

Some of the recent research themes undertaken within the AD field include the inclusion of technology in the AD product or in the process of AD creation: some studies research whether AD can be voiced by means of speech synthesis (Fernández-Torné and Matamala 2015), as this solution could minimize costs related to recording the AD script by a human, and some other studies focused on the inclusion of technologies such as machine translation or speech recognition (Ortiz-Boix and Matamala 2013) within the lengthy process of AD creation (Fernández-Torné 2016). Other themes found in recent research into AD are new creation workflows, such as the translation of audio description scripts from other languages (Jankowska 2015) and crowdsourcing audio description (Jankowska 2018), or research related to non-standard approaches to AD scripting (Szarkowska and Wasylczyk 2014; Udo and Fels 2009; Walczak 2017b). As well as researching AD product, process and reception, some studies also focus on the context of audio description (Baker and Saldanha 2019), such as legal aspects (Jankowska et al. 2017a), legislation or guidelines.

Over the last decade, reception studies have proliferated within the AD field (Chmiel and Mazur 2016; Di Giovanni 2018a). Both large-scale (Rai 2009) and small-scale studies have been conducted, particularly within research projects at the international (Mazur and Kruger 2012; Mazur and Chmiel 2012) and national level (Jankowska et al. 2017b), or as part of doctoral theses (Cabeza-Cáceres 2011; Fresno 2014).

Since the first publications on AD, various methodologies have been employed in reception studies (Chmiel and Mazur 2012), learning about users' preferences (Szarkowska and Wasylczyk 2014; Szarkowska and Jankowska 2012), their comprehension (Cabeza-Cáceres 2011) and recall (Fresno 2016), measuring their emotional response to the stimuli

presented (Ramos and Rojo López 2014; Ramos 2015; Ramos Caro 2016) or presence (Fryer and Freeman 2012, 2013, 2014; Walczak and Fryer 2017). Some of the studies complemented subjective data collection with objective methods, such as eye-tracking (Di Giovanni 2014; Orero and Vilaró 2012), electroencephalography (Kruger et al. 2016), or heart rate measurements (Ramos, 2015; Ramos Caro 2016).

1.3.4. Presence and psychological immersion

Because of the immersive nature of film (Wissmath and Weibel 2012), psychological immersion attired the attention of AVT scholars in the recent years. This concept is used as an umbrella term for several other terms deriving from literary studies, media psychology, film studies and information technology. It refers to the experience of the reader or viewer of becoming engrossed by a fictional world (Kruger, Doherty and Ibrahim 2017).

Presence is one of the concepts included within the broad term of psychological immersion, together with such concepts as flow (Wissmath et al. 2009; Wissmath and Weibel 2012) and perceived realism (Cho et al. 2014). Presence originated in the field of Virtual Reality (Steuer 1992) in which the user is surrounded solely by computer-generated images. It describes "the immersion of an audience in a mediated environment to the extent that they experience it as unmediated" (Wilken and Kruger 2016: 258) and it serves as an "experiential quality metric to evaluate, develop, and optimize both advanced broadcast and virtual environment (VE) media systems" (Lessiter et al. 2001: 282). Presence relates in the first instance to the spatial dimension of immersion although it can also refer to the social dimension of having the sense of being in the presence of virtual characters. The experience of presence is not common in traditional media except for cinema (Burch 1979). It is associated with interactive media and the studies conducted to date show that "Virtual Reality or 3D games seem to be a reliable source of this experience" (Schubert, Friedmann and Regenbrecht 2001: 268), as they present users with a three-dimensional space.

This thesis defines presence as "the sense of being in a virtual space that is presented by technological means" (Slater and Wilbur 1997; Witmer and Singer 1998). It follows a four-component model of presence (Regenbrech and Schubert 2002; Schubert 2003; Schubert, Friedmann and Regenbrecht 1999, 2001) that distinguishes between spatial presence, involvement and judgement of realness. Spatial presence, as defined by the authors

(Regenbrecht and Schubert 2002: 427), relates to the sensation that "the own body is actually located in the virtual space". The second component refers to how much of the viewer's attention is absorbed by the VE and how much attention is still paid to the real environment. The third factor assesses the user's judgement as to "whether the VE seems as real as a real environment" (ibid.). The last component is related to the general sense of presence.

In AD reception research, presence has been used extensively in recent years (Fryer and Freeman 2012, 2013, 2014; Fryer, Pring and Freeman 2013; Romero-Fresco and Fryer 2013; Walczak 2017b; Wilken and Kruger 2016) because of an increasing understanding that certain styles of description may have an impact on the extent to which an audience is immersed in the filmic story. The results of the series of studies conducted by Fryer et al. (2012, 2013, 2014), in which the researchers compared the degree to which blind and sighted viewers become immersed in the story, show that audio description enables persons with sight loss to experience at least as high a sense of presence as a sighted audience, and in some cases an even higher sense of presence. Other studies within MA that measured psychological immersion focused on subtitles (Kruger, Doherty and Ibrahim 2017; Kruger, Doherty and Soto-Sanfiel 2017) or subtitles and dubbing (Wissmath et al. 2009).

1.4. Methodology

This thesis follows a user-centered methodology, with a direct and active involvement of participants. For all the studies conducted within this thesis, two profiles of participants were defined:

— Professional users (audio describers).
— Audio description users (persons with sight loss).

Table 1 gives an overview of where the information related to each study is to be found:

Aim	User profile	Method	Number of participants	Location
To obtain feedback from AD target users about their needs regarding AD in 360° videos and define different presentation modes.	AD users	Focus group	6	Article 1 and 2 (Chapters 2 and 3)
To obtain feedback from professional audio describers about their needs regarding the production of AD in 360° videos.	Professional audio describers	Focus group	6	Article 1 and 2 (Chapters 2 and 3)
To test an AD editor for describing 360° videos with professional describers.	Professional audio describers	Usability study	48 (24+24)	Article 3 (Chapter 4)
To compare the levels of presence and preferences of AD target users when exposed to different presentation modes.	AD users	Reception study	30	Article 4 (Chapter 5)

Table 1. Summary of the methodological aspects related to each test

All the studies included within this thesis were carried out following ethical considerations, found in Annex 2. Ethical approval for all the studies was given by the Universitat Autònoma de Barcelona, Spain. All the studies were anonymous, as the questionnaires were coded. Together with the consent forms, they will be securely stored for three years after the completion of the studies.

1.4.1. Focus groups

Two focus groups were conducted: in Barcelona and Kraków, Poland. The aim of conducting focus groups at an early stage of this thesis development was to gather

qualitative feedback from two profiles of users. This qualitative research method, already employed in AVT research (ITC 2000; Iturregui-Gallardo and Méndez-Ulrich 2019; Agulló and Matamala 2019) was chosen for collecting users' data because of two reasons. Firstly, choosing this method allowed researchers to familiarize participants with immersive technology and answer participants' questions regarding various technological aspects. Secondly, it allowed for gathering qualitative feedback from two profiles of users (audio describers and AD users), who confronted each other's views during the discussions in order to reach common conclusions. All participants were presented with an information sheet before the focus group (Annex 2.1) and gave their written consent in order to participate (Annex 2.2).

1.4.1.1. Questionnaires

Before the discussion, the demographic information was gathered by means of the demographic questionnaire (Annex 3.2), read aloud to participants. It included ten questions, gathering information about: sex, age, main language, level of finished studies, degree of sight loss (question only for AD users), the age at which the visual impairment started (question only for AD users), technology usage, VR usage, and assistive technologies usage when accessing online content (questions only for AD users).

1.4.1.2. Participants

A total of six audio describers and five persons with sight loss (two partially-sighted and three blind) took part in both focus groups. Following best practice on qualitative research design (Barbour 2008; Krueger 1998; Bryman 2008), a reduced number of participants was recruited to allow for effective discussion. Participants were recruited through academic contacts and via social media (Twitter, Facebook). Details regarding participants can be found in Chapter 2 and 3.

1.4.1.3. Procedure

Both focus groups followed the same structure, which can be summarized as follows:

1. Participants are welcomed.

2. Participants are presented with the information sheet and informed consent form.
3. Demographic questionnaire is administered.
4. Participants take part in a group discussion, followed by the approval by all participants.

The group discussions were structured, as they consisted of two tasks. The first task was addressed to professional describers taking part in the study. In the second task, AD users were asked for their feedback on the technology of spatial sound. For a detailed description of the procedure followed for each study, see Chapter 2 and 3.

1.4.2. Usability study

This section describes the methodological aspects of the usability study conducted with professional describers. What follows is a succinct account of the measurement tool, the participants, the stimuli and the test development followed during the first round of this study.

1.4.2.1. Measurement tool

The study followed a mixed-method research strategy, which combined quantitative and qualitative data collection. Data was gathered by means of two questionnaires: demographic questionnaire and post-questionnaire, both administered in an online form. Demographic questionnaire gathered the following information about the participants: sex, age, main language, current job, previous experience with audio describing 360º videos, experience of working in the field of AD, language used when audio describing, level of qualifications, previous training in AD, usage of electronic devices, frequency of watching VR content and interest in VR content.

The post-questionnaire was organized in the following way:

1. System Usability Scale (SUS) questionnaire (Sauro and Lewis 2016).
2. Preference questions (eight open questions).
3. Additional users' comments.

The questionnaires were prepared in an online form, in English, which allowed for gathering feedback from audio describers from various countries. Versions of the demographic questionnaire and post-questionnaire are found in Annexes 4.1 and 4.2 respectively.

1.4.2.2. Participants

Twenty four professional audio describers took part in the first iteration of the usability test. Participants were recruited through academic contacts. The call for participating in the study was posted on social media (Facebook, Twitter) and sent via several mailing lists. Volunteers willing to take part in the test had to fulfil two inclusion requirements: they had to be professional audio describers and, secondly, they had to be able to complete the required tasks and answer the questionnaires in English. All the participants were presented with information sheet before the study (Annex 2.1) and gave their consent before taking part (Annex 2.2). Details regarding participants from this iteration are presented in Chapter 4.

1.4.2.3. Stimuli

The video used in the first iteration of the study was fictional content, which presented a linear story. The reasons behind choosing this video as a stimulus are two-fold. Firstly, it provided enough spaces between the dialogues to insert audio descriptions without strict time constraints. Secondly, the presented story developed at different places within the storyworld, which allowed for testing a function of the online tool related to placing the spatial sound within the storyworld. More detailed information regarding the video can be found in Chapter 4.

1.4.2.4. Procedure

The procedure for this study included the following steps:

1. Presentation of the information sheet and informed consent form.
2. Administration of the demographic questionnaire.
3. Testing procedure:

a) Presentation of the user guide.
b) Presentation of instructions for participants.
c) Performing several tasks in the editor.

4. Administration of the post-questionnaire.

All questionnaires were sent to participants via e-mail upon their agreement to participate in the study. The details about the procedure can be found in Chapter 4.

1.4.3. Second iteration of the usability study

It should be noted that a second iteration of this study took place, but it is not reported in the articles included in this thesis. However, due to its relevance, it has been decided to briefly summarise the main methodological aspects here and discuss the results in the conclusions in relation to the first iteration. A thorough description of the participant profiles and results is included as Annex 4.3.

1.4.3.1. Measurement tool

The same questionnaires were administered as in the first iteration of the study (see section 1.4.2.1). The only changes were made in relation to the demographic questionnaire because they allowed for a more precise data analysis. One question, added at the beginning of the questionnaire asked about the participation in the previous iteration of the test. Another two questions were added, which asked participants about their previous experience with recording their descriptions and the usage of software. The post-questionnaire, consisting of usability and preference questions was the same as in the first iteration of the study.

1.4.3.2. Participants

Similarly to the first iteration, twenty four professional audio describers took part in the second iteration of usability study. Seven participants from the previous iteration took part in this test. Participants were recruited through academic contacts, mailing lists and social media (Facebook, Twitter). Volunteers had to fulfil the same requirements as in the first iteration of this study. Although they were the same, the sample was not homogenous in the two tests. All participants were presented with an information sheet before the study

(Annex 2.1) and gave their consent before taking part (Annex 2.2). Details regarding participants are presented in Annex 4.3.

1.4.3.3. Stimuli

The same video materials were used in both iterations of the study. The only difference was that a different excerpt from the same video was used because seven participants had taken part in the previous iteration. Including the same excerpt could therefore have impacted the results of the study. The excerpt chosen for this iteration was shorter than the excerpt in the previous iteration because some participants raised concerns about the total duration of the study. Table 2 provides a comparison of the excerpts used in both tests:

	First iteration	**Second iteration**
Length	1 minute 10 seconds	50 seconds
Timecodes	00:00 – 01:10	03:30 – 04:20

Table 2. Excerpts of the video clip used in the usability studies

1.4.3.4. Procedure

The same procedure was followed as in the first iteration of the study (see section 1.4.2.4). However, as the tools were improved based on the feedback from the first iteration, participants were presented with improved instructions on how to use the tool. For a detailed description of the procedure, see Annex 4.3.

1.4.4. Reception studies

The studies conducted with the AD users followed a two-level structure, comprising the pilot test and the main test. As the methodology for both tests is detailed in the Chapter 5, this section will describe only the key methodological elements of both studies, including

the measurement tools, the participants, the materials, and test development. Both studies conducted with AD users followed a mixed-method approach. Quantitative and qualitative analysis allowed for an in-depth analysis of the matters under investigation.

1.4.4.1. Measurement tool

The research tool used for measuring user experience is the Igroup Presence Questionnaire (IPQ), one of five canonical questionnaires used to measure presence (Rosakranse and Oh 2014). IPQ has fourteen questions and measures four facets of media experience: general presence, spatial presence, involvement and perceived realism. Spatial presence refers to the sense of being there in the virtual environment; involvement refers to the attention given to the virtual environment, and experienced realism refers to the reality judgment of the virtual environment (for details see Chapter 5).

The IPQ was chosen as a measurement of user experience, as it has been validated in different forms of virtual environments, (Regenbrecht and Schubert 2002; Brown et al. 2003; Krijn et al. 2004; Hartanto et al. 2014; Kinateder et al. 2015), including a head-mounted display (HMD), and it is recommended as a measure of presence because of its high reliability (Schwind et al. 2019).

The demographic questionnaire included questions that asked participants about their: sex, age, first language, education, degree of sight loss, age at which the visual impairment started, usage of technologies, frequency of watching VR content, the reasons behind not using this technology, level of interest in immersive content, types of content watched on TV or online, daily usage of AD and the usage of assistive technologies. A copy of the demographic questionnaire in two language versions, English and Catalan, constitutes Annex 5.1 to this thesis.

To measure users' presence and gather preferences, two questionnaires were used: a presence questionnaire and preference questionnaire. A detailed description of both questionnaires is contained in the corresponding article in Chapter 5. A copy of the IPQ questionnaire used in two studies can be found in Annex 5.2. The preference questionnaire used in the pilot study is presented in Annex 5.3. and the preference questionnaire used in the main study is described in Annex 5.4.

1.4.4.2. Participants

The only inclusion requirement to be met was that participants should be visually-impaired adults speaking Catalan and Spanish, but no other limitation was established. A total of thirty six participants took part in two experimental studies conducted with AD users (fourteen blind, twenty two partially-sighted). They were recruited through the Catalan Association for the Blind and Visually-impaired (B1B2B3) and through personal contacts. All participants were presented with an information sheet before the study (Annex 2.1) and gave their consent before taking part (Annex 2.2). Details regarding the participants are presented in Chapter 5.

1.4.4.3. Stimuli

What follows is a brief presentation of the stimuli used in the pilot study (section 1.4.4.4) and the main study (section 1.4.4.5). More details can be found in Chapter 5.

The common motivation behind selecting the clips for both studies was that they had to present independent stories. As such, participants could understand them when played in a random order. The clips had to be comparable and the density of dialogues could not be too high in order to provide audio description in pauses between them.

The research material for both studies consisted of three initial clips from *Holy Land* series produced by Jaunt Ryot[5]. In this travel documentary, viewers are transported to various cultural sites in Israel, guided by a main narrator. The videos were stand-alone narrative pieces, comparable in length. Table 3 shows the durations of all episodes:

Episode 1	Episode 2	Episode 3
5:05	3:56	4:02

Table 3. Length of the clips used for both studies with AD users

[5] I would like to thank Ryot and Jaunt VR for letting the ImAc project use *Holy Land* content.

1.4.4.4. Stimuli used in the pilot study

The aim of this study was to test different presentation modes of spatial sound. Therefore, actions had to develop at different places within the storyworld in the clips used for this study. For each episode, three AD presentation modes were created. The variable that was modified for this test was sound.

Classic presentation mode is a standard sound option in which AD was heard from above participant's head. Audio description delivery in this presentation mode can be assimilated with standard AD for 2D audiovisual content.

Static presentation mode in an innovative sound option in which AD is delivered from participant's left or right side, as if a friend is sitting close and telling the story.

In the Dynamic presentation mode, the sound of AD is placed in the location of the main action or other important visual elements being described.

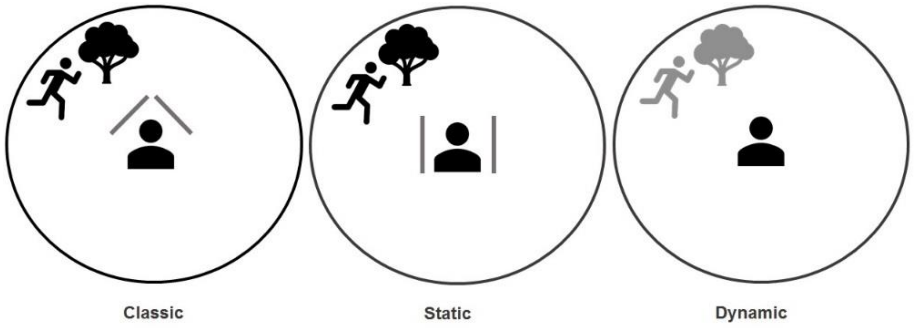


Figure 1. Presentation modes used in the pilot test

1.4.4.5. Stimuli used in the main study

The main study used the same clips as the pilot study. Only the presentation modes differed in this study from the previous test, as the experimental methodology was revised. Based on the results of the pilot study, and taking into account the interest of users in new scripting approaches, three presentation modes were created for this study.

The Classic mode was the same that was used in the pilot study. It features a standard, neutral narrator, and it can be assimilated with standard audio description used for standard 2D content.

The Radio mode follows a non-standard scripting style, in which the narrator adapts the voicing of AD to the given scene. For instance, if the scene takes place inside a church, the narrator whispers. If the scene takes place at a busy market, AD is delivered in a more lively way.

The Extended presentation mode contained extended descriptions that could be activated at the user's will. The participants were asked to play all the extended descriptions, which were indicated by a short bell sound. Upon playing them, the main video would pause until the end of the additional track. The duration of extended descriptions varied between 20–45 seconds. While the main AD description track gave details that were crucial to understand the plot, extended descriptions contained information that could allow users to create a better mental picture of the characters and places presented. Such detailed descriptions would not be otherwise possible, due to time constraints.

1.4.4.6. Procedure

The procedure adopted for both studies conducted with AD users was identical. It included the following steps:

1. Presentation of information sheet and informed consent.
2. Administration of the demographic questionnaire.
3. Testing procedure:
    a) Screening: clip 1.
    b) Measures taken: questionnaire, part 2.
    c) Screening: clip 2.
    d) Measures taken: questionnaire, part 2.
    e) Screening: clip 3.
    f) Measures taken: questionnaire, part 2.
4. Administration of the questionnaire, part 3.

Before presenting participants with the presentation modes, audio introductions were played in both studies. The order of the videos was randomized across participants. For a detailed description of the procedure followed for each study, see Chapter 5.

Chapter 2. Article 1

Audio description in 360° videos:
Results from focus groups in Barcelona and Spain

2. Article 1

Abstract

This article[6] discusses how audio description could be integrated into 360° videos by reporting the results from two focus groups conducted in the initial stages of the ImAc project. To involve participants in the research process, the project adopted a user-centered methodology, and a series of focus groups was conducted with professional audio describers and end users to gather feedback about their needs and expectations regarding the implementation of audio description and, secondarily, audio subtitling. Results indicate that content selection in this medium raises concerns for audio describers, and needs to be further researched. The results obtained from the end users not only highlight the need to audio describe the main action, but also their interest in having different parts of the visual scene audio described. Results also indicate that auditory cues would allow end users to orient themselves in the scene, and feel more immersed in the content presented.

Keywords: audiovisual translation, media accessibility, user-centered methodology, audio description, immersive media, 360° videos, virtual reality

2.1. Introduction

Rapid developments in the field of virtual environments can be seen around the world (Manjoo 2014). Although the medium is new, and its possibilities still need to be defined, immersive technologies are being applied in different industries. They are already used in video conferencing, language learning, e-commerce, architecture, the medical field, filmmaking and video games (Gleb, n.d.; EC 2017). It is very likely that immersive technologies will permeate other areas of the technological landscape in the near future. Although in many cases 360° content, which is defined in this article as a type of Virtual Reality (VR), is still being produced for tests and experimental purposes (EBU 2017, 8),

[6] This is a post-peer review version of the article: Fidyka, Anita, and Anna Matamala. 2018. "Audio description in 360º videos: Results from focus groups in Barcelona and Kraków." Translation Spaces 7(2): 285–303. https://doi.org/10.1075/ts.18018.fid.

and most existing 360° videos released are "supplementary content for on-air programmes" (EBU 2017, 9), more than half of European broadcasters have begun to offer 360º content, or will offer it in the near future (EBU 2017, 8). This content will be used to entertain, inform and educate audiences, and will need to be made accessible to all of them.

Article 27 (1) of the Universal Declaration of Human Rights states that: "everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits." Moreover, Article 9 (g) of the UN Convention on the Rights of Persons with Disabilities states that appropriate measures should be taken by member parties to "promote access for persons with disabilities to new information and communications technologies and systems, including the Internet." Taking into account this framework, all types of new technology, including 360º videos, should provide accessible content and platforms, thus catering for the needs of all members of society. Access services, such as audio description, audio subtitling, subtitles for the deaf and hard of hearing, and sign language interpreting, studied within the field of Audiovisual Translation (AVT) and Media Accessibility (MA), can be seen as instruments to ensure human rights, benefitting not only persons with disabilities, but also other groups, such as the elderly, migrants, foreign language speakers and language minorities (Greco 2016). This paper will focus on audio description (AD), and how it could be integrated in immersive environments, both from the perspective of the end consumer of access services, and the professional who creates them. Since audio subtitling (AST) often coexists with AD, some comments will also be made in relation to the former.

Audio description is an audiovisual transfer mode that represents visual images in words, and thus makes them accessible for those who cannot access the visuals. As defined by Snyder (2008, 192), AD "provides a verbal version of the visual." This access service allows visually impaired persons to access audiovisual material and cultural property autonomously. As far as audiovisual content is concerned, AD is defined as "an additional narration that fits in between dialogues to describe action, body language, facial expressions, scenery, and costumes – anything that will help a person with a visual impairment to follow the plot of the story" (Whitehead 2005). Closely related to AD is AST (Braun and Orero 2010), which provides an "aurally rendered and recorded version of subtitles" (Reviers and Remael 2015, 52). Audio subtitles can be offered as an independent

access service, but are often integrated with AD when text is present on screen, especially in the form of subtitles (Matamala 2014).

In 2D audiovisual products such as films, AD is delivered between the dialogues, and it is expected to not interfere with music and other important sound effects (Jankowska 2015). Empirical studies on this type of audiovisual translation conducted to date (Perego 2016) have included eye-tracking studies (Mazur and Chmiel 2016, Szarkowska et al. 2013), and reception studies aimed at determining users' comprehension (Cabeza-Cáceres 2011), preferences (Chmiel and Mazur 2016), emotions (Ramos Caro and Rojo López 2014), and presence (Wilken and Kruger 2016). However, empirical studies carried out to date in relation to the subject of AD in more immersive environments are almost non-existent.

Given this context, the aim of our research was to gather user feedback, through a series of focus groups, on how AD (and secondarily AST) could be integrated in immersive content, both from the perspective of producers and consumers. Although it could be argued that immersive technologies are still at a very early stage, we believe that it is the right moment to approach users, and ask them how they think accessibility should be taken into account. In short, we believe that a user-centered design methodology should be adopted when developing new technologies, and AVT and MA scholars should make a key contribution in defining user needs. As has often been advocated in AVT and MA studies, especially in papers related to accessible filmmaking (Romero-Fresco 2013; Udo and Fels 2010a, 2010b), accessibility needs to be considered as part of the production process, and not as an afterthought.

This research is framed within the Immersive Accessibility project, ImAc, a 30-month initiative funded by the European Commission within the H2020 framework. The aim of ImAc is to research how access services can be integrated in 360º technology following a user-centered design methodology in which user input is sought at every stage in the process, and accordingly influences the next. One of the first actions in the project was to gather user feedback on various access services in different countries through a series of focus groups. More specifically, three focus groups on AD were conducted: in Great Britain, Spain and Poland.

This article will report on the focus groups devoted to AD carried out in Spain (Barcelona) and Poland (Kraków). Section 2.2 gives an overview of virtual environments. Section 2.3

reviews previous work in the field of AD in high-immersive environments. Section 2.4 outlines the methodological aspects of the focus groups, and section 2.4.3 offers a discussion of the results. Finally, conclusions and future research possibilities are presented in section 2.5.

2.2. Virtual environments: defining 360º videos

This section aims to define 360º videos within a taxonomy of virtual environments. As defined by Slater and Usoh (1993, 221), a virtual environment is "an environment created by an interaction of a human participant with a world displayed by the computer." The authors also propose the term "immersive virtual environments" for those in which "sensory input to the user from the external world is, ideally, wholly provided by the computer generated displays" (Slater and Usoh 1993, 221). Presence is a concept used by some authors to describe and measure users' sense of immersion in audiovisual content, and is central to experiencing virtual reality. This multi-construct concept encompasses different definitions, and has been so far defined by scholars as a "perceptual illusion of non-mediation" (Lombard and Ditton 1997, 9), a "psychological sense of immersion in any mediated environment" (Fryer and Freeman 2012), and an "experiential quality metric employed to evaluate broadcast and virtual environment media systems" (Lessiter et al. 2001, 282). As stated by Slater and Usoh (1993), both external and subjective factors may contribute to users' sense of presence in immersive virtual environments.

Under the umbrella term of virtual environments, it is possible to differentiate content belonging to the following environments: VR, augmented reality (AR) and mixed reality (MR). Taking into consideration the definition of immersive virtual environments proposed by Slater and Usoh (1993), VR environments in which sensory input is wholly computer-generated can be characterized by a higher degree of immersive capacity than environments in which computer-generated input is mixed with the images of the real world, such as in AR or MR. In general, VR, as defined by Sherman and Craig (2003, 13), is "a medium composed of interactive computer simulations that sense the participants' position and actions and replace or augment the feedback to one or more senses […]." In other words, VR is a medium through which we can experience a computer-generated reality that simulates realistic experiences. However, as stated by Sherman and Craig (2003, 6), the definition of VR is still in flux, as it is a new medium.

Currently, users can access VR content by means of two types of head-mounted displays: one providing 6 degrees of freedom, the other providing 3 degrees of freedom (EBU 2017, 14). The difference between the two lies in the user's movement options: in the case of 6 degrees of freedom, users are able to move their bodies in the visual scene, while in the case of 3 degrees of freedom headsets, users are limited to one bodily position and discover the surrounding visual scene by head movements.

Current VR systems use motion sensors for head tracking, hand tracking or body position tracking (Sherman and Craig, 77), and screens for stereoscopic displays (Sherman and Craig 2003, 132). In the past, stereoscopic (3D) images, which create a 3D illusion by using a pair of 2D images, were used commercially in 3D movies, by means of polarized glasses (e.g. IMAX 3D).

As enumerated by Sherman and Craig (2003, 6), "the key elements in experiencing virtual reality – or any reality for that matter – are a virtual world, immersion, sensory feedback (responding to user input) and interactivity." A virtual world, the first key element, is defined by the authors as "the content of a given medium" (Sherman and Craig, 6). The authors specify that when it is VR, "it brings those objects and interactions in a physically immersive, interactive presentation" (Sherman and Craig, 7). The second key element of VR listed by the authors is immersion, which authors use in two ways, differentiating between physical (sensory) and mental immersion (Sherman and Craig, 9). The third element listed by the authors as essential to VR is sensory feedback, provided to users according to their physical position. Finally, the fourth key element is interactivity, as VR should respond to users' actions to seem authentic. Within the term "interactivity," the authors differentiate the ability to affect a computer-based world as well as change one's viewpoint within a world (Sherman and Craig, 10).

360° videos, referred to by some as spherical, omnidirectional, or surround videos (Bleumers et al. 2013, 800), are considered by many to be a form of VR. As classified by the European Broadcasting Union (EBU 2017), within VR, it is possible to distinguish: (1) computer-generated VR when "the content is primarily rendered from a 3D model in real time and on the user's device" (EBU, 6); (2) 360° videos when the content is primarily video-based; (3) combinations of the above, which can be placed between 360° videos and computer-generated input, when an immersive experience is created by using both content

types; and (4) panoramic 2D (monoscopic) or stereoscopic images viewed on head-mounted displays.

360° videos provide a vision which unfolds 360° horizontally and 180° vertically relative to the observer's physical location. While viewing with them, users stand in one physical location, and trigger the content by head movements (3 degrees of freedom). As stated by Bleumers et al. (2013, 800), "people can freely choose the viewing angle while [the omnidirectional video] plays, as if they are turning and controlling the camera. As such, [omnidirectional video] provides viewers with a new form of interactivity." Bleumers et al. (2013) also note that omnidirectional videos need to be distinguished from multi-angle videos. With the former, users can choose only their "viewing direction from a given viewpoint," while with the latter, users are presented with "the opportunity to choose between alternate video streams, often showing a single event […] from different viewpoints […]." Video content can be viewed on the flat screen of a personal computer; in such cases surrounding actions and scenes can be discovered by the viewer by clicking on an arrow cursor, or by means of head-mounted displays.

Other types of virtual environments are AR and MR, as indicated above. In contrast to VR, AR does not immerse users fully in computer-generated content, but overlays it on real-world images, and these two types of content cannot interact with each other. According to Sherman and Craig (2003, 18), "they give the user additional information about the physical world not perceived by unaided human senses," therefore the amount of information available to users is increased compared to their usual sensory perception. Usually, it is the visual sense that is augmented: "augmented reality [is] a type of virtual reality in which synthetic stimuli are registered with and superimposed on real-world objects; often used to make information otherwise imperceptible to human senses perceptible" (Sherman and Craig, 18). Head-mounted displays or mobile devices are used to access VR, while in AR portable devices, such as smartphones or tablets, special glasses or headsets are also used (Gleb, n.d).

Similarly to AR, the real world is also enhanced with digital objects in MR. The difference is that computer-generated content is combined with real-world content, while being anchored to it, and thus it interacting with it (EBU 2017). As virtual content is anchored to the real world, a headset needs to track it, and adjust virtual content accordingly. Holographic devices or head-mounted displays similar to VR headsets are required to

experience MR. Such devices can be translucent glasses that allow real surroundings to be seen, and in which virtual experiences are created with holograms. Alternatively, they can feature non-translucent displays that completely block the real world (Gleb, n.d).

According to the Slater and Usoh (1993), virtual environment displays can provide information in visual, auditory, and kinaesthetic modalities. As direct sensory feedback is an essential ingredient of virtual environments, it is provided to users experiencing these environments based on their physical location, and mostly through the visual sense. There are, however, environments that provide touch experiences, i.e., haptic technologies, currently under study. To give an example, haptic technologies are being developed by the Walt Disney Company that could allow persons with sight loss to interact with flat surfaces of digital media, giving them an impression that such surfaces are three-dimensional. As suggested by Booton (2013), thanks to such touch screens – that is "screens that not only look but actually feel 3D," persons with sight loss are able to feel the textures and edges of flat digital objects. The emergence of haptic technology will open up new possibilities for more engaging forms of AD, as described in the section 3 of this article.

To sum up this section, one could conclude that 360º videos are one type of VR alongside other forms such as computer-generated VR, and they are defined by Bleumers et al. (2013, 800) as "a form of video that has been captured so that […]: the viewer can look around in a 360°, camera-registered, moving image."

2.3. AD in immersive contents

Research on AD in 360° videos is, to the best of our knowledge, non-existent. This may be due to the fact that 360° video is relatively new compared to other media. It could also be due to the fact that it is a largely visually-driven medium, as in most cases it is the visual sense that receives sensory feedback (Sherman and Craig 2003, 10). Most research on AD is still very much related to television and cinema (Chmiel and Mazur 2014, 43; Ofcom 2000), and existing standards and guidelines on AD (ACB 2003; Ofcom 2000; Remael et al. 2015) generally focus on AD in low-immersive 2D audiovisual products.

One could argue, as Fryer and Freeman (2012, 15) do, that more interactive forms of AD exist in contexts other than immersive media. First of all, descriptive audio guides at museums and galleries include audio instructions, which provide orientation information,

or guide “a blind person’s fingers around a raised, tactile image” (Fryer and Freeman, 15). Secondly, in live AD contexts, such as theatres, museums and art galleries, AD users are invited to visit the stage, or to touch costumes, settings and props during ‘touch tours’, while these objects are being described. Additionally, some of the ADs invite persons with sight loss to interact with presented objects; for example, to operate a bell in the theatre context (Fryer and Freeman, 15). Research on AD in this type of event, and on audience participation in live performances (Di Giovanni 2018b), may be a good source of information when designing strategies to audio describe immersive content.

The question of AD in relation to immersive media, although not explored extensively in AD practice, has been addressed in some experimental studies to date. The emergence of 3D cinema in the last decade resulted in the need to address the question of the AD of 3D effects. To that end, a small-scale study in the form of focus groups was organized in the UK (Greening 2011). The results obtained suggest, however, that there would be little user interest as far as explicit descriptions of such effects are concerned; participants unanimously considered that there is no need for 3D effects to be described (Greening, 3). The study consisted of two focus groups, and involved 10 persons with sight loss who regularly watched audio described programmes on TV and DVDs. The procedure included explaining to the participants the history and technique of 3D as well as performing a task consisting in viewing 3D video clips. The reason behind this is that when 3D effects are audio described, less time is left for AD of other significant visual elements. The participants in the study gave more importance to issues such as facial expressions, location of characters, other actions taking place on the screen, age of characters present in the scene, costumes and physical appearance of characters.

Another approach was taken in studies researching the possibility of incorporating haptics with AD in order to convey the sense of touch, as in certain VR systems, auditory and kinaesthetic senses receive most of the feedback (Sherman and Craig 2003, 14). Viswanathan et al. (2010) focus on how haptic descriptions can add relevant information, and facilitate comprehension of video materials. The authors suggest that the lack of time between dialogues leaves certain information undescribed, and this could be overcome by the incorporation of haptics with AD. In that study, participants were shown several audio described scenes from various movies, and experienced tactile cues for the relative position of two actors in a scene through a vibrotactile belt, and facial expressions of the actors

through a vibrotactile glove. The vibrotactile cues, experienced by participants as vibrations around their waists, corresponded to location and distance information of two actors conversing in a movie scene. The vibrotactile glove, however, allowed participants to experience vibrations on the back of their hand. These touch cues corresponded to the facial expressions of actors on the screen. In each scene, multiple actors were present, and their facial expressions were preceded by locating the actor through the belt. The glove provided six basic human emotions – happiness, sadness, surprise, anger, fear, and disgust – in addition to neutral expressions. The authors of the study claimed that it maintained the suspense of the movie by not interpreting the expressions, but by mapping them to its nine vibration motors.

Other research has focused on how VR can be made accessible to users with sight loss, especially in gaming situations, but the focus has been on technical aspects beyond audio description and will not be discussed in this paper (Colwell et al. 1998; Ghali et al. 2012; Picinali et al. 2011).

2.4. Methodology

Bearing in mind that the question of implementing AD in 360° videos has not been researched yet, the aim of conducting a series of focus groups was to provide the basis for the development of AD in 360° videos by analysing the needs of potential end users.

The rationale for choosing this type of qualitative research was manifold: (1) focus groups enable participants to become familiar with the immersive technology, and to explain its different facets and possibilities; (2) focus groups allow participants to share different points of view, negotiate senses, revise their opinions and reach common conclusions (Barbour 2007); and (3) they allow researchers to ask additional questions to clarify confusing aspects.

Both focus groups described here were conducted with a limited number of participants, as recommended in the literature on qualitative research design (Krueger 1998; Barbour 2007; Bryman 2008). We decided to involve both professional users (service producers such as audio describers or technical experts) and home users (service consumers) with some technological expertise, hence called ‘advanced’ home users here.

As 360° videos offer a 360-degree field of view horizontally and a new way of interaction, in a sense that users can choose which contents to trigger by head movements, our assumption was that content selection when audio describing in this medium would be deemed more problematic by professional audio describers than in 2D media. Regarding home users, our assumption was that they would be interested in triggering audio descriptions of different parts of the visual scene by head movements, but would also find it equally important to be offered the audio description of the main action to allow them to follow the plot. Another assumption regarding content consumption was that users with sight loss would need to be guided in the visual scene, and immersive sound could prove particularly useful in this regard.

The focus group in Barcelona was led by a facilitator and drew on the services of two note takers; the first note taker followed the discussion among participants and took note of their responses, while the second structured the notes in the form of conclusions. The focus group in Kraków was moderated by a facilitator, who also took structured notes from the discussion in the form of conclusions. Both studies were carried out in accordance with ethical guidelines and approval was obtained from the Ethics Committee of the Universitat Autònoma de Barcelona. The studies were anonymous and privacy was ensured.

2.4.1. Participants

Data in Barcelona were obtained from 6 participants (aged between 25 and 51), made up of: 2 advanced end users (partially sighted), 3 audio describers and 1 technical expert. None of the professional audio describers suffered from sight or hearing loss. All participants declared themselves to be frequent users of the Internet and technological devices. A laptop was the most frequently used technology by the participants on a daily basis (5), followed by TV (4) and mobile phone (4), tablet (3) and PC (2). None of the participants possessed a device to access 360° content. All participants were familiar with AD and AST.

Data in Kraków were obtained from 6 participants (3 end users and 3 audio describers). They were 2 males and 4 females, with ages ranging 25–46. The end users were blind participants: with vision impairment from birth (2) and between 5–12 (1). All participants had university educations. One of them reported having a device to access VR content. Mobile phone was the most used technology by the participants on a daily basis (6),

followed by laptop (5), PC (3), tablet (2) and TV (1). Similar to participants in Barcelona, all participants in the focus group in Kraków were familiar with AD and AST.

2.4.2. Procedure

The focus group in Barcelona was conducted on 24 November 2017, and the focus group in Kraków was conducted on 28 December 2017. The study in Barcelona lasted approximately 90 minutes, and consisted of a number of consecutive steps. First, participants were welcomed and the ImAc project was introduced. Participants were familiarized with the aim of the study, 360° technology, and the glasses to access it. Second, all participants signed informed consent forms, and filled in pre-questionnaires with demographic data before the discussion commenced. Alternative oral consent forms were read aloud to participants with visual impairments. The study conducted in Kraków used the same methodology, and lasted approximately 120 minutes.

The pre-questionnaire contained 11 open and closed questions, organized in the following blocks (1) socio-demographic profile (age, sex, educational level); (2) useable vision and age at which visual impairment began; (3) use of mobile and web technologies; (4) questions related to the use of screen readers (e.g. JAWS, VoiceOver, TalkBack), magnifiers (e.g. Zoomtext) and voice commands; and (5) a question on immersive media exposure. The questionnaires were coded and, together with informed consent forms, they will be securely stored for three years after the completion of the project.

To trigger the discussion, the facilitator had prepared two tasks and a list of guiding questions. In Task 1, which lasted approximately 15 minutes in Barcelona and approximately 30 minutes in Kraków (together with the leading questions asked by the facilitator), participants watched a short 360º video. The input chosen for the focus group in Barcelona was an episode of *Polònia*, a TV comedy show broadcast by the Catalan public broadcaster TV3. The reason for choosing this input was that, as the story develops, new characters appear in different parts of the visual scene. Users can thus follow the main plot or move their heads to different parts of the 360º scene. The input chosen for Task 1 in Kraków was a 5-minute 360º video: an interview with a Polish ski jumper on the premises of a ski jump. The rationale behind selecting this input was that the interviewer

and interviewee change locations in the course of the clip. This allows users to follow the main plot, or to ignore it and choose to watch surrounding landscape.

In Task 1, one of the professional audio describers was asked to produce live audio description addressed to the end users present in the room. The aim of this activity was for audio describers to indicate the main challenges they faced, and how these could be addressed in terms of production, and for end users to indicate the main challenges in terms of consumption, and so that both groups could suggest how 360º technology could be rendered accessible.

In Task 2, which lasted approximately 15 minutes in Barcelona and 20 minutes in Kraków, participants were asked to listen carefully to an audio input. The audio input presented the technology of object-based audio (IRT Lab, n.d.) using an orchestra as an example in both focus groups. This sound technology allows users to hear where the sound comes from, and changes according to a user's head movements. In the case of audio input in this task, this sound technology rendered particular instruments more or less audible depending on the user's head position. After the listening activity, participants were asked to discuss whether this type of audio could be used, and how, in providing audio description for 360º video content.

Following the two activities described above, there was a discussion based on a list of guiding questions related to the provision of access services for 360º video content, and finally conclusions were agreed. At the end, the researchers answered additional questions asked by participants.

2.4.3. Discussion of results

The remainder of this article discusses the results of the focus groups in Barcelona and Kraków. However, given the scope of this article, the emphasis is placed on analysing the needs of both the professional audio describers and home users of AD and AST in 360° media, and the results concerning the technical aspects of AD production are mentioned only briefly in the next subsection.

2.4.3.1. Results regarding the production of audio description

Results regarding the production of AD can be grouped into two main categories: the amount of visual information that needs to be described, and the specificities of the software for producing AD, which will not be discussed in detail here, as already indicated. As a general remark, professional audio describers in the focus group in Barcelona pointed out that it is challenging to describe the visual scene, as, in 360° media, there is much more visual information to convey than in a standard AD. A visual metaphor used by one describer is that one should describe the scene "as if you were inside a sphere." The difficulty regarding content selection was confirmed in the focus group in Kraków, as professional describers considered that, in this medium, sighted users can choose which parts of the visual scene to consume, and users with sight loss should also be given this possibility.

Audio describers in Barcelona suggested that to allow the user to look around and discover the visual scene, there should be an option to pause the video. Then, different AD tracks related to different sections of the visual scene would be triggered by head movements. This option creates the possibility of watching the content several times, and listening to different AD tracks of the visual scene each time. This possibility would, however, mostly concern the content consumed at home, and not that presented in public venues such as museums as more time would be needed to watch content in this way. This approach was also raised in the focus group in Kraków, where the participants compared it to 'choose-your-own-adventure' books. Participants in both focus groups indicated that this approach would increase the number of AD units, and consequently the workload. Also, according to participants in both groups, this approach would impact on the cost of producing AD. Professionals in Kraków stated that the question of how to remunerate audio describers in this medium is crucial.

As far as technology is concerned, professionals in Barcelona suggested that 360° video content should be divided into different sections on screen, and ADs could be provided for each section. They expressed the need for a general view of a visual scene in a flat view as well as the possibility to open particular sections of the visual scene in new windows for producing AD. They even considered that a minimum of 4 sections, ideally 6, would be needed. Audio describers in Poland confirmed these results by stating that a general view

of the visual scene would be needed, with an option to write the AD in windows linked to different sections of the visual scene.

Finally, professionals in Barcelona suggested that it would be helpful to check the final version of the AD using immersive glasses, but they would prefer to produce the AD in a content manager displayed on a flat screen. The Polish participants also prefer to work on a flat screen, and only to check the final result with immersive glasses because then it would be difficult to write on the keyboard and mark time codes. To this end, they consider that they would need a text-to-speech module that would allow them to proofread the final version of the ADs with glasses.

2.4.3.2. Results regarding the consumption of audio description

The results obtained on the consumption of AD, both in Barcelona and in Kraków, can be divided into two major categories: comments concerning how to access the services, and comments on the specific features of the AD, and to a lesser extent AST.

As far as the question of accessing the services is concerned, the participants expressed their views in terms of (1) activation and deactivation of audio description; (2) screen magnifiers and screen readers; (3) audio subtitling personalization; (4) identification of user preferences and parameters; and (5) viewing immersive content with or without glasses.

Conclusions reached during the discussion both in Barcelona and in Kraków suggested that end users prefer to open personalization options using voice commands, hence this is an important feature to be added when developing any interface. As suggested by one participant, voice commands would be needed to give instructions such as ‘play’, ‘stop’, ‘pause’, ‘forward’, ‘rewind’, and ‘switch AD/AST on and off’. Users also requested that the immersive player integrate screen readers as well as screen magnifiers to enable the enlargement or zoom of menus, another feature that proves useful to many users with sight loss.

Concerning AST, users in the focus group in Barcelona expressed the view that it is a service that should be activated or deactivated by means of voice commands. They also indicated that they prefer to be offered AST rather than an option to enlarge the text of the subtitles. Participants also indicated that they would like user preferences and parameters to be automatically remembered and transferred between different devices, which was

confirmed by the responses provided by the participants in the focus group in Poland, who added that they should be able to mark their preferences in check boxes. When asked about accessing immersive content in head-mounted displays, or by means of a smartphone with a sensor tracking the user's head movements and headphones, participants in both focus groups indicated that there is no one-size-fits-all solution, and that using one option or the other depends very much on each end user's specific needs. Additionally, end users in Poland were in favour of accessing immersive content by means of a smartphone with a sensor tracking the user's head movements and headphones.

Regarding the results obtained from end users on the actual access services under analysis, they can be grouped into the following categories: (1) describing main action and secondary scenes; (2) returning to the main action; (3) using immersive sound; and (4) prioritizing information according to the volume.

The conclusions reached both in Barcelona and in Kraków show that users need AD linked to the main action, as it allows them to follow the plot. Users also suggested that, beside the main action, they should be able to discover different parts of the visual scene by turning their heads. They would like to activate additional ADs describing secondary actions or surroundings by head movements, which confirms our assumptions. As proposed by the participants, a film could be stopped at any time to listen to secondary AD units. It could even be watched several times, and each time a different viewing path could be chosen. One user suggested, "there should always be a predominant AD and other secondary ADs available, so that each user can have their own experience."

Regarding immersive sound, users in Barcelona stated that it could be helpful to position oneself in the visual scene, and identify the place where the main action is happening. The responses also indicate that immersive sound helps persons with sight loss feel more immersed in the content presented. For example, one user noted that: "immersive sound can help you position yourself (who is where in the scene, to your right/to your left). If you feel involved, you can feel more present in the scene and guide yourself through the scenario much better." This is confirmed by the responses provided by the participants in the focus group in Poland, who considered that object-based audio deepens the sensation of being in the centre of the action.

As far as the question of returning from secondary scenes to the main action is concerned, end users in Barcelona indicated that a specific sound effect could instruct them that they are looking at the main action, and not in a different direction. Although this challenge was also indicated in Poland by both end users and professional audio describers, no specific solution was proposed in this regard.

As to the prioritizing of information, in the focus group in Barcelona it was suggested that the volume of AD could help users differentiate the main action from surrounding actions and scenes. As one participant commented, "depending on what you are looking at, the volume of the sound could be modified. If you are looking at a scene, the volume should be higher to indicate that this is the main action you are perceiving."

Moreover, an innovative idea was put forward during the focus group in Barcelona: information could be delivered in the form of 'headlines' or 'highlights', which could encourage users to turn their heads to the area from where the sound comes. In other words, a suggestion made by the users was that if users were interested in that so-called headline, they could turn their heads towards that action. Then, the volume would automatically increase. It remains to be seen, though, how this could actually be put into practice.

In contrast to the results obtained in Barcelona, in which end users suggested that they could be guided inside the visual scene by means of short 'headlines', end users in Poland stated that they would like to be guided in the visual space by means of immersive sound. In other words, although the option of 'headlines' was suggested by professionals in Poland, end users seemed more attracted to the possibilities of immersive sound, as they considered it would allow persons with sight loss to know where to turn their heads to receive the AD.

Finally, participants in both groups also indicated the possibility of AD being voiced by a female and male voice: one being applied for the main action, and the second one for additional AD units, as a means to differentiate between the overlapping visual input that necessarily is present in a virtual environment.

Additionally, as voice over is often used in Poland, participants in Kraków were asked which transfer mode would be preferable when consuming 360º foreign-language content: dubbing or voice over. The responses provided show that users are strongly in favour of dubbing instead of voice over when consuming 360º foreign-language products.

2.5. Conclusions

This article has presented the results of focus groups conducted in Barcelona and Kraków in the initial stages of the ongoing European project ImAc. In order to contextualize the focus groups, the state of the art of the limited research in the field of AD in virtual environments was outlined.

Our initial assumption regarding content selection was confirmed by professional describers; it was deemed more problematic than in 2D content. Also, spontaneous responses provided by both professional describers and advanced home users, and agreed in the form of conclusions in both focus groups, suggest that while watching 360º content users should have the opportunity to follow the main action by listening to corresponding AD, and should also be able to consume additional AD tracks, which confirms our assumptions. With that aim in mind, it was suggested in both focus groups that the video material should be paused to enable users to discover different parts of the visual scene. The resulting challenge consists in increasing the number of AD units available, but which will not always be activated.

In terms of the application of immersive sound, participants expressed their interest in its implementation in AD in 360º media, as it may serve to make them feel more immersed in the world presented, to know where to turn to receive AD and to help them to orient themselves in the visual scene. This also confirmed our initial assumptions.

All things considered, end users voiced their interest in 360º technology, and consuming AD in 360º content in the future. Also, professional audio describers and end users in both focus groups stressed the need to implement access services now that the technology is being developed. It is in this context that participant-oriented methodologies such as focus groups are a necessary first step. Although focus groups rely on a small number of participants, and results cannot be generalized to a wider population, they nonetheless constitute a qualitative research method that has “few rivals in terms of method” (Saldanha and O’Brien 2013, 170) when it comes to finding out about people’s conscious thoughts about a certain topic in the field of translation and Media Accessibility.

All the knowledge gained during the project is contributing to building a critical mass, which is much needed in this new field of AD in VR. Although the direction of the development of 360º technology and contents will determine the possible lines of research

– and possible ways of implementing access services – the recommendations provided by advanced home users and professional audio describers in Barcelona and in Kraków, along with the results from a focus group replicated in the UK, provide a solid basis for the development of access services in 360º media in the next stages of ImAc, and for future extensive experimental testing with wider population samples. The question of AD in 360º still needs, however, to be researched thoroughly, especially concerning storytelling – so as to provide some first insights regarding content selection – and concerning the application of immersive sound. All this will allow guidelines to be drawn up for audio describers of 360º media to ensure the quality of AD in such environments, and, in turn, the quality of the user experience.

Funding

ImAc has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement 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 Anita Fidyka'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

ACB (American Council of the Blind). 2003. *ADI AD Guidelines.* Accessed July 26, 2018. http://www.acb.org/adp/guidelines.html

Barbour, Rosaline. 2007. *Qualitative Research Kit: Doing Focus Groups.* London: SAGE Publications. https://doi.org/10.4135/9781849208956

Bleumers, Lizzy, Wendy Van den Broeck, Bram Lievens, and Jo Pierson. 2013. "Extending the Field of View: A Human-Centered Design Perspective on 360° TV." *Behaviour & Information Technology* 33 (8): 800–814. https://doi.org/10.1080/0144929X.2013.810780

Booton, Jennifer. 2013. “By Tricking the Brain, Disney’s Bringing Digital Sight to the Blind.” Accessed July 27, 2018. https://www.foxbusiness.com/features/by-tricking-the-brain-disneys-bringing-digital-sight-to-the-blind

Braun, Sabine, and Pilar Orero. 2010. “Audio Description with Audio Subtitling: An Emergent Modality of Audiovisual Localization.” *Perspectives: Studies in Translatology* 18 (3): 173–188. https://doi.org/10.1080/0907676x.2010.485687

Bryman, Alan. 2008. *Social Research Methods*. Oxford, New York: Oxford University Press.

Caro-Cáceres, C. 2011. “Intonation in AD: Does it Affect Users’ Comprehension?” *Paper presented at 4th Media For All*, London, June 28–July 1.

Chmiel, Agnieszka, and Iwona Mazur. 2016. “Researching Preferences of Audio Description Users. Limitations and Solutions.” *Across Languages and Cultures* 17 (2): 271–288. https://doi.org/10.1556/084.2016.17.2.7

Colwell, Chetz, Helen Petrie, Diana Kornbrot, Andrew Hardwick, and Stephen Furner. 1998. “Haptic Virtual Reality for Blind Computer Users.” In *Proceedings of the Third International ACM Conference on Assistive Technologies – Assets ’98, Marina del Rey, April 15–17,* 92–99. New York: ACM Press. https://doi.org/10.1145/274497.274515

Di Giovanni, Elena. 2018. “Audio description for live performances and audience participation.” *JosTrans: The Journal of Specialised Translation* 29: 189–211. Accessed July 26, 2018. https://www.jostrans.org/issue29/art_digiovanni.pdf.

EBU (European Broadcasting Union). 2017. *Virtual Reality. How are Public Broadcasters Using it?* Accessed July 26, 2018. https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it

EC (European Commission). 2015. “Europe 2020 Strategy Policy.” *European Commission Website*. Accessed July 26, 2018, https://ec.europa.eu/digital-single-market/en/europe-2020-strategy

EC (European Commission). 2017. “Virtual Reality at the Service of Healthcare.” *Executive Agency for Small and Medium-sized Enterprises (EASME) Website*. Accessed July 26, 2018. https://ec.europa.eu/easme/en/news/virtual-reality-service-healthcare

Fryer, Louise, and Jonathan Freeman. 2012. “Presence in Those with and without Sight: Audio Description and its Potential for Virtual Reality Applications.” *Journal of CyberTherapy & Rehabilitation* 5 (1): 15–23. Accessed July 26, 2018. https://www.acb.org/adp/docs/cybertherapy%20article%20final%20draft%20March%201.pdf

Ghali, Neveen I., Omar Soluiman, Nashwa El-Bendary, Tamer M. Nassef, Sara A. Ahmed, Yomma M. Elbarawy, and Aboul Ella Hassanien. 2012. “Virtual Reality Technology for Blind and Visual Impaired People: Reviews and Recent Advances.” In *Advances in Robotics and Virtual Reality*, edited by Tauseef Gulrez, and Aboul Ella Hassanien, Intelligent Systems Reference Library Series 26: 363–385. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-23363-0_15

Gleb, B. n.d. “VR vs AR vs MR: Differences and Real-Life Applications.” *Ruby Garage Blog*. Accessed July 26, 2018. https://rubygarage.org/blog/difference-between-ar-vr-mr

Greco, Gian Maria. 2016. “On Accessibility as a Human Right, with an Application to Media Accessibility.” In *Researching Audio Description: New Approaches*, edited by Anna Matamala, and Pilar Orero, 11–33. London: Palgrave Macmillan.

Greening, Joan. 2011. “Do We Need 3D Audio Description Guidelines: Recommendations from Focus Group Study Report.” Accessed July 26, 2018. http://audiodescription.co.uk/uploads/general/3D_film_and_tv_report_2.pdf

IRT Lab. n.d. “Interactive Content.” *Interactive Lab Website*. Accessed July 28, 2018. https://lab.irt.de/demos/object-based-audio/interactive/

Jankowska, Anna. 2015. *Translating Audio Description Scripts: Translation as a New Strategy of Creating Audio Description*. Frankfurt am Main: Peter Lang. https://doi.org/10.3726/978-3-653-04534-5

Krueger, Richard A. 1998. *Focus Group Kit: Moderating Focus Groups*. Thousand Oaks: SAGE Publications. https://doi.org/10.4135/9781483328133

Lessiter, Jane, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. 2001. “A Cross-media Presence Questionnaire: The ITC-Sense of Presence Inventory.” *Presence: Teleoperators, and Virtual Environments* 10 (3): 282–297. https://doi.org/10.1162/105474601300343612

Lombard, Matthew, and Theresa Ditton. 1997. “At the Heart of it All: The Concept of Presence.” *Journal of Computer-Mediated Communication* 3 (2). https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Manjoo, Farhad. 2014. “If You Like Immersion, You’ll Love This Reality.” *The New York Times* April 2, 2014. https://mobile.nytimes.com/2014/04/03/technology/personaltech/virtual-reality-perfect-for-an-immersive-society.html

Matamala, Anna. 2014. “Chapter 6. Audio Describing Text on Screen.” In *Audio Description. New Perspectives Illustrated*, edited by Anna Maszerowska, Anna Matamala, and Pilar Orero, 103–120. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.112.07mat

Mazur, Iwona, and Agnieszka Chmiel. 2016. “Should Audio Description Reflect the Way Sighted Viewers Look at Films? Combining Eye-Tracking and Reception Study Data.” In *Researching Audio Description. New Approaches*, edited by Anna Matamala, and Pilar Orero, 97–121. London: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-56917-2_6

Ofcom. (Office of Communications). 2000. *ITC Guidance on Standards for Audio Description*. Accessed July 26, 2018. http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf

Perego, Elisa. 2016. “History, Developments, Challenges and Opportunities of Empirical Research in Audiovisual Translation.” *Across Languages and Cultures* 17 (2): 155–162. https://doi.org/10.1556/084.2016.17.2.1

Picinali, Lorenzo, Christopher Feakes, Andrew Etherington, and Timothy Lloyd. 2011. “VR Interactive Environments for the Blind: Preliminary Comparative Studies.” In *Joint Virtual Reality Conference of euroVR and EGVE*, 113–115.

Ramos Caro, Marina, and Ana María Rojo López. 2014. “Feeling Audio Description: Exploring the Impact of AD on Emotional Response.” *Translation Spaces* 3 (1): 133–150. https://doi.org/10.1075/ts.3.06ram

Remael, Aline, Nina Reviers, and Gert Vercauteren (eds). 2015. *Pictures Painted in Words. ADLAB Audio Description Guidelines*. Trieste: Edizioni Università di Trieste. Accessed July 26, 2018. http://www.adlabproject.eu/Docs/adlab%20book/index.html

Reviers, Nina, and Aline Remael. 2015. “Recreating Multimodal Cohesion in Audio Description: A Case Study of Audio Subtitling in Dutch Multilingual Films.” *New Voices in Translation Studies* 13 (1): 50–78. https://www.iatis.org/images/stories/publications/new-voices/Issue13-2015/Articles/Reviers_New_voices_PUBL.pdf

Romero-Fresco, Pablo. 2013. “Accessible Filmmaking: Joining the Dots Between Audiovisual Translation, Accessibility and Filmmaking.” *JosTrans. The Journal of Specialised Translation* 20: 201–223. Accessed July 26, 2018. http://www.jostrans.org/issue20/art_romero.pdf

Saldanha, Gabriela, and Sharon O’Brien. 2013. *Research Methodologies in Translation Studies*. Manchester: St Jerome Publishing.

Sherman, William R., and Alan B. Craig. 2003. *Understanding Virtual Reality: Interface, Application, and Design*. The Morgan Kaufmann Series in Computer Graphics. Amsterdam: Morgan Kaufmann Publishers.

Slater, Mel, and Martin Usoh. 1993. “Representations systems, perceptual position, and presence in immersive virtual environments.” *Presence. Teleoperators and Virtual Environments* 2 (3): 221–233. Cambridge: Massachusetts Institute of Technology Press. https://doi.org/10.1162/pres.1993.2.3.221

Snyder, Joel. 2008. “Audio Description: The Visual Made Verbal.” In *The Didactics of Audiovisual Translation*, edited by Jorge Díaz-Cintas, 191–198. Amsterdam/Philadelphia: John Benjamins Publishing. https://doi.org/10.1075/btl.77.18sny

Szarkowska, Agnieszka, Izabela Krejtz, Krzysztof Krejtz, and Andrew Duchowski. 2013. "Harnessing the Potential of Eye-Tracking for Media Accessibility." In *Translation Studies and Eye-Tracking Analysis*, edited by Sambor Grucza, Monika Pluzycka, and Justyna Alnajjar. Warschauer Studien zur Germanistik und zur Angewandten Linguistik Series, 153–183. Frankfurt am Main: Peter Lang. https://doi.org/10.3726/978-3-653-02932-1

Udo, John-Patrick, and Deborah I. Fels. 2010a. "The Rogue Poster-Children of Universal Design: Closed Captioning and Audio Description." *Journal of Engineering Design* 21 (2–3): 207–221. https://doi.org/10.1080/09544820903310691

Udo, John-Patrick, and Deborah I. Fels. 2010b. "Universal Design on Stage: Live Audio Description for Theatrical Performances." *Perspectives: Studies in Translatology* 18 (3): 189–203. https://doi.org/10.1080/0907676X.2010.485683

United Nations General Assembly, Resolution 217A, Universal Declaration of Human Rights (UDHR). Accessed July 26, 2018: http://www.un.org/en/universal-declaration-human-rights

United Nations Department of Economic and Social Affairs. Division For Inclusive Social Development, Convention A/RES/61/106 2006, Convention on the Rights of Persons with Disabilities (CRDP). Accessed July 26, 2018. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html

Viswanathan, Lakshmie Narayan, Troy McDaniel, Sreekar Krishna, and Sethuraman Panchanathan. 2010. "Haptics in Audio Described Movies." In *IEEE International Symposium on Haptic Audio Visual Environments and Games, Phoenix*. 1–2. https://doi.org/10.1109/have.2010.5623958

Whitehead, Jill. 2005. "What is Audio Description?" In *Vision 2005: Proceedings of the International Congress held between 4 and 7 April 2005 in London, UK.* International Congress Series 1282: 960–963. https://doi.org/10.1016/j.ics.2005.05.194

Wilken, Nicole, and Jan-Louis Kruger. 2016. "Putting the Audience in the Picture. *Mise-en-Shot* and Psychological Immersion in Audio Described Film." *Across Languages and Cultures* 17 (2): 251–270. https://doi.org/10.1556/084.2016.17.2.6

Chapter 3. Article 2

Retelling narrative in 360° videos: Implications for audio description

3. Article 2

Abstract

The aim of this article[7] is to question whether the approach to producing audio description in 2D films needs to be revisited for 360° narrative videos, a new media format characterized by its immersive capacity. A two-step research methodology was designed. First, an extensive literature review was performed. The data obtained during the first step was then used to design and conduct focus groups. The first part of the article discusses the findings from the literature review, comparing standard narratives with 360° narrative videos. It draws a number of conclusions for audio describers in relation to content selection, a key task in the translation of visuals into words. In the second part, data obtained from the focus groups held with describers and users is presented. The results suggest possible approaches to audio description for 360° content, such as the use of spatial sound and elements of interaction.

Keywords: audiovisual translation, media accessibility, immersive environments, Cinematic Virtual Reality, user-centered methodology, audio description

3.1. Introduction

Virtual environments (Slater and Usoh 1993) are increasingly present in our society to inform, educate, or entertain audiences. Under this umbrella term, three main types of content can be found: Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR) (Fidyka and Matamala 2018). Within the field of VR, one can find 360º videos, which contain content based primarily on videos as opposed to computer-generated images (EBU 2017). 360° videos offer a 360-degree horizontal field of view and 180-degree vertical field of view and are becoming a new storytelling medium (Damiani and Southard 2017) for their interactive potential. Storytelling techniques in 360° videos are currently being

[7] Acknowledgement: The Version of Record of this manuscript has been published and is available in Translation Studies: 10 March 2021
https://www.tandfonline.com/doi/abs/10.1080/14781700.2021.1888783

researched by different scholars and filmmakers who are experimenting with so-called Cinematic Virtual Reality (Gödde et al. 2018). 360° videos can have one linear story or they can propose navigational alternatives (Reyes 2017; Lee 2017). In such videos, the viewer decides which viewing path to take; for instance, what decision the protagonist has to make. It is, however, the first group, i.e. 360º videos with one linear story, that will be the focus of this article.

To date, most of the 360º content are non-fiction, location-based programmes, such as the news and current affairs (Jones 2017, 171), whereas fictional content – such as animations, short films or episodic series (Dooley 2017, 164) – is being developed to a lesser extent. Thanks to this media format, audiences can follow important social events from their location (EBU 2017, 9), or interact with these locations in a way that would not be possible in real life. For example, a journey to the centre or orbit of our planet can be made. This is why the concept of presence, defined as "psychological sense of being there" (Biocca 1997, 18) is central to experiencing this media format. In other words, users interacting with 360º videos are expected to feel immersed in the content presented.

Taking into account the human right to access culture and information, Virtual Reality content, including 360º videos, should be rendered accessible. This access can be provided by a variety of different services which can be viewed as an instrument to ensure that human rights are being upheld (Greco 2016).

360º videos are multimodal (Braun 2008, 3), which means that they create meaning from the combination of and interaction between different semiotic channels: images, dialogues, sounds and on-screen text. As the visual channel is involved, they are not fully accessible for persons with sight loss. Audio description (AD) and additional access services such as audio subtitles (AST) could then be used to render this type of audiovisual content accessible to the visually impaired target audience. In general terms, AD is a type of intersemiotic translation that transfers the visual channel to a verbal mode (Maszerowska, Matamala and Orero 2014; Fryer 2016). In audiovisual products, such as films, theatre, opera, sports or music events, a description of the visual components is provided along with certain sound components that would be difficult to recognize without their verbal description (Remael 2012). It is primarily created for audiences with vision loss as it enables them to create a mental representation of the visual layer, but it can also be

beneficial for other groups, e.g. foreign language learners. In terms of narrative productions, it allows its users to recreate the original story.

As far as the production of AD in films is concerned, and similarly for any type of translation process, ADLAB guidelines (Remael, Reviers and Vercauteren 2014) state that audio describers should conduct a source text analysis before drafting a script. Such analysis is a two-step process. First, the describer focuses on the story the filmmaker intends to convey, trying to determine what specific narrative elements are used. Then decisions are made on which elements to include in the description to allow the audience to recreate the story. This second step, content selection, is identified as one of the main issues in AD (Vercauteren 2016). Current guidelines offer help on how to select content for standard films, but this issue has not yet been addressed in relation to 360º narrative video content.

In order to assess whether the current practices for producing AD in 2D films have to be revisited for 360º narrative videos, a two-step research methodology was chosen. As an initial research tool, a literature review was carried out to show how 360º narrative videos change the experience of traditional narratives, and to offer initial insights into content selection in the new media format. Based on the data obtained during the literature review, focus groups were designed and held. The data obtained through these two methodological tools will help to define how AD can be integrated into this media format and help to expand knowledge on this type of intersemiotic translation, eluding traditional distinctions of genres and content types.

After a brief discussion of the principles of story-creation and story-recreation, we offer an overview of existing studies on storytelling in 360° narrative videos. We then discuss the current state of the art on narratology in AD and draw some conclusions on content selection for audio describers. In the last two sections, we describe the methodology of the focus groups and discuss the results obtained in relation to storytelling in 360º narrative videos.

3.2. From traditional film narrative to narrative in 360° videos

Narratology is a discipline that studies all types of narrative, including verbal and audiovisual texts and helps one to “understand, analyse and evaluate narratives” (Bal 2009,

3). Mieke Bal differentiates between three layers in a narrative text: the fabula, the story and the text (ibid., 5). She defines the fabula as "a series of logically and chronologically related events caused or experienced by actors" (ibid.). This layer consists of the following elements: events, actors, time and locations. They are later ordered into a story, which is defined as a fabula "presented in a certain manner" (ibid.). In other words, when creating a filmic story, filmmakers first decide what actions the characters will perform or undergo and later, how the story will be told. At this second stage they can decide, for instance, the order of the actions (chronological or non-chronological), their frequency (the actions can be shown more than once from the perspective of different characters or they can be omitted), or their speed (actions can be shown at their normal speed, in fast-forward, or in slow-motion). At this stage, the filmmakers also decide the specific traits that the characters will have and the details of the presented places. Moreover, they decide what amount of time will be given to the various elements of the fabula (ibid., 8). The third layer distinguished by Bal is "a narrative text" (ibid., 9). It is a story "told" (ibid.) in a medium, i.e. converted into signs. In other words, in the last stage of story-creation, the filmmakers make a concrete decision about how the story will be presented, i.e. which film techniques are going to be used.

Another process of which audio describers should be aware of is audience story-reconstruction. In general terms, the audience is presented with a concrete narrative text and must recreate the abstract construct used by the filmmaker to create the story. When watching a film, the audience members process and interpret information by combining actions with other information and establishing temporal relations between them. Important to the story reconstruction are mental frames (Herman 2002) created and updated by the audience. They initially contain basic information about the characters and settings in which the actions are to take place (e.g. "2019"). Later, as the audience members constantly update their mental model of the story, more details are added (e.g. "a rainy day in October 2019"). Only when they are able to create a new frame for a given event or update an already existing frame are they able to follow the story.

Knowledge about these two processes is valuable when conducting the source-text analysis. While the first process allows describers to establish the most relevant narrative elements in the content, the latter allows them to know what cues are necessary to facilitate the reconstruction of the filmic story by blind and partially-sighted audience members

(Vercauteren 2012; Remael, Reviers and Vercauteren 2014). In the remaining part of this section, we will look at existing studies on storytelling in 360° videos in an attempt to compare traditional films with 360° narrative content.

Studies on storytelling in traditional 2D films are numerous and the approaches are diverse. However, research on narratives in 360º videos conducted to date is scarce as they only recently have received scholarly attention (Gödde et al. 2018). As the full spectrum of possible storytelling techniques in this media format are still being defined, it is not possible to delineate them.

The first relevant aspect is the role of the audience and their level of control. Dolan and Parets (2016) enumerate four narrative forms based on user experience within the virtual world: a user can be an observer or participant and be simultaneously active or passive. Similarly with film narratives, 360° productions fall into the passive observer category (Jones 2017), as the audience members are placed at the centre of the events, with the option to explore the virtual world without influencing the development of the narrative. In other words, narratives in 360° media formats are pre-constructed by a film director or content creator. A second aspect worth mentioning is how the content is watched. Dooley (2017) notes the main difference is the rectangular screen on which traditional audiovisual media material is typically viewed, whilst in 360° content, the storyworld surrounds the viewer. In this regard, the process of viewing 360° videos is similar to watching content in a planetarium, the difference being that members of the audience are not seated inside a dome but wear head-mounted displays and discover the storyworld through head movements. As a result, the attention of viewers towards narratologically-relevant elements cannot be directed in the same way as can be in films (Dooley 2017; Gödde et al. 2018). A study by Rothe, Hußmann and Allary (2017) demonstrated that it is especially difficult to guide the user's attention at the beginning of a new scene. As viewers can explore the entire visual space, they can omit significant parts of the main narration, looking instead at other parts of the storyworld. This can result in a lack of understanding of the causality of plot (Jones 2017; Syrett, Calvi and van Gisbergen 2017).

To guide users, dispersing the main action across several parts of the storyworld is recommended (Jerald 2015), along with the use of auditory and visual cues. Rothe, Hußmann and Allary (2017) tested several traditional methods of guiding the viewers'

attention (salient objects, sounds, lighted objects and movement) in order to see whether they could be transferrable to Cinematic Virtual Reality. Although the authors note that further investigations are necessary to find the most appropriate guiding methods, the results of their study show that moving objects or lights can guide the viewing direction within the storyworld even without any sounds. Similarly, Nielsen et al. (2016) compared diegetic cues (guiding viewers' attention implicitly by encouraging them to follow a firefly with their gaze) with non-diegetic cues (controlling their body orientation) and no guidance. The results indicate that diegetic cues can guide viewers effectively within the storyworld, whilst assuming control of the viewers' action may decrease their perceived presence in the storyworld. These two preliminary studies show that diegetic cues related to movement may be used often in 360° narrative videos to indicate where the audience should look.

As far as film language is concerned, techniques such as crossfades and fade-outs may be used more frequently in this media format than in standard films; indeed, some consider them to be a better method for changing the scene than cuts (Gödde et al. 2018), as changing location rapidly can be disorienting. Recently, however, filmmakers have begun to use cuts more often, as viewers become more accustomed to them in VR (ibid.).

As cuts are being used less frequently than in standard films, the space in 360° content undergoes a gradual transformation around the viewer, by different elements being added or removed (Damiani and Southard 2017). Additionally, the space can contain elements of the internal lives of the characters such as emotions, dreams or thoughts (ibid.) by means of text, visuals or colours, which appear in the storyworld. For this reason, 360° content is thought to be an appropriate media format to convey subjective, impressionist first-person narratives (ibid.).

All in all, storytelling in VR is still in the process of being defined. This preliminary analysis shows that although the audience in 360° content follows a linear story, pre-constructed by the filmmaker in a similar way to standard films, the differences in narrative construction still exist, such as lack of framing and a different use of guiding cues or scene transition techniques.

3.3. Narratology in AD research: from standard films to 360º content

Based on narratological knowledge, models for content selection in AD were developed by Jan-Louis Kruger (2010) and Gert Vercauteren (2016). Kruger (2010, 233) differentiated two approaches for rendering visual information accessible for audiences with sight loss: traditional AD and audio narration (AN), located at two ends of a continuum. While AD reflects the visual channel objectively, subjective audio narration reflects the narratological sense of the production. AD with elements of narration and subjectivity is placed in the middle of the spectrum.

Vercauteren (2016) adopted a narratology, both structuralist and cognitive, to develop a systematic method for audio describing filmic stories, based on their most relevant narrative constituents. His research presented a "determine-decide" approach, considering what narrative elements could be included in AD and how to prioritize these narrative elements based on audience needs in order to interpret and recreate stories.

Other research conducted in the field of AD has also adopted a narratological approach. In a study for the Pear Tree Project, Kruger (2012) analysed eye-tracking and written data to discern how to select AD content that would allow the audience to reconstruct the original story in a narratologically more meaningful way. The results of his study show that elements that occupy a lot of space on the screen do not necessarily have more relevance compared to elements that are visually insignificant. Similarly, by applying a cognitive approach, Jeroen Vandaele (2012) focused on how to select the most relevant AD elements. The results of his research suggest that aside from the information that allows AD users to follow the plot, the AD script should contain information about the events that can or could have happened, to allow AD users to experience suspense, curiosity and surprise. Anna Matamala and Aline Remael (2015) analysed the narrative of films driven by visual effects to determine whether audio descriptions written for such films require a different approach. Thanks to the analysis of "the cinema of spectacle", the authors conclude that because such narratives aim to dazzle their audience with effects, the emphasis in the description should be placed on the prosody of the AD script, the use of metaphors and interaction with sound. In other words, how to describe should be given importance in this type of filmic production.

Anna Maszerowska (2014) studied the visual codes of film language such as light and contrast and how they should be rendered in AD. Her doctoral thesis provides a preliminary compilation of the existing solutions for rendering lighting in AD. It concludes that more consistency is needed in the rendering of light in AD script and that knowledge about the different functions of light can prove helpful when drafting AD scripts. Taking a cognitive approach, Nazaret Fresno (2014) examined the reception of characters in fiction films. The results of her investigation show that user recall is facilitated by limiting the information in the descriptions and dividing it into short units for delivery at different stages.

The studies related to standard content are multiple, but it remains to be seen how the current approach to providing AD could be transferred into new 360º formats. In the following sections we will focus on the main constituents of film narrative: events, spatiotemporal settings, characters and film language, and we will discuss the strategies used to describe them in 2D content. This will be done with a view to assessing what strategies could be used in 360º narrative videos to allow the audience to reconstruct their stories.

3.3.1. Describing events

As previously mentioned, audio describers should pay attention to order, duration and frequency when describing action (Remael, Reviers and Vercauteren 2014). For example, they are advised to signal any instances of events that are out of chronological order, to prioritize information or describe the visuals in advance in case of events being shown on the screen simultaneously by means of a split screen or deep space composition. Also, in case an event is shown more than once at different points during a film, describers should determine whether it is repeated unmodified or with a different status and, in turn, decide whether to include the repeated event in the script (ibid.; Vercauteren 2016).

In 360° content, events are rarely shown out of chronological order, meaning that the action can be described in the AD without any temporal signalling (e.g. “a year earlier”). Actions can also happen simultaneously, not by means of a split screen, but by locating them in different parts of the storyworld. In this regard, as is the case with standard films, describers must decide during the text analysis which events guarantee the coherence of the story and which complement the main narrative. The possible challenges regarding the description of events may be related to the signalling via which part of the storyworld they are

presented. This information must be provided to allow persons with sight loss to have a more immersive viewing experience.

3.3.2. Describing spatiotemporal settings

Bal's theory of narrative space (2009) defines settings as an evolving constituent of narrative. According to Bal, when a story returns to settings that were previously introduced, they can be repeated in an unchanged way (repetition), new elements can be added (accumulation) or they can undergo a transformation. In addition to providing a background, settings can have a symbolic function.

When analysing a source text, audio describers must determine whether settings are new or known, whether they are only a background or if they have a symbolic function. If settings are already known, describers have to determine whether they are the same as before or whether they have changed. In circumstances where the settings serve only as a background, a description with basic spatiotemporal information is sufficient. Regarding settings with a symbolic function, a more detailed description should be provided. If the settings return to a place that is already known, describers have to determine whether something has changed (Vercauteren and Remael 2014).

When describing settings in 360° videos, describers will have to follow the same rules as they would when describing standard content, i.e. determine where the action takes place, in which period and the function of any given setting. The changes in settings that can occur around the viewer, i.e. addition or disappearance of visuals within the storyworld were defined by Bal as "accumulation" (2009, 139) and should be similar for those in standard films. Such elements should be signalled in the script. In a case where a given element reappears, describers must determine whether it is unmodified or whether its status has changed.

The challenge will probably relate to the prioritization of visual information in the storyworld. As users have a 360º view, the storyworld includes more elements than standard 2D content. AD should, however, fit in between the dialogues, which may not give describers enough time to describe all relevant elements.

3.3.3. Describing characters

The issue with describing characters has been studied by several authors (e.g. Fresno 2012; Orero 2011; Vercauteren and Orero 2013). This topic has also been tackled in the AD-Verba and ADLAB projects. ADLAB guidelines (Remael, Reviers and Vercauteren 2014) distinguish between three categories of characters: focal, secondary and background ones. Each requires a different approach in AD. Focal characters are the most important to the narrative, secondary characters usually have a supporting function, and background characters will be less detailed in the script. When writing an AD script, more attention is normally given to dynamic characters, i.e. those that develop remarkably throughout the narrative, as opposed to those that develop only marginally (Vercauteren 2016, 227). Audio describers also have to decide whether the physical or mental dimension of a character is more important when describing them (ibid.). Additionally, describers need to determine whether the character being presented is new or already known (Remael, Reviers and Vercauteren 2014; Vercauteren 2016, 233), as new characters must be introduced with information about their age, appearance and characteristic traits. For already known characters, describers should determine whether they have changed. Another relevant aspect is whether characters are authentic or fictional, real or unrealistic and whether they serve a symbolic function, i.e. if they represent a certain group of people, social class, profession, or stereotype.

It may be assumed that the challenges faced with regard to describing characters in 360° videos will be very similar to those in standard films. However, in 360° videos two or more characters can appear in different parts of the storyworld. Without adequate guidance, persons with sight loss may therefore find it more challenging to know where they are located. Moreover, as this media format seems better suited to convey the internal lives of characters and subjective narratives, a need may arise for description of emotions to be further explored, as done by Braun (2007), Mazur (2014), and in AD guidelines (e.g. American Council of the Blind 2009, 6). Additionally, as close-ups are not possible in this media format, identification of characters' emotions may prove more challenging.

3.3.4. Describing film language

The term *film language* refers to what the authors of the ADLAB guidelines (Remael, Reviers and Vercauteren 2014) define as "the accepted systems, methods or conventions

through which a film story comes to the audience." Under this umbrella term, four broad categories can be found: mise-en-scène, cinematography, editing and sound design. The rendering of film language in audio description has been studied by various authors (e.g. Perego 2014; Fryer and Freeman 2013; Vercauteren 2016). This topic has also been studied in the ADLAB project, based on the example of the film *Inglourious Basterds (2009).* When writing AD for films, describers should determine the narrative function of the techniques used by the filmmakers (Remael, Reviers and Vercauteren 2014), and may want to adopt another strategy, such as including the actual filmic techniques in the description, as done for art-house films (Szarkowska and Wasylczyk 2014; Walczak and Fryer 2017).

Although 360° videos are still in their infancy, certain aspects differ from those in standard films and may impact on descriptions. As far as cinematography is concerned, the main challenge for describers could be the lack of framing, which results in a larger amount of visual detail, more than typically found in standard films. Similarly, as different types of shots such as close-ups are not possible, prioritizing the information presented in the storyworld may prove challenging. As regards editing, crossfades and fade-outs – which tend to appear more often in 360° videos than standard films – should be signalled in the AD script (e.g. "the screen fades to black"). Other aspects such as mise-en-scène comprise "all the elements placed in front of the camera to be photographed: the settings and props, lighting, costumes and make-up, and figure behavior" (Bordwell and Thompson 1990, 410) and will be used in a similar way to standard films but the amount of visual detail could be challenging for describers.

It can be concluded from this section that content selection in 360° videos will bear many similarities to content selection in standard films. It can also be asserted that the possible challenges when describing such content may be related to a) guiding the attention of AD users towards the relevant narrative elements, and b) prioritizing information in the storyworld. In the following sections, the methodology applied in the focus groups and the participant feedback on storytelling are discussed and serve as a necessary first step in a potential reshaping of AD in new media.

3.4. Focus group methodology

The previous sections have discussed how storytelling is changing in 360º videos and how this can impact on audio descriptions for this new media format. However, when designing

new solutions it is advisable to include the actual users in the making process in order to bridge the maker-user gap (Greco 2018). Based on the data collected during the literature review, focus groups were designed to involve end users. This user-centric approach has been applied at the core of the ImAc project, in which the implementation of access services in 360º videos is being researched alongside the development of the technology. This approach is similar to the view put forward in the field of accessible filmmaking (Udo and Fels 2010a, 2010b; Romero-Fresco 2013), stating that access services should already be considered in the content production process and not as an afterthought.

The focus groups proved of value in the early stages of the project as they enabled participants to become familiarized with 360º technology. This qualitative research method allowed participants to ask additional questions about the possibilities of this technology as well as to confront their views on the implementation of access services with other participants.

The first focus group took place in Barcelona, Catalonia (Fidyka and Matamala 2018), and the second in Kraków, Poland. Both studies had the same structure and consisted of two tasks followed by a discussion and agreed conclusions. The studies were conducted in accordance with ethical procedures approved by the Ethics Committee of Universitat Autònoma de Barcelona. Participants were informed about the aim and context of the study and were asked to sign consent forms. All data was confidential and the identity of the participants was made anonymous.

The focus group in Barcelona (24 November 2017) was conducted with two end users, three audio describers and one technical expert. The focus group in Kraków (28 December 2017) was conducted with three audio describers and three AD users. As the literature on qualitative research design recommends a limited number of participants for conducting focus groups (Barbour 2008, 60), the inclusion of six participants proved an optimal number for the discussion. Aside from participants, a facilitator moderated each focus group.

3.4.1. Profile of participants

Before the discussion, participants filled in a pre-questionnaire aiming at determining their demographic profile. It contained eleven questions regarding age, sex, educational

background, visual impairment, use of technologies, exposure to immersive media and access to online content. Participants were aged between 25–51 years old. Three end users defined themselves as blind. Of these three, two defined themselves as blind from birth and one, as becoming blind between 5–12 years old. The remaining two end users defined themselves as partially-sighted. Almost all (11) participants had university education and one participant reported having "further education". The responses indicated that all participants were frequent technology users. Most of the participants used mobile phones (10) and laptops (10) on a daily basis, followed by PCs (5), tablets (5) and TVs (5). All participants were familiar with AD. Only one participant owned a device to access VR content.

3.4.2. Procedure

The actual discussion in each of the focus groups held in Barcelona and Kraków was preceded by two tasks. Task 1 began with showing participants a five-minute 360° video and asking one of the professional audio describers to provide a live AD. The input in the focus group in Barcelona was an episode of "Polònia", a comedy show broadcast on public Catalan television. The input used in Kraków was an interview with a ski jumper. The reason behind choosing these materials was that both comprised actions happening all around the storyworld which allowed the users to follow the main action or to ignore it by looking at their surroundings. In Task 2, participants watched the clips with spatial sound (IRT Lab n.d.). Asking the end users about this technology was deemed important, as it enables placing AD in different locations within the storyworld. Thanks to said sound technology in AD, persons with sight loss will be able to localize actions and other audio elements being described.

3.4.3. Discussion of results

In the remainder of this article, the results of the focus groups are discussed, with special emphasis on the aspects related to storytelling. Other relevant aspects, such as more general comments relating to the production of AD in this media format, integration of audio subtitles or interaction by voice are discussed in Fidyka and Matamala (2018).

First of all, professional audio describers taking part in the study in Barcelona suggested that the 360º storyworld be much larger than the standard display for 2D films and that special attention be given to setting. To this end, it was suggested that the description of the main action should be provided as a priority and that AD for the surroundings would also be needed. Therefore, a new approach for producing AD was suggested: the storyworld could be divided into four to six sections, with an AD of the surroundings provided in every section. It was later added that the main narration should be paused to listen to additional recordings.

Audio describers who took part in the study in Kraków, however, deemed this format interactive, which means that the user can choose which parts of the storyworld to see. Therefore, they considered that persons with sight loss should also be able to watch 360° content in an interactive way. They expressed an interest in the solution proposed in the focus group in Barcelona (pausing the main narration and listening to additional AD tracks), as this creates the possibility of interaction.

All in all, the end users present in both focus groups agreed that an AD needs to be provided for the main action. The results from the focus group in Barcelona suggest that secondary AD tracks triggered by head movements when wearing a head-mounted display could also be implemented. This approach is interesting as it could provide a solution to the prioritization of content discussed earlier in this article. With this solution, audio describers could include more information that could later be activated at the will of the user.

Interestingly, although not strictly related to the topic of storytelling, in both focus groups, blind participants asked about the possibility of not wearing a head-mounted display. For example, two blind people in the focus group in Kraków discussed the possibility of wearing headphones with a sensor that would track their head movements instead of a head-mounted display.

Regarding content selection, audio describers in both focus groups considered it challenging. Audio describers in the focus group in Kraków agreed that deciding which elements of the storyworld should be described would depend on the content. In some situations, priority should be given to conveying the intention of the director or content creator, while in others, e.g. travel simulations or city tours, describing the surroundings should be given importance. Audio describers in the focus group in Kraków agreed that

taking into account the amount of possible work, the materials for AD should be carefully selected. One participant in this focus group added that the AD should be produced in close cooperation with the content creators or it would fall entirely under the purview of the creative team.

Regarding directing audience attention toward the narratologically-relevant elements, the responses obtained from persons with sight loss suggest that the implementation of spatial sound would be of value; it could allow them to orientate themselves and locate different elements inside the storyworld. Furthermore, the responses suggest that spatial audio could improve their immersion, which is central to experiencing 360° content.

An interesting solution was proposed by one participant in the focus group in Kraków with the suggestion of preparing visiting paths to overcome the challenge of orientating themselves within the storyworld. More specifically, participants could choose a "guide me" option which would guide them through the main action and the most important descriptions of the surroundings. The advantage of this solution is that persons with sight loss would be guided through the visual scene towards the elements that are important from a narratological point of view.

Overall, participants in both focus groups were interested in the integration of AD in 360° videos and in the possibility of activating it in an interactive way. They were mostly interested in educational or training content to be employed by museums and other institutions. They also deemed this technology more adequate for short-duration video material.

3.5. Conclusions

This article sought to discuss whether AD requires a specific approach in 360° content. To this aim, a two-step research methodology was designed in which a bibliographical review was followed by focus groups, and a user-centered model of translation was adopted (Suojanen, Koskinen and Tuominen 2015).

The analysis in the first part of the article shows that the main challenges faced when describing narrative in 360° videos relate to guiding users with sight loss within the storyworld and prioritizing visual information when conducting a source text analysis.

The results obtained from the focus groups suggest possible approaches to AD in 360° content that could provide solutions to the aforementioned challenges. Such approaches could include the option of pausing the video to learn more about the storyworld by listening to additional AD tracks triggered by head movements. Another possible approach that could be combined with the one mentioned above would be the integration of spatial sound, as it can enable persons with sight loss to be guided towards the relevant narrative elements. Additionally, the results from the focus groups in terms of content selection suggest that due to the complexity of such content, cooperation with content creators should be sought when providing descriptions.

The findings discussed in this article are just a starting point in this new dynamic field. The study is limited in scope due to the early adoption status of this media format and the narrative possibilities of 360° videos not being fully delineated yet. It paves the way for more complex analyses with larger samples, which would combine theoretical approaches with experimental studies. Further studies could include but not be restricted to: measuring the effect of spatial sound in AD and audio subtitling on actual audiences, measuring user presence when faced with diverging AD strategies and testing the new production workflow with professional audio describers. As the theme caught the interest of users in both Barcelona and Kraków, elements of interactivity could be tested in future reception studies.

As the field of Audiovisual Translation Studies is being reshaped by immersive content appearing in the market, the knowledge of accessibility in such content could be integrated in the following years in updated guidelines for audio describers in addition to courses offered by institutions for the training of future professionals. The ultimate aim is that access services – AD in this case – are produced in a way that meets quality standards, and enable all audiences to enjoy innovative media such as 360º videos.

Funding information

ImAC has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement 761974 (http://www.imac-project.eu/). 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 Anita Fidyka'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

American Council of the Blind. 2009. "Audio Description Standards." American Council of the Blind. Accessed November 21, 2018. http://www.acb.org/adp/docs/ADP_Standards.doc

Bal, Mieke. 2009. *Narratology: Introduction to Theory of Narrative*. Toronto: University of Toronto Press.

Barbour, Rosaline. 2008. *Doing Focus Groups*. London: SAGE Publications.

Biocca, Frank. 1997. "The Cyborg's Dilemma: Embodiment in Virtual Environments." *Journal of Computer-Mediated Communication* 3 (2). DOI: 10.1111/j.1083-6101.1997.tb00070.x

Bordwell, David, and Kristin Thompson. 1990. *Film Art: An Introduction.* New York: McGraw-Hill College.

Braun, Sabine. 2007. "Audio Description from a Discourse Perspective: a Socially Relevant Framework for Research and Training." *Linguistica Antverpiensia* 6: 357–369.

Braun, Sabine. 2008. "Audio Description Research: State of Art and Beyond." *Translation Studies in a New Millennium* 6: 14–30.

Damiani, Jesse, and Dylan Southard. 2017. "Writing for VR: The Definitive Guide to VR Storytelling." *VR Scout*, October 2, 2017. Accessed November 21, 2018. https://vrscout.com/news/writing-vr-definitive-guide-vr-storytelling/

Dolan, Devon, and Michael Parets. 2015. "Redefining the Axiom of Story: The VR and 360 Video Complex." *Medium*, December 20. Accessed November 21, 2018. https://medium.com/@devon.michael/redefining-the-axiom-of-story-the-vr-and-360-video-complex-bee3c20d69df

Dooley, Kath. 2017. "Storytelling with Virtual Reality in 360-Degrees: A New Screen Grammar." *Studies in Australasian Cinema* 11 (3): 161–171.

European Broadcasting Union. 2017. "Virtual Reality. How Are Public Broadcasters Using It?" Accessed November 21, 2018. https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it

Fidyka, Anita, and Anna Matamala. 2018. "Audio Description in 360° Videos: Results from Focus Groups in Barcelona and Kraków." *Translation Spaces* 7 (2): 285–303.

Fresno, Nazaret. 2012. "Experimenting with Characters: An Empirical Approach to the Audio Description of Fictional Characters." In *Audiovisual Translation and Media Accessibility at the Crossroads*. *Media for All 3*, edited by Aline Remael, Pilar Orero and Mary Carroll, 147–171. Amsterdam: Rodopi.

Fresno, Nazaret. 2014. "La (re)Construcción de los personajes fílmicos en la audiodescripción." PhD diss., Universitat Autònoma de Barcelona.

Fryer, Louise. 2016. *An Introduction to Audio Description. A Practical Guide*. London: Routledge.

Fryer, Louise, and Jonathan Freeman. 2013. "Cinematic Language and the Description of Film: Keeping AD Users in the Frame." *Perspectives* 21 (3): 1–15.

Gödde, Michael, Frank Gabler, Dirk Siegmund, and Andreas Braun. 2018. "Cinematic Narration in VR – Rethinking Film Conventions for 360 Degrees." In *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage and Industry*, edited by Jessie Y. C. and Gino Fragomeni, 184–201. Cham: Springer.

Greco, Gian Maria. 2016. “On Accessibility as a Human Right, with an Application to Media Accessibility.” In *Researching Audio Description: New Approaches*, edited by Anna Matamala and Pilar Orero, 11–34. London: Palgrave.

Greco, Gian Maria. 2018. “The Nature of Accessibility Studies.” *Journal of Audiovisual Translation* 1: 205–232.

Herman, David. 2002. *Story Logic: Problems and Possibilities of Narrative*. Lincoln: University of Nebraska Press.

IRT Lab. n.d. “Interactive Content.” *Interactive Lab Website*. Accessed December 11, 2018. https://lab.irt.de/demos/object-based-audio/interactive/

Jerald, Jason. 2015. *The VR Book: Human-centered Design for Virtual Reality*. San Rafael: Morgan & Claypool.

Jones, Sarah. 2017. “Disrupting the Narrative: Immersive Journalism in Virtual Reality.” *Journal of Media Practice* 18 (2–3): 171–85.

Kruger, Jan-Louis. 2010. “Audio Narration: Re-narrativising Film.” *Perspectives* 18 (3): 232–248.

Kruger, Jan-Louis. 2012. “Making Meaning in AVT: Eye Tracking and Viewer Construction of Narrative.” *Perspectives* 20 (1): 67–86.

Lee, Nicole. 2017. “Pixar’s VR Debut Takes You Inside the Entrancing World of ‘Coco’.” *Engadget*, November 15. https://www.engadget.com/2017/11/15/pixar-coco-vr/

Maszerowska, Anna. 2014. “The Verbalizations of the Effects of Light and Contrast in Audio Description.” PhD diss., *Autonomous University of Barcelona.*

Maszerowska, Anna, Anna Matamala and Pilar Orero, eds. 2014. *Audio Description: New Perspectives Illustrated*. Amsterdam: John Benjamins.

Matamala, Anna, and Aline Remael. 2015. “Audio Description Reloaded. An Analysis of Visual Scenes in 2012 and Hero.” *Translation Studies* 8 (1): 63–81.

Mazur, Iwona. 2014. “Gestures and Facial Expressions in Audio Description.” In *New Perspectives Illustrated,* edited by Anna Matamala and Pilar Orero, 179–197. Amsterdam: John Benjamins.

Nielsen, Lasse T., Matias B. Møller, Sune D. Hartmeyer, Troels Ljung, Niels C. Nilsson, Rolf Nordahl and Stefania Serafin. 2016. “Missing the Point: An Exploration of How to Guide Users’ Attention during Cinematic Virtual Reality.” *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 229–232. New York: Association for Computing Machinery. DOI: 10.1145/2993369.2993405

Orero, Pilar. 2011. “Audio Description for Children: Once upon a Time there Was a Different Audio Description for Characters.” In *Entre Texto y Receptor: Accesibilidad, Doblaje y Traducción*, edited by Elena di Giovanni, 169–184. Frankfurt am Main: Peter Lang.

Perego, Elisa. 2014. “Film Language and Tools.” In *Audio Description. New Perspectives Illustrated,* edited by Anna Maszerowska, Anna Matamala and Pilar Orero, 81–101. Amsterdam: John Benjamins.

Remael, Aline. 2012. “For the Use of Sound: Film Sound Analysis for Audio Description: Some Key Issues.” *MonTI* 4: 255–276.

Remael Aline, Nina Reviers, and Gert Vercauteren, eds. 2014. *Pictures Painted in Words. Audio Description Guidelines*. Trieste: Edizioni Università di Trieste. Accessed November 22, 2018. http://www.adlabproject.eu/Docs/adlab%20book/index.html

Reyes, María Cecilia. 2017. “Screenwriting Framework for an Interactive Virtual Reality Film.” Paper presented at the Immersive Learning Research Network, Coimbra, June 26–29.

Romero-Fresco, Pablo. 2013. “Accessible Filmmaking: Joining the Dots between Audiovisual Translation, Accessibility and Filmmaking.” *Journal of Specialised Translation* 20: 201–223.

Rothe, Sylvia, Heinrich Hußmann and Mathias Allary. 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: Association for Computing Machinery. DOI: 10.1145/3139131.3143421

Slater, Mel, and Martin Usoh. 1993. “Representations Systems, Perceptual Position, and Presence in Immersive Virtual Environments.” *Presence: Teleoperators and Virtual Environments* 2 (3): 221–233.

Suojanen, Tytti, Koskinen, Kaisa, and Tiina Tuominen. 2015. *User-Centered Translation*. London and New York: Routledge.

Syrett, Hannah, Licia Calvi, and Marnix van Gisbergen. 2017. “The Oculus Rift Film Experience: a Case Study on Understanding Films in a Head Mounted Display.” In *Intelligent Technologies for Interactive Entertainment*, edited by Ronald Poppe, John-Jules Meyer, Remco Veltkamp and Mehdi Dastani, 197–208. Cham: Springer.

Szarkowska, Agnieszka, and Piotr Wasylczyk. 2014. “Audiodeskrypcja autorska.” *Przekładaniec* 28: 48–62.

Udo, John-Patrick, and Deborah Fels. 2010a. “The Rogue Poster-Children of Universal Design: Closed Captioning and Audio Description.” *Journal of Engineering Design* 21 (2–3): 207–221.

Udo, John-Patrick, and Deborah Fels. 2010b. “Universal Design on Stage: Live Audio Description for Theatrical Performances.” *Perspectives* 18 (3): 189–203.

Walczak, Agnieszka, and Louise Fryer. “Creative Description: The Impact of Audio Description Style on Presence in Visually Impaired Audiences.” *British Journal of Visual Impairment* 35 (1): 6–17.

Vandaele, Jeroen. 2012. “What Meets the Eye. Cognitive Narratology for Audio Description.” *Perspectives* 20 (1): 87–102.

Vercauteren, Gert. 2012. “A Narratological Approach to Content Selection in Audio Description. A Strategy for the Description of Narratological Time.” *MonTI* 4: 207–231.

Vercauteren, Gert. 2016. “A Narratological Approach to Content Selection in Audio Description.” PhD diss., University of Antwerp.

Vercauteren, Gert, and Pilar Orero. 2013. “Describing Facial Expressions: Much More than Meets the Eye.” *Quaderns. Revista de Traducció* 20: 187–199.

Vercauteren, Gert, and Aline Remael. 2014. “Spatio-temporal Settings.” In *Perspectives Illustrated*, edited by Anna Matamala, 61–80. Amsterdam: John Benjamins.

Chapter 4. Article 3

Production of access services in immersive content: Understanding the needs of audio describers

4. Article 3

Abstract

360° content, offered more and more frequently by various broadcasters and culture institutions, should cater for the needs of all members of our society, including persons with sight loss. So far, however, the question of providing audio description (AD) in such content, has been under researched. This study aims to report the results of the usability study of the prototype AD editor developed within the Immersive Accessibility (ImAc) project, which allowed us to gain insights into the needs of professional audio describers when working with 360° content. The editor is an online tool which allows describers to choose an appropriate sound type for AD, and place AD segments in the 360° sphere. The study was conducted online and data was collected by means of a demographic pre-questionnaire and a post-questionnaire, consisting of a System Usability Scale and additional preference questions. The results obtained provide valuable feedback on how to improve the functionality of the tool to meet the needs of its users. They also indicate the need for guidance when selecting content to be described in this media format, which suggests that AD in immersive content could be integrated into AD guidelines or specific courses offered by training institutions. The results of this study are just a starting point in the field of immersive accessibility, hence the recommendation for further research on the subject of accessibility in this media format.

Keywords: Audiovisual Translation, Media Accessibility, audio description, 360° videos, usability

4.1. Introduction

Audio description (AD) is an access service that is considered as a way of retelling a story: it translates the visual channel to a verbal mode (Maszerowska, Matamala, & Orero, 2014; Fryer, 2016; Snyder, 2008). It is used widely to render standard, two-dimensional audiovisual (AVT) products, such as films and TV programs, accessible. It is also used to describe, for instance, artworks in museums or live events, rendering cultural property accessible for those who cannot access the visuals, and to other groups at risk of social

exclusion (Greco, 2016). So far, AD has been mainly provided for two-dimensional content (Fidyka and Matamala, 2018a). Although more interactive forms of AD exist, such as AD in the theatrical environment (Fryer and Freeman, 15) or AD in planetariums – in which content is displayed in a dome, surrounding the viewer – research on AD in immersive media is practically non-existent nowadays (Fidyka and Matamala, 2018a). The few studies conducted to date include audio description in 3D cinema (Greening, 2011), or the integration of haptics in audio description (Viswanathan, McDaniel, Krishna, & Panchanathan, 2010). Regarding other access services, the implementation of subtitles has been researched (Agulló, Matamala, & Orero, 2018; Agulló and Matamala 2019; Rothe, Tran, & Hußmann, 2018) and some solutions on the implementation have been proposed by major distributors (Brown, 2017). However, the issue of implementing AD in 360° content has not been addressed so far.

All new technologies appearing on the market should be accessible to ensure that all members of our society have access to culture and arts, as specified in documents on Human Rights, such as the Universal Declaration of Human Rights and the UN Convention on the Rights of Persons with Disabilities. To ensure that such highly visual environments are accessible for persons with sight loss, devices used to consume immersive content (eg. glasses) should be made accessible and access services such as audio description should be provided.

Although immersive media are still emerging on the market, they already have a wide array of creative formats (Allen and Tucker, 2018). One of them is 360° videos, belonging to Virtual Reality (VR). These videos are typically between 5 and 15 minutes long, and they are «the most tightly authored among all VR formats» (Allen and Tucker, 2018:17) that have particular market potential. It means that the story – which is told with a central protagonist's «rise, fall and resolution» (ibid.) – is linear, often driven by dialogue and pre-scripted by a director or content creator (ibid.).

In 360° videos, also referred to as omnidirectional or spherical videos, the main challenge lies in storytelling techniques, which are still being defined. Sighted users can access 360° content by means of head-mounted displays, i.e. special glasses to access the Virtual Reality content, which allow them to feel as if they are inside a sphere, while the linear story pre-scripted by a director unfolds all around them (Fidyka and Matamala, 2018b; Dooley 2017). As different events can occur at different angles of the visual scene, guiding the sighted

users inside the virtual world towards the main action is considered challenging (Rothe, Hußmann, & Allary, 2017; Jones, 2017; Syrett, Calvi, & van Gisbergen, 2017), and various storytelling techniques are being researched (Gödde, Gabler, Siegmund, & Braun, 2018; Jerald, 2015). In any case, sighted audiences can look around the sphere, following the main story or ignoring it.

The media format is mostly visually-driven, with images being triggered by head movements. Finding a way to guide persons with sight loss effectively inside the sphere increases the challenge (Fidyka and Matamala, 2018a, 2018b). The technology of spatial sound, which gives audiences a three-dimensional soundscape of the elements presented in the content may prove of value, as well as different scripting styles. The implementation of spatial sound, including ambisonics (Johansson, 2019) in AD provided for 360° videos is being researched within the Immersive Accessibility (ImAc) project. Its application in AD is also being investigated by other researchers (López, Kearney, & Hofstädter, 2016; Portillo, 2018).

ImAc is an H2020 project based on a user-centred methodology (Suojanen, Koskinen & Tuominen, 2015). Thanks to this methodology, end users and professional audio describers collaborate with project partners at every stage of the project. In its early stages, feedback from users was gathered through a series of studies based on focus groups (Fidyka and Matamala 2018a, 2018b). These qualitative studies allowed researchers to define the implementation of access services and editing tools, and their results suggest that spatial sound can serve as a tool to facilitate orientation within the sphere. As far as the production of AD in immersive content is concerned, the development of an editing tool for this access service is one of the aims of the project.

To respond to the challenge of guiding persons with sight loss in this new medium, a new method of producing audio description needed to be proposed prior to the development of the AD web editor. Based on the early feedback from end users, it was decided that when providing AD by means of the web editor, audio describers would be able to choose between three different audio description modes, associated with specific sound and scripting features, as discussed in detail in the next section. This choice will allow describers to choose the most appropriate style for each content (Allen & Tucker, 2018). This new approach also needed to be tested with professional describers, as it impacts on their current workflow.

This article will discuss the methodology and results of the usability tests of the AD web editor in its first prototype version, which allowed us to understand the needs of professional audio describers when producing AD in this media format. To contextualize the study, the main features of the prototype editor, developed by Anglatècnic, including the functions specific to immersive environments, will be described in the next section. Section 4.3 explains the methodology of the test, and section 4.4 offers the discussion of its results. Although the article reports on the results of a usability test, the interest lies in the interesting insights provided by professionals on the specific challenges the immersive content poses when audio describing, and how this can impact their current practices.

4.2. Features of the web editing tool

The following section describes different features of the first version of the AD editing tool. Firstly, standard features that can also be found in other existing editors are discussed. In the second part, features specific to the medium under discussion are explained.

4.2.1. Standard features

The AD editor is an online tool, and is comprised of different sections. In the central section of the editor, audio describers can edit the AD script, add timecodes and preview the video with AD, as shown on Figure 2.

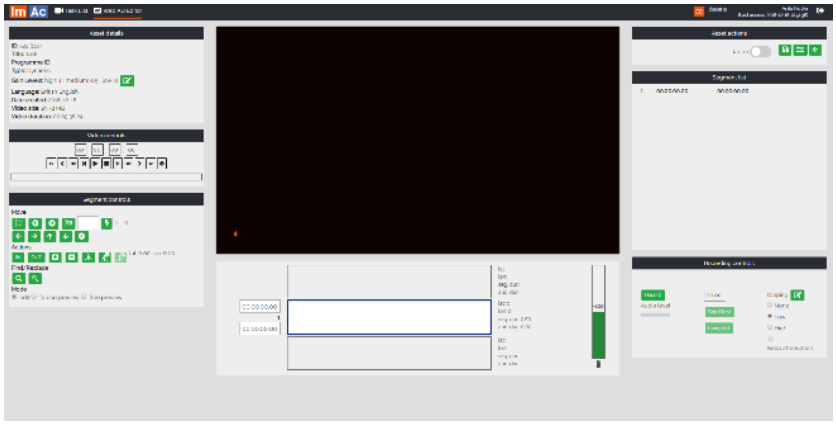


Figure 2. General view of the AD web editor

Another section, «Asset details», displays the basic information about the video, such as its name, size and language. Also, specific messages appear in that section in case of errors.

Figure 3 shows the «Video controls», which allow audio describers to play, pause or stop the video, and navigate through it. Video controls have their own shortcuts, which were pre-set for the test, but will be customizable in future versions of the AD editor.

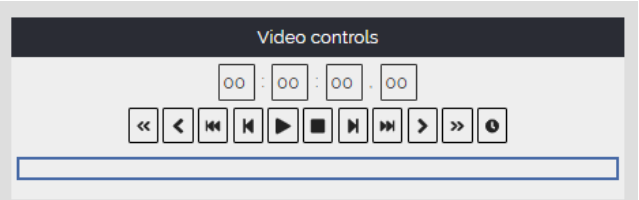


Figure 3. Video controls

The next section, «Segment controls», is related to audio description. It allows audio describers to add, remove and navigate through AD segments as well as add or remove timecodes (Figure 4). It also allows audio describers to set the angle, as described in section 4.2.2. Similarly to «Video controls», «Segment controls» also have their own shortcuts.

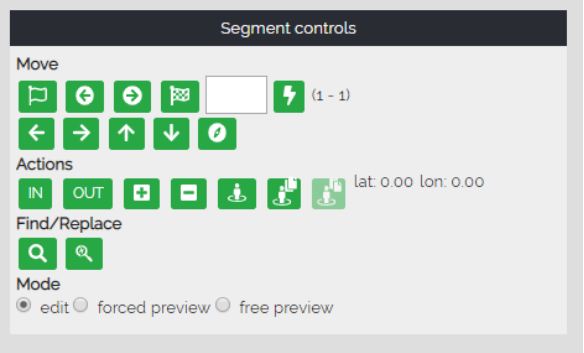


Figure 4. Segment controls section

As already mentioned, audio describers edit the script below the video player (Figure 5). To the left of the script, audio describers can set timecodes for a given AD segment. The number of the segment is also displayed in this section. To the right of the script editing area, the longitude and latitude of the current segment can be found. Also, the duration of the segment is displayed below. In the corner of the right side of this section, reading speed is displayed. When the reading speed is appropriate, the colour is green. When the number of characters per second is too high, it becomes red.

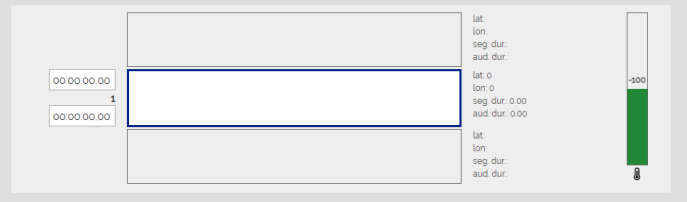


Figure 5. Script editing area

On the right side of the editor, there are three sections, named respectively: «Asset action», «Segment list» and «Recording controls» (Figure 6). The section «Asset action» allows to save audio files, and go back to the main page of the editing interface. «Segment list» contains the AD script with timecodes and a segment number. When AD for a given segment is recorded, the colour of this segment changes to green. When all segments are green, it means that all segments are recorded correctly. AD segments can be recorded by pressing the «Record» button. A countdown is provided for the recording to show audio describers how much time is left according to the timecodes set by them. Below the «Record» button, the audio level of the recording can be checked. After the AD is recorded, the recording can be previewed in two tests: one starting two seconds before the timecode, and another one starting 5 seconds before the timecode.

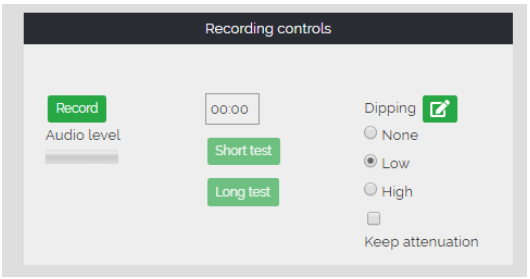


Figure 6. Recording controls

Additionally, fading can be chosen by audio describers. It refers to the decrease in the volume of the video when the AD is playing. In this regard, audio describers can choose between: none, low or high attenuation. The «None» option does not reduce the volume of the video. Conversely, when the option «High» is chosen, it will result in the significant reduction of the volume of the video. When the box «Keep fading» is checked, the volume of the main audio will be lowered until the next AD segment.

4.2.2. Features specific to immersive content

The AD web editor allows professional users to produce, preview and record audio description for 360° videos. As such, it includes functions which are specific to immersive content, which happens all around the 360° sphere.

Because of the characteristics of this media format, three types of AD can be created for each 360° video clip when working with the editor. They differ regarding the placement of spatial sound, and it is also recommended that audio describers use a different way of scripting for each of them. For the purposes of this project, these three AD types are provisionally referred to as: «Classic», «Static» and «Dynamic» (Figure 7). When professional users access the web editor, they choose one of them. The first sound option allows users to hear the AD as if it was coming from above their heads. It is the one that more closely mimics what is usually provided in standard, 2D, audio described content, and such way of scripting is recommended. The second one, «Static», is heard from the left or right side, as if someone was standing beside the user. As the AD is heard as if coming from a short distance from the user's location, it is recommended to write the AD in a non-standard way, with a stronger involvement of the describer. The third AD type, «Dynamic» allows audio describers to place the descriptions at different angles of the sphere, and it can be used to locate a character, object or event. In other words, when a user plays a video with this type of AD, they will hear the AD coming from a specific point in the space. We believe that this option can guide users effectively within this highly visual medium, and will also allow them to feel more present in the content. This is an innovation in the field of audio description, as it can provide additional information on the location of the visuals being described by means of audio cues. As far as the scripting is concerned, a minimum audio description is recommended to allow users to be guided by the sound of the video.

Figure 7. AD types

Although many of the AD web editor's features are similar to the features of others subtitling or audio description editors, there are some new features that needed to be

implemented, taking into account the specificity of the environment. One of such functions is «Set angle», a function implemented within the «Dynamic» AD type. This function allows users to assign a given AD segment to a specific angle within the sphere, specified by latitude and longitude. It means that the user, when watching the content with AD, will hear an AD of a given object, or event from the direction set by an audio describer, as it stays 'tied' to that part of the 360º sphere. This function, which can help users of access services to orient themselves inside the sphere, was implemented based on requirements defined in the focus groups (Fidyka and Matamala, 2018a). When an audio describer wants to create a «Dynamic» AD type, they need to look inside the sphere for a desired angle, and set it by clicking on a special button (Figure 8), or by a combination of keyboard shortcuts (customizable in the future version of the editor).

This new option will change the current workflow of producing AD. So far, audio describers needed to write AD instances between the dialogues, and mark their timecodes. Those audio describers whose task was also to voice the AD, needed to record it and preview it. When producing AD with a «Dynamic» AD type, besides following all the aforementioned steps, audio describers need to set an angle for every AD instance. In other words, they need to decide from which angle the linear AD will be heard by the end users.

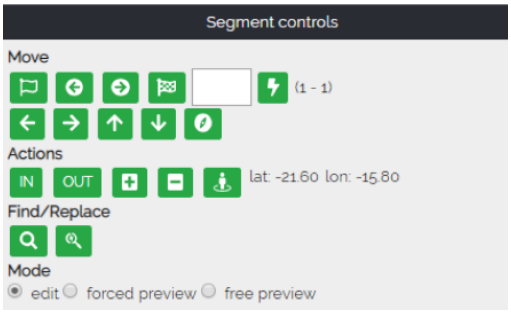


Figure 8. Set angle function and preview modes

As far as the preview of AD is concerned, two modes can be used for verification in the «Dynamic» AD type, namely «Forced preview» and «Free preview». In the first verification mode, AD and angle are bound to the video, which means that the video changes the angle during playback in the AD web editor. In other words, audio describers do not need to navigate through the video to find the angles they have set, as the software does it automatically. Thanks to this preview mode, audio describers are able to check whether the angles they set are correct. In the «Free preview», however, only the AD

segments are bound to the timecodes. It means that audio describers can freely navigate the video during the playback, as during the editing.

As the functionalities described above are new, further testing is required to ascertain that they meet the needs of actual audio describers. This is why usability tests were conducted on the «Dynamic» AD type, in which they are implemented. The overall goal of our study was to assess the progress made in the development of the audio description editor. In particular, we wanted to learn whether the tool meets the needs of its actual users, so that its future version can cater for them more effectively. Additionally, our aim was to gather feedback on the work of the describers in relation to immersive environments and learn how such content will impact on their workflow.

4.3. Method

The usability test of the AD editor was performed online from the 24th of September to the 19th of October 2018. It aimed at receiving feedback from audio describers from different countries. The test instructions were given in English, and participants were asked to fill in the questionnaires in the same language. They were, however, requested to provide audio description in their native languages, which allowed us to gather feedback from users from diverse countries, with different AD traditions. The response rate was 70.59%. The study was conducted voluntarily by professionals in accordance with ethical procedures approved by the Ethics Committee at Universitat Autònoma de Barcelona (UAB). Participants were informed about the aim and context of the study, and gave their consent before the test. Data are confidential, and the privacy of participants is ensured.

4.3.1. Participants

24 participants completed the test, 15 females and 8 males, plus one participant who preferred not to reply to this question. Their ages ranged from 25 to 64 years. The mean age was 36.71 years old. Their main languages were Catalan (1 respondent), Spanish (6 respondents), both languages (2 respondents), Bosnian (1 respondent), English (6 respondents), Dutch (2 respondents), Polish (3 respondents), German (2 respondents), and Swedish (1 respondent). They were mainly AV translators, audio describers, AD and Media Accessibility supervisors, researchers and project managers. 21 participants had completed

university studies, one participant had further education, and two participants had secondary education.

4.3.2. Materials

Measures included usability and preferences. To obtain demographic data, two online questionnaires were prepared: a pre-questionnaire, and a post-questionnaire. They were sent to participants through separate e-mails upon agreeing to participation, as explained in section 4.3.

The post-questionnaire consisted of two parts: the System Usability Scale (SUS) questionnaire (Sauro & Lewis, 2016) and a preference questionnaire, with both closed and open questions which aimed at gathering additional user feedback. The SUS questionnaire chosen as a measure of usability contains 10 items, each with five steps: from strongly disagree to strongly agree (Sauro & Lewis, 2016, p. 198). It was available in English.

Regarding preferences, a specific questionnaire containing nine items was developed. It asked what participants appreciated or disliked about the editor, the possible improvements and missing functionalities, the level of difficulty related to the «Set angle» function, and the usefulness of preview modes. It also gathered the opinions of participants regarding whether or not describing 360º content is more time-consuming than that of standard content, and whether 360º videos will impact on their work in the coming years. The last question provided space for additional comments.

Each participant was assigned with one clip to be audio described with a «Dynamic» AD type. The video chosen was an initial 1-minute excerpt of a fictional story suitable for audio description: there was enough space to produce audio description between the music and other sounds and, most importantly, the story developed at different angles of the sphere, which allowed us to test the «Set angle» function. The video was played in low resolution (720s) to avoid overloading the server, and to make the audio describing task smoother.

Other materials used in the study included a document with detailed instructions on how to perform the tasks, and a step-by-step user guide with screenshots on how to use the web editor. They were provided to participants by e-mail, as discussed in the next section.

4.3.3. Procedure

Prior to the study, a small-scale pilot test was performed with two users to evaluate the experimental protocol to be used on a larger scale. The participants met the criteria for inclusion in the sample, as they were professional audio describers. However, no changes were made to the final methodology because no problems arose in the pilot test development.

The study was developed online, and consisted of the subsequent steps. Participants were contacted through social media and personal contacts. Upon agreeing to participation, they received separate e-mails from the researchers, outlining the broader context and the procedure of the study. These e-mails included: (1) access to the editor, (2) a link to instructions on how to complete the test, (2) a link to a user guide, (4) links to the online questionnaires. Participants were informed about the exact order in which they should complete the tasks and fill in the questionnaires.

Participants completed the study in the following order. First they gave their informed consent to take part in the study. Then, the information that their data would be kept confidential was provided by the researchers. They were then asked to fill in the demographic questionnaire, perform a series of tasks in the editor, and fill in the post-questionnaire.

Participants were asked to complete the following tasks: (1) log in to the editor and open the assigned video, (2) audio describe the video excerpt in the user's native language, (3) preview the video in forced preview mode, (4) preview the video in free preview mode, and (5) save the AD and go back to the main window. Task 2 mentioned above («Audio describe the video excerpt») consisted of a series of sub-tasks: (1) add AD instances with correct timecodes, (2) set the angle for each AD instance, (3) record the AD segments produced, (4) insert one AD segment between two existing ones, and (5) delete two AD segments.

Once the tasks were completed, participants were asked to fill in the post-questionnaire, consisting of SUS and preference questions. At the end of the study, participants were thanked, and information on how to obtain feedback was provided.

4.4. Results and discussion

Below we present results obtained from pre- and post-questionnaire as well as their discussion.

4.4.1. Results from pre-questionnaire: user profile

The pre-questionnaire focused on gathering socio-demographic information. Regarding the previous experience of participants in the AD field, most of the participants had received training in AD (91.67%). Nine participants (37.5%) had produced more than 300 hours of AD content, four participants (16.67%) had produced between 151 and 300 hours of AD content, four participants (16.67%) had produced between 51 and 150 hours, and seven participants (29.2%) had produced less than 50 hours of audio description. However, as little as four participants had had previous experience in audio describing a 360º video.

When asked about the usage of the AD editors, results show that not all audio describers use specific editing tools on a daily basis, with 10 of the participants reporting using only text processors and video players when producing AD.

The pre-questionnaire also aimed at determining cyber potential of research participants. When asked which devices they use on a daily basis, almost all participants (23) confirmed using mobile phones (95.83%); 21 participants use laptops (87.5%); 15 participants use TVs (62.5%), 14 participants use PCs (58.33%), eight of them use tablets (33.33%), one uses HMD (4.17%), and one participant chose the option «Other» (4.17%).

Unsurprisingly, most of participants do not use VR on a daily basis. To establish how frequently participants use VR, we asked them: 'How often do you watch Virtual Reality content (for instance, 360º videos)?' The results show that most of the participants (21, 87.5%) have never watched such content on a smartphone plugged into a head-mounted display or in a head-mounted display (20, 87.5%). Only one participant declared occasionally consuming Virtual Reality content on their smartphone, one participant occasionally uses a tablet to consume VR content and, regarding PCs, two participants (8.33%) use this device occasionally.

The next question asked about the reasons behind not using VR content or using it only occasionally. In this question, six participants (25%) replied that they are not interested,

three participants (12.5%) replied that it is not accessible, 12 participants (50%) replied that they have not had the chance to use it, two participants (8.33%) chose the option «Other reasons», and one participant (4.17%) did not provide any answer to this question. One of the participants provided an additional comment: «I don't normally access this content, I thought there were just a few, although I was surprised when accessing the project.»

Research participants were also asked to state their level of agreement with the statement «I am interested in Virtual Reality content (such as 360º videos)». The results show that four participants strongly agree with the statement (16.67%), eight participants agree (33.33%), nine participants neither agree nor disagree (37.5%), one participant disagrees (4.17%), and two participants strongly disagree (8.33%). Finally, when asked if they own any device with which to access Virtual Reality content, 10 participants replied that they do not (41.67%), five replied that they do not know or prefer not to reply (20.83%), and eight replied that they do (33.33%).

4.4.2 Results on usability

Regarding the results on usability, the score obtained in the SUS questionnaire is 55.9, which is considered below average, with a score of 68 or more considered as average. The obtained score corresponds to the percentile rank of 19%, and when converted to the letter grades, the obtained mark is D (Sauro & Lewis, 2016, pp. 203–204). This shows that the prototype web AD tool still has a lot of potential for improvement, and makes user testing at this stage even more relevant, as users are contributing to the definition of requirements as the tool is being developed.

The second part of the post-questionnaire focused on gathering data on users' preferences, and the results will be discussed question by question in the following section.

4.4.3. Results from preference questions

As far as the first question is concerned («What did you like the most about the AD editor?»), two of the participants (8.33%) highlighted the fact that the whole process of producing AD takes place in one piece of software. Four (16.66%) of the comments pointed to the fact that the most important functions (video, AD segments and recording) are displayed on one page, which facilitates the production of AD. In this regard, one of the

participants commented: «It is quite easy, it has shortcuts and everything is visible and easily accessible on one page». Nine participants (37.5%) commented positively on the interface and its layout. They used the following words to describe it: «very clear», «simple», and «easy to understand». One of the participants (4.17%) also positively assessed that the software is available online.

In the second and third questions, we asked participants which functions are the least useful, and how they could be improved. Nine audio describers (37.5%) pointed to the second question («What did you like less about the AD editor?») the problems encountered in the recording and preview modules. Also in the third question («What do you think could be improved, and how?»), problems with the video and buttons which would freeze or play with delay, were mentioned by five describers (20.83%). One comment (4.17%) in the third question suggested that a better video quality would be needed in order to describe all details. In this case, this was due to the testing conditions, as a higher quality video could overload the editor, and slow down the AD production process. Additionally, five (20.83%) participants suggested in both questions that they would prefer a different, more intuitive configuration of the shortcuts – or that they would prefer to customize the shortcuts themselves. They reported that this change is essential, as using the shortcuts they are accustomed to would allow them to work more efficiently. As already mentioned, although the shortcuts were not customizable in the prototype version of the editor, audio describers will be able to adapt them to their needs in the final version. Importantly, one comment (4.17%) in the second question pointed out that it would be necessary to preview the produced AD not only in the web editor, but also in a head-mounted display. We also believe that this option would be worth implementing, as it would allow audio describers to preview their work in the way in which end users will be consuming the content.

The following question asked participants about other functionalities that could be implemented. In this regard, participants suggested many improvements that are technically feasible. Among them, a waveform that would indicate music and other sounds, was suggested by two participants (8.33%) as a way to time audio description more precisely. This question was put forward in one comment (4.17%) as follows: «Perhaps one [function] in which you could see the lines corresponding to sound (…) is very useful to ensure that AD does not interfere with dialogues, or other sounds in the film.» Regarding combining AD with the other sounds of the video, one of the participants suggested that more options

for fading the audio of the video would be needed, as there are currently only three options which are possible: «none», «low» and «high fading». This comment seems particularly relevant, as the audio volume of the AD track is a critical aspect when ensuring that descriptions can be heard clearly throughout the video or film.

An important improvement was suggested by another participant, who considers that a synchrony between AD segments and video would be needed to facilitate the work. It means that if someone clicks on an AD segment on the «Segment list», the video should move to its timecode.

Other suggested improvements focused on both standard functionalities, and functions specific to the medium under discussion. They included: moving back or forward 5–10 frames at a time, a map in which all the locations with ADs assigned to them would be located, being able to merge or separate segments, setting a minimum separation between AD segments and dialogues or other sounds, and seeing the actual reading speed of the AD. Another participant suggested that using earphones during recording would be needed.

Importantly, an option to export the script to a text file for a professional recording was suggested in one comment (4.17%), as not all audio describers record AD themselves. For example, in Poland, audio describers send their script to a professional responsible for recording (Chmiel & Mazur 2014). Audio describers who do not record the AD themselves, mark parts of the dialogue in the script or sounds that proceed, or follow a given AD segment. They also indicate in brackets how the given instance should be read, for example slowly or rapidly. The need to mark the part of dialogue in the script intended for recording by a professional was put forward in the comment of one of the participants: «I am used to using sound and dialogue cues rather than 'In times' only. I find it helpful to see the dialogue cue that leads into a description».

Regarding the innovative «Set angle» function, the results suggest that most of the participants (75%) found it easy to use, despite one participant reporting technical problems with their laptop. Another response suggests that one participant would like to set the angle for only some events, and not for all AD segments: «(...) I would like to have more freedom. The tutorial tells me we need an angle for each segment. I would like to have an angle only for very important situations». Therefore, it would be worth exploring a combination of the so-called «Dynamic» and «Classic» audio descriptions. Moreover, it would be worth

exploring other spatial sound possibilities that allow one not only to set the angle from which the audio description will be heard, but also a specific position, ideally linked to an object.

As far as the preview modes – crucial to check if the set angles are correct – are concerned, five participants (20.83%) did not encounter any problems while using them, e.g. «(...) one allows you to move, the other one makes you see your fixed angles», but for 11 participants (approximately 45.8%) one or both preview modes did not work properly, and delay and visualisation problems were reported in the prototype version. Five participants (20.83%) did not provide an answer to this question.

Beyond the specific questions linked to the editor, it was interesting to gather general feedback on the experience of audio describing 360º content. When asked about whether it takes longer to describe 360º videos than standard content, most participants (79.2%) replied positively, adding that setting an angle for each AD segment needs additional time.

Although not entirely within the scope of this article, some participants addressed the question of content selection in this emerging medium, the question discussed in depth in Fidyka and Matamala (2018b). Similarly to the results of the focus groups organized to provide a first insight into this question, it was deemed challenging, as explained in one of the comments: «technically, I consider it easy to use, the problem is which angle is the most important to describe.» It was added in other comments that audio describers will need special guidance when deciding which elements are the most relevant and should be audio described. This comment is similar to the comments put forward in the focus group organized in Poland, during which participants suggested that cooperation with content creators should be sought when audio describing 360º videos (ibid.). In this regard, guidelines that would provide audio describers with guidance on what to describe, when to describe it and how would prove particularly useful, but they have not yet been created.

The last question in the post-questionnaire asked participants whether they consider that 360º videos would impact on their AD practice in the coming years. In this regard, participants expressed varying opinions. 58.3% of the participants answered affirmatively to this question. One of the comments further explained that the application for this medium is vast, and as more content is created, it should be provided with access services. Another comment pointed to the fact that this medium changes the current approach for producing

AD, which means that specific training, or guidelines should be offered to audio describers producing AD in this medium. As explained by one of the participants: «I wouldn't feel comfortable taking on a job like this without proper instructions from the client and/or relevant training». This opens an opportunity for training institutions, which should accept the challenge of training professionals not only in the traditional modes, but also in more innovative media.

Additionally, one of the participants provided a detailed report after completing the test. This participant's overall assessment was positive, but they suggested some technical changes. Most of the comments pointed to improvements that could be made in the standard functions of the prototype editor (such as including a more responsive video control bar, improving the navigation by timecode, moving between AD segments), but one comment focused on the functions specific to immersive environments: audio describers producing AD in this medium would need a globe («Mercator map projection») on which all the set locations could be seen on one map. Moreover, echoing the opinions voiced in the focus group in Kraków, discussed in Fidyka and Matamala (2018b), this participant suggested that there should be the opportunity to read the audio description text via speech synthesis during the preview. Importantly, previewing in a full-screen mode should be possible.

All in all, the results of both parts of the questionnaire have provided valuable feed-back to improve the preliminary version of the tool, thanks to a user-centric methodology in which users are asked for feedback during the development process. Although quantitative data regarding usability in these stages generally falls short of what is expected from a market tool, it provides a benchmark for comparison in future releases, and contributes to qualitative feedback.

4.5. Conclusions

The usability test of the AD web editor, developed so as to respond to the characteristics of 360º content, allowed us not only to assess the progress made in the development of the tool, but also to better understand the habits and needs of professional audio describers.

The results of the usability test show that although many of the features of the AD web editor have been assessed positively by the respondents taking part in the study, there is still room for considerable improvement in order to meet the needs of professional users.

Thanks to participants' suggestions, we found several aspects of the editor which could be improved. These include, for instance, customizing the shortcuts, previewing AD with speech synthesis or previewing it by means of HMD. The technical feasibility of some other proposals, such as a map on which AD segments could be seen, needs to be explored.

Additionally, the results confirm the findings of the previous studies conducted in the form of focus groups (Fidyka and Matamala, 2018a, 2018b): content selection in this media format is considered challenging by audio describers who need guidance on how to describe. Beyond the specific analysis of the tool, the test has shown some attitudes of current audio describers towards new immersive media: in this regard, they state that this new production system will be more time-consuming, and request specific guidelines and training.

The limitations of this study need to be acknowledged. One limitation is the sample size of 24 participants. The future usability study testing the final version of the tool could be conducted with a higher number of participants to obtain more reliable results. Secondly, the fact that the 360º video used as a stimulus in the study was watched on the flat screens of laptops or PCs with which the audio describers worked might have had impact on the results. Although this was due to the fact that the study was conducted online to reach professional audio describers from different backgrounds, watching the stimulus first in the HMD would be beneficial for audio describers to better understand the presented content. Another limitation could be related to the audio describers' lack of previous experience with AD software, which could have an impact on the obtained results.

Regarding future studies, more research will be needed on the implementation of spatial audio in AD in this media format to better respond to the needs of the end users of the access services. Additionally, it should be explored how spatial sound could enhance user experience when used in audio subtitles (AST), which serve audiences who cannot access the written subtitles and cannot understand the language of the original (Braun & Orero, 2010; Reviers & Remael, 2015). Also, more studies would be needed on AD in more complex VR types. Finally, an additional set of empirical tests will need to be conducted with end-users in order to assess whether the access services produced cater to their needs, allowing them to understand, enjoy, orient themselves, and feel present in the 360º content.

Acknowledgements

ImAC has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement 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 Anita Fidyka'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., Matamala, A., & Orero, P. (2018). From Disabilities to Capabilities: Testing Subtitles in Immersive Environments with End Users. Hikma, 17, pp. 195–220. doi: 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. Journal of Specialised Translation, 32.

Allen, C., & Tucker, D. (2018). Immersive Content Formats for Future Audiences. Retrieved from Digital Catapult on February 14, 2019: https://assets.ctfassets.net/nubxhjiwc091/5WypbBxFEQqS8I6WaKUQgg/ca25eb800e9f88a740079dbca623ca2c/Immersive_Content_Formats_for_Future_Audiences.pdf

Braun, S., & Orero, P. (2010). Audio Description with Audio Subtitling: An Emergent Modality of Audiovisual Localization. Studies in Translatology, 18(3), pp. 173–188. doi: https://doi.org/10.1080/0907676x.2010.485687

Brown, A. (2017). User Testing Subtitles for 360º Content. Retrieved from BBC (Research & Development) on February 14, 2019: https://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr

Chmiel, A. & Mazur, I. (2014). Audiodeskrypcja [Audio description]. Poznań: Wydział Anglistyki UAM.

Dooley, K. (2017). Storytelling with Virtual Reality in 360-degrees: a New Screen Grammar. Studies in Australasian Cinema, 11(3), pp. 161–171. doi: 10.1080/17503175.2017.1387357

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

Fidyka, A., & Matamala, A. (2018b). Retelling Narrative in 360º Videos: Implications for Audio Description. Manuscript submitted for publication.

Fryer, L. (2016). An Introduction to Audio Description. A Practical Guide. London: Routledge.

Fryer, L., & Johnatan, F. (2013). Cinematic Language and the Description of Film: Keeping AD Users in the Frame. Perspectives: Studies in Translatology, 21(3), pp. 1–15. doi: 10.1080/0907676X.2012.693108

Gödde, M., Gabler, F., Siegmund, D., & Braun, A. (2018). Cinematic Narration in VR – Rethinking Film Conventions for 360 Degrees. Virtual, Augmented and Mixed Reality. 10th International Conference, VAMR 2018. (pp. 184–201). Las Vegas: Springer. doi: 10.1007/978-3-319-91584-5_15

Greco, G. M. (2016). On Accessibility as a Human Right, with an Application to Media Accessibility. In A. Matamala & P. Orero (Eds.), Researching Audio Description: New Approaches (pp. 11–34). London: Palgrave.

Greening, J. (2011). Do we Need 3D Audio Description Guidelines: Recommendations from Focus Group Study Report. Retrieved from (February 20, 2019): http://audiodescription.co.uk/uploads/general/3D_film_and_tv_report_2.pdf

Jerald, J. (2015). The VR Book: Human-centered Design for Virtual Reality. Sant Rafael, CA: Morgan & Claypool. doi: 10.1145/2792790.2792827

Johansson, M. (2019). VR for Your Ears: Dynamic 3D Audio Is Coming Soon. Retrieved from IEEE Spetrum on February 19, 2019: https://spectrum.ieee.org/consumer-electronics/audiovideo/vr-for-your-ears-dynamic-3d-audio-is-coming-soon

Jones, S. (2017). Disrupting the Narrative: Immersive Journalism in Virtual Reality. Journal of Media Practice, 18(2–3), pp. 171–185. doi: 10.1080/14682753.2017.1374677

López, M., Kearney, G., & Hofstädter, K. (2016). Enhancing Audio Description: Sound Design, Spatialisation and Accessibility in Film and Television. Proceedings of the Institute of Acoustics, 38(2). Retrieved from (December 16, 2018): http://enhancingaudiodescription.com/wp-content/uploads/2016/11/RS2016-paper-Lopez-et-al.pdf

Maszerowska, A., Matamala, A., & Orero, P. (Eds.) (2014). Audio Description: New Perspectives Illustrated. Amsterdam: John Benjamins.

Nielsen, L. T., Møller, M. B., Hartmeyer, S. D., Ljung, T., Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Missing the Point: An Exploration of how to Guide Users' Attention during Cinematic Virtual Reality. Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology (pp. 229–232). New York: ACM. doi: 10.1145/2993369.2993405

Portillo, C. (2018). Films Unseen: Approaching Sound Design Alternatives to Enhance Perception and Mental Imagery Models among Audio Description Audiences on Sci-fi Films. Paper presented at Accessibility in Film, Television and Interactive Media Conference. Retrieved from (December 17, 2018): http://enhancingaudiodescription.com/conference/

Reviers, N., & Remael, A. (2015). Recreating Multimodal Cohesion in Audio Description: a Case Study of Audio Subtitling in Dutch Multilingual Films. New Voices in Translation Studies, 13(1), pp. 50–78. Retrieved from (February 20, 2019): https://www.iatis.org/images/stories/publications/new-voices/Issue13-2015/Articles/Reviers_New_voices_PUBL.pdf

Rothe, S., Hußmann, H., & Allary, M. (2017). Diegetic Cues for Guiding the Viewer in the Cinematic Virtual Reality. Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. New York: ACM. doi: 10.1145/3139131.3143421

Rothe, S., Tran, K., & Hußmann, H. (2018). Dynamic Subtitles in Cinematic Virtual Reality. Paper presented at ACM TVX 2018. Seoul. doi: https://dl.acm.org/citation.cfm?id=3213556

Sauro, J., & Lewis, J. (2016). Quantifying the User Experience: Practical Statistics for User Research. Amsterdam: Morgan Kaufmann.

Snyder, J. (2008). Audio description: The Visual Made Verbal. In J. Díaz-Cintas (Eds.), The Didactics of Audiovisual (pp. 191–198). Amsterdam/Philadelphia: John Benjamins. doi: 10.1075/btl.77.18sny

Suojanen, T., Koskinen, K., & Tuominen, T. (2015). User-centered translation. London: Routledge.

Syrett, H., Calvi, L., & van Gisbergen, M. (2017). The Oculus Rift Film Experience: a Case Study on Understanding Films in a Head-mounted Display. In R. Poppe, J.J. Meyer, R. Veltkamp, & M. Dastani (Eds.). Springer. doi: 10.1007/978-3-319-49616-0_19

Viswanathan, L. N., McDaniel, T., Krishna, S., & Panchanathan, S. (2010). Haptics in Audio Described Movies. IEEE International Symposium on Haptic Audio Visual Environments and Games, (pp. 1–2). Phoenix. doi: 10.1109/have.2010.5623958

Chapter 5. Article 4

Audio description in 360° content: Results from a reception study

5. Article 4

Abstract

The ImAc project was the first European initiative aiming to propose and test the model of implementing access services in 360° videos, paving the way for future studies in the under-researched field of immersive accessibility. This article reports on the methodology and results of a pilot study and a small-scale reception study, conducted in the last months of the project. The results show a favourable reception of extended audio descriptions by AD users. They also indicate interest in the implementation of spatial sound in AD provided for 360° content, which could be tested in future reception studies.

Keywords: Audiovisual Translation, Media Accessibility, audio description, 360° videos, reception study, presence, extended AD.

5.1. Introduction

The media landscape is being reshaped by new technologies and media formats, which are becoming more and more personalised and interactive (Allen and Tucker 2018). Likewise, new user-created content and interactive ways of storytelling, such as object-based media (Hanson 2019) are increasingly available. One emerging media format is 360° videos. They have become a new outlet for journalists, artists and videomakers, offering an interactive way of conveying the story to the viewer. They belong to Virtual Reality (VR), characterised by a high level of immersiveness (Slater and Usoh 1993: 221). Users can access 360° content, sometimes referred to as omnidirectional or spherical videos by means of headsets called head-mounted displays, as well as on smartphones equipped with gyroscopes, PCs, and tablets. 360° videos include both videos with six degrees of freedom and videos with three degrees of freedom (Fidyka and Matamala 2018), the latter being the focus of this article. 360° videos with three degrees of freedom allow users for certain, yet not complete, interactivity; when watching 360° content with three degrees of freedom, users stand in one physical position, triggering images with their head movements, but they are constrained to a fixed viewpoint. In other words, users remain at the centre of the action

and have an impression of being surrounded by the storyworld. As they are immersed in the story, the concept of presence, defined as the "perceptual illusion of non-mediation" (Lombard and Ditton 1997: 9) is essential in this media format and it serves as a quality metric employed to evaluate virtual environment content (Lessiter et al. 2001: 282).

In line with an effective legal framework (CRPD 2006) and European directives, namely the Audiovisual Media Services Directive (2010/13/EU) and the European Accessibility Act (EAA), audiovisual media products should be made accessible for all European citizens. In recent years, audio description (AD), a translation form that conveys the visual code of audiovisual productions in words (Braun 2008: 14), has been researched extensively in various European training institutions (Reviers 2016). Nevertheless, with the emergence of immersive content, the need to propose and test a model for implementing AD in 360° content has arisen. Such model should ensure that the viewing experience of AD users is more interactive than the experience of watching regular content on TV or in the cinema. In other words, as presence is crucial to a satisfactory user experience in this media format, AD should not only grant access to the visual content, but it should also engage its users on a more immersive level.

The first model of implementing AD in immersive environments was proposed within the EU-funded ImAc project. The project started in 2017, together with the early adoption of 360° videos by European broadcasters (EBU 2017: 9). As the project followed a user-centric approach, in its early stages a series of focus groups was organised in order to involve AD users and learn about their needs and preferences. The results of these qualitative studies, discussed in Fidyka and Matamala (2018), show the interest of the actual users in the integration of spatial sound, an audio technology already researched in the AD field (López, Kearney and Hofstädter 2016; Portillo 2018) and the elements of interaction in AD in this media format.

Based on the obtained feedback, a pilot reception study was designed and carried out in the next stages of the project, testing different presentation modes of spatial sound (see section 5.2). However, as the results from the pilot were inconclusive (see section 5.2.5), the methodology for the actual reception study was reconsidered, testing both a non-standard approach to AD scripting and extended audio description, which offers users a possibility of interaction (see section 5.3.2).

In order to evaluate the experience of participants when consuming audio described 360º content, presence measures were used in both studies, as they proved effective for assessing the experience of AD users (Fryer and Freeman 2012b; Walczak and Fryer 2017). What follows is a brief summary of the methodology and results of the pilot study (section 5.2) and the actual reception study (section 5.3). Finally, suggestions for further research are discussed in section 5.4.

5.2. Pilot study

This section discusses the methodology and results of the pilot study, conducted in Barcelona between 11–13 June 2019 in the form of individual testing. The aim of this study was to test the implementation of spatial sound in AD produced for 360° videos and to test the methodology with a reduced sample before the main study.

5.2.1. Participants

Six participants aged between 23 and 34 (2 blind, 4 partially-sighted) took part in the preliminary test. Only one participant was blind from birth (1) and other participants reported the beginning of their sight loss between the ages of 0-4 (1 participant), 5-12 (1 participant), and 13-20 (3 participants). All participants were frequent users of technological devices, such as smartphones, laptops and tablets, but only two participants reported watching 360º content occasionally on a smartphone and one participant (16.67%) by means of a head-mounted display. When asked about the reasons behind it, two participants pointed to the novelty of this media format (“I have not had the chance to use it”), two participants to the lack of access services (“It is not accessible”), and the remaining two to the lack of interest in immersive content (“I am not interested”). Similarly, none of the participants reported having a device on which to watch immersive content. All participants were familiar with AD and 50% of participants reported using this access service daily (two participants for 2-3 hours a day and one participant for less than 1 hour). Regarding accessing online content, three participants reported using screen readers, one participant using magnifiers, one participant both devices, and another one none of these tools.

5.2.2. Measures

Three online questionnaires were developed for this study in order to measure presence and preferences. These measures were chosen as presence and spatial realism are two important goals in the field of spatial audio research (Herre et al. 2015: 770). Presence was measured by means of the Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, Regenbrecht 2001), which includes 14 items on a 7-point scale. It consists of four components: (1) spatial presence – the sense of being physically present in a virtual environment, (2) involvement – attention devoted to the virtual environment, (3) experienced realism, defined as the subjective experience of realism in a virtual environment (Igroup n.d.; Regenbrecht and Schubert, 2002), and (4) a last component related to the general definition of the sense of presence: "I had a sense of being in the virtual environment" (Slater and Usoh 1993). IPQ was chosen as a measurement of user experience, as it has been used in previous studies on presence in virtual environments (Regenbrecht and Schubert 2002; Brown et al. 2003; Krijn et al. 2004; Hartanto et al. 2014; Kinateder et al. 2015), and it is recommended as a measure of presence because of its high reliability (Schwind et al. 2019).

The preference questionnaire, administered at the end of the study, included four questions. It asked participants to (1) rank the AD modes in order of preference, (2) explain the reasons behind their choice, and (3) suggest ways of improving AD. The last question provided space for additional comments. Ethical approval for the study was given by the Universitat Autònoma de Barcelona (UAB). Consent forms and coded questionnaires will be securely stored at the Universitat Autònoma de Barcelona for three years after the completion of the project.

5.2.3. Materials

Because of the novelty of this media format, one challenge related to this study was the limited availability of 360° videos that would meet testing requirements. Three initial episodes of the series "Holy Land" by Jaunt Ryot were chosen as a stimuli, as they were stand-alone narrative pieces, comparable in length. In this travel documentary, viewers are transported to various cultural sites in Israel, guided by the main narrator. Each episode chosen provided enough time to insert AD within the constraints of dialogue and made testing spatial sound possible, as the action develops at various angles of the 360° scene. A Catalan voiced-over version was created for the test, using a professional voice talent.

Another challenge related to the selection of stimuli was related to their length, as although there is no recommendation for an ideal stimuli length when measuring presence, a duration of 10–15 minutes is recommended in gaming context (ITU 2018:12). The 360° videos currently available on the market are shorter, those with a linear narrative oscillating around 5–15 minutes (Allen and Tucker 2018; Agulló 2019). Because of this, the three complete episodes were chosen, as clips should have a duration of typical 360° videos to reach ecological validity (Bryman 2008: 48).

For each episode, three AD modes were created (Classic, Static, Dynamic), referred to for the test purposes as AD-C, AD-S and AD-D.

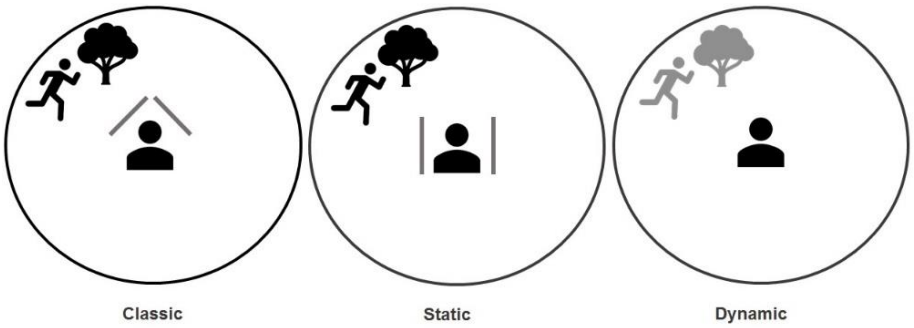


Figure 9. AD presentation modes used in the preliminary study

In the first presentation mode, the AD sound was placed above the user's head, while in the second presentation mode it was located on the user's side, as if someone was standing or sitting close to them, telling the story. In the Dynamic presentation mode, the AD sound was placed at different angles of the scene, depending on where the main action or other visual elements relevant to the plot were located. As the last presentation mode allows users to locate the events within the storyworld, our assumption was that it could guide viewers effectively within the storyworld and have a better viewing experience, which in turn would be reflected in higher presence scores.

The AD script for three episodes was originally written in English using the ImAc project editor, following existing AD guidelines (Ofcom 2000; Remael, Reviers and Vercauteren 2015), the reason being that the same test was conducted in the UK. This English AD was then translated into Catalan and voiced by a Catalan professional voice talent (female). An audio introduction was also created and voiced by a professional.

5.2.4. Procedure

The test was developed with a head-mounted display, and it was administered by the main researcher and a research assistant. Firstly, participants were welcomed, then presented with the ImAc project and the aim of the test. Secondly, they were assigned an individual participant's code and asked to sign informed consent forms. In the next step, participants replied to the questions in the demographic pre-questionnaire and listened to one general audio introduction (sound only). The actual AD test was comprised of: watching three video clips with randomised audio presentation modes, and replying to the IPQ questionnaire after each of them. At the end of the study, participants answered a preference questionnaire. All participants had the questionnaires read aloud and their responses were recorded by researchers on written online forms. For the purposes of this paper, participants' responses were translated into English.

5.2.5. Results

In the next sections, results from presence and preference questions will be presented, followed by a discussion.

5.2.5.1. Results from the IPQ questionnaire

Regarding presence, the median values for IPQ in the three conditions for each subscale are shown in Table 4 below:

	General presence	**Spatial presence**	**involvement**	**Experienced realism**
Classic	3.00	3.40	3.00	1.75
Dynamic	4.00	3.70	3.00	2.13
Static	4.00	2.70	4.00	2.50

Table 4. IPQ scores – preliminary test

Non-parametric Friedman tests reveal no statistically significant differences between the scores of any subscale between conditions: general presence (Chi-Square(2)=.200; N=6; p=.905); spatial presence (Chi-Square(2)=.087, N=6, p=.957); involvement (Chi-Square(2)=1.810, N=6, p= .405); experienced realism (Chi-Square(2)=1.00, N=6, p=.607).

5.2.5.2. Results from the preference questionnaire

Regarding the preference questionnaire, the results from questions 1 and 2, which asked participants about their preferred sound option and the reasons behind their choice, show that participants based their choice on script characteristics, rather than audio presentation. This is confirmed in comments such as the following: "It is hard for me to distinguish the three videos. [...] The criterion has been the videos that I have enjoyed the most" (participant 1), "There were details that captivated me more" (participant 2), "AD of places was better, and you could hear the noise of the environment, streets, music" (participant 5), "I liked it more because it had more details than other [videos]. I have noticed things that I would not have noticed otherwise" (participant 6). Similarly, one answer in additional comments (participant 5) pointed to the difficulties in differentiating between the three sound modes: "There are no differences in the three videos, between the types of sound. You only notice changes in the content of the audio description which could be improved by adding more details."

In questions 3 and 4, participants made several suggestions on how to improve AD. Firstly, the following comments suggest that two participants prefer more detailed descriptions to create a more complete mental image of the storyworld:

> "[...] perhaps by creating a more specific description of the places. I understand that there may be not enough time to describe more things. But everything is very general. I missed more elements, a more specific way of explaining [...]" (participant 5).

> "By adding more. It gave me the feeling that there was little description, and at certain times I did not know what was happening on the screen" (participant 1).

Secondly, related to the question of how AD could be improved to allow participants to be more immersed in the story, two comments pointed to the interest in listening to the original music of the video or background sounds: "I would also like to hear more music or more

ambient sound from the scene" (participant 2), "The ambient sound is very important" (participant 5).

Thirdly, some comments focused on the improvements which could be implemented to make 360° content more immersive. Although these comments are beyond the scope of this article, as they focus on storytelling techniques in 360° content, they can, however, serve as a recommendation for future content creators who wish to integrate access services already at the production stage. In this regard, participants suggested more slowly-paced content, with less sudden shifts of location, and more hearable ambient sounds.

5.2.6. Discussion

The results from the IPQ and preference questionnaires are inconclusive, as they demonstrate that participants were not able to clearly perceive the differences among the three audio treatments. Regarding the Dynamic presentation mode, none of the users noticed that the sound of AD was placed at different locations, depending on where the action took place. The reasons for this could be content-related. It is possible that participants could not perceive the differences in the AD correctly because the original videos were not recorded in spatial sound. It is also possible that the AD instances in these episodes were too short and the differences in audio would be more perceivable in content with longer pauses between the dialogues. In spite of the inconclusive results as regards preference for audio treatments, qualitative feedback on users' needs was gathered thanks to the adopted methodology. Based on participants' preferences, the AD presentation modes were reconsidered for the main reception study.

5.3. Main study

The main reception study followed the same methodology as the pilot test. To respond to users' preferences, a solution had to be found regarding the need of a more detailed AD. This posed a challenge; although the 360° storyworld is larger than standard content and it can contain more narratologically-relevant elements, AD is time-constrained, as it needs to fit in between the dialogues. Therefore, it was decided to test the Extended presentation mode which included additional descriptions, activated at the user's will (ISO/IEC 20071−11, WCAG 2017).

Secondly, it was decided to test an unconventional approach to AD scripting in order to see if it could have a positive impact on AD users' presence. Previous studies in the AD field have researched non-standard approaches to AD, including first-person AD (Fels et al. 2006), AD with elements of film language (Fryer and Freeman 2012a), or AD based on the production's screenplay (Szarkowska 2013; Walczak 2017). Previous reception studies have shown that unconventional AD scripting can increase a sense of immersion in the presented story for persons with sight loss compared to standard AD (Walczak and Fryer 2017). However, this question has not yet been tested in relation to more immersive content.

5.3.1. Participants

30 participants took part in the main reception study, with ages ranging from 22 to 78. Most participants had a university degree at an undergraduate (20) or postgraduate (7) level. 18 participants defined themselves as partially-sighted and 12 participants as blind. 10 participants taking part in the study reported onset of sight loss from birth. All participants reported using mobile phones on a daily basis, followed by television (20), laptop (14), PC (12) and tablet (9), which suggests that participants are frequent users of technological devices. As far as immersive technologies are concerned, only one participant reported using a head-mounted display on a daily basis, and most of the participants had never watched Virtual Reality content before. Similarly, only two participants reported having a device to access VR content (PS4 and PlayStation VR). The most frequent reason behind not watching such content was not having had the chance to use it (20), followed by a lack of accessibility (5) or lack of interest (2). Regarding the level of interest in immersive content, most participants were strongly interested (12) or interested (11), followed by a neutral attitude ("neither agree nor disagree" – 6), and only one participant chose the option "strongly disagree". All participants were frequent AD users. Regarding the usage of assistive technologies, most participants reported using screen readers (13), two participants reported using magnifiers, eight participants chose the option "both" and seven chose the option "none".

5.3.2. Materials

The same comparable clips were used as in the pilot test, with new presentation modes, referred to as Classic (AD-C), Radio (AD-R) and Extended (AD-E). The original English

script was written by RNIB (the Royal National Institute for the Blind) and the Catalan translation was prepared by the Universitat Autònoma de Barcelona. The text was rephrased and adapted, where necessary, to fit in the time slots between the dialogues. While the first presentation mode was the conventional AD, the second presentation mode followed an unconventional scripting style, which may be seen as a combination of the first-person narration introduced by Udo and Fels (2006) and audio drama (Fryer 2010). Audio description in Radio presentation mode was presented by a guide who accompanied the viewer, pointing to the most relevant visual elements of the storyworld. The viewer was addressed directly, in a conversational manner, and with a use of colloquialisms, nominal phrases, and discourse markers of casual speech (Table 5):

Classic	**Radio**
Una botiga ven articles de pell. En una altra, una dona mira collarets de granadura. [back translation] 'One shop sells leather goods. In another, a woman looks at beaded necklaces.' (ep. 1)	Practiqueu la cara de pòquer i el regateig abans d'entrar en aquest mercat. Prepareu-vos! [back translation] 'Practice your poker face and haggling skills before entering this market. Get ready!' (ep. 1)
Més pelegrins esperen en una llarga cua a l'entrada d'un santuari. [back translation] 'More pilgrims wait in a long line outside the entrance to a shrine.' (ep. 1)	Veniu d'hora si voleu calma i no haver de fer cua. L'any passat van venir aquí 4 milions de turistes, i cada cop en són més. [back translation] 'Get here early if you want to be calm and avoid the queues. Last year, 4 million tourists travelled here, and each year the number is going up.' (ep. 1)

Table 5. Classic and Radio scripting style

The rationale for choosing this scripting approach was two-fold. Firstly, while the first presentation mode describes the events, characters, and surroundings following the existing AD guidelines, the second presentation mode positions the viewer as an observer inside the presented world, as shown in Table 6 (discussed elements in bold):

Una dona de mitjana edat amb un mocador blanc al cap i una faldilla llarga i grisa contempla les pintures religioses de les parets del passadís del costat. La gent va amunt i avall. Alguns estan asseguts, esperant el seu torn. Una dona puja pels esglaons de la gruta, que són molt alts, mentre un home s'agenolla i pressiona el front contra l'esglaó superior. Un altre home es descalça en senyal de respecte. [back translation] A middle-aged woman in a white headscarf and long grey skirt gazes up at the religious paintings on the walls in the adjoining corridor. People go up and down. Some sit on chairs to await their turn. A woman climbs the steep steps out of the grotto, which are very steep, while a man kneels and presses his forehead to the top step. Another man takes off his shoes as a mark of respect. (ep. 3)	Per entrar a la gruta, us heu d'ajupir i passar per una petita porta amb un nom ben adequat: Porta de la Humilitat. A dins, està molt decorat: llums brillants, frescos i cortines de vellut. De seguida s'omple; haureu de fer cua. Fora, Betlem és una ciutat animada, però no heu d'anar gaire lluny per recordar la història de Maria i Josep. Si ets el típic turista, hi ha molt per fer. Hi ha un antic basar ple de vida. [back translation] To go into the grotto, you have to duck and go through a tiny door aptly called the Door of Humility. Inside, it's very ornate: sparkling lanterns, frescoes, velvet curtains. But it gets busy here so you have to wait your turn. Outside, Bethlehem is a pulsing city, but you don't have to go far to be reminded of the story of Mary and Joseph. Still, there is plenty to do if you're a regular tourist. There's a lively Old bazaar. (ep. 3)

Table 6. Classic and Radio scripting style

Secondly, this approach includes sentences that convey the atmosphere of the presented scenes. Examples of such sentences include the following:

> Com he dit: increïble! Jerusalem és nova i antiga: edificis moderns, carreteres transitades, palmeres tropicals. Tot això més enllà dels merlets de la ciutat vella i la dona que toca l'arpa a la porta de Jaffa. Una estranya harmonia.
>
> [back translation] Like I said: unreal! Jerusalem is both old and new: modern buildings, busy roads, tropical palm trees. All this beyond battlements of the old city and the woman who plays the harp at Jaffa Gate. **There's a strange harmony.** (ep. 2)
>
> El monestir de Sant Jordi és una meravella! Penja entre penya-segats al desert de Judea. Us semblarà un lloc perfecte per trobar una mica de pau.

> [back translation] The monastery of Saint George is a marvel! It is perched on the cliff in the Judean desert. **It will seem like a perfect place to find some peace.** (ep. 3)

Regarding the Extended presentation mode, it consists of the main script written in the style of Radio presentation mode and additional descriptions. These descriptions can be optionally activated by the user after hearing a special bell sound, informing about the availability of an additional commentary. This option was suggested in focus groups organised at the beginning of the ImAc project (Fidyka and Matamala 2018). After hearing such an audio cue, participants have a 5-second interval to play the AD by clicking on the controller of the head-mounted display. Upon activation, the main video is paused and the extended track is played until the end. In non-testing conditions, the activation of the extended description is optional, and it can also be activated by voice interaction with devices such as Amazon Echo Dot. However, for testing purposes, participants were asked to activate every additional track with a controller. Additionally, in non-testing conditions, the return to the main AD would be possible earlier, without having to listen to the extended track until the end.

The extended descriptions were inserted whenever a new landscape or an architectural object relevant to the plot were introduced in the episodes, allowing for a description of the visual elements that would not be described otherwise due to time constraints. For example, the first extended track in the first episode included the description of the presenter, Naomi Darg, who was also present in the remaining two episodes. Including such detailed description that allows persons with sight loss to visualise the character would not be possible without pausing the main narration:

> Hola, gràcies per treure el cap darrere l'escena! Aquí és on us donem tota la informació addicional. Comencem per la protagonista. Naomi Darg, en pantalla, presenta Holy Land. Diria que té entre 30 i 40 anys, és de complexió normal i té els cabells arrissats. Li arriben fins a les espatlles. Com que som a Jerusalem i visitem llocs sagrats, duu roba còmoda i més aviat discreta, de colors neutres, adequada per al clima càlid: vestits llargs, faldilles, samarretes de màniga llarga i mocadors de cotó. A la Cúpula de la Roca, porta un mocador fosc que li cobreix el cap.
>
> [back translation] Hi, thanks for joining me behind the scenes! This is where we give you all the extra bits of information. Let's start with the on-screen talent. On screen, Holy Land is led by Naomi Darg. I'd say, she is in her 30s, average build, with curly hair that

> falls to her shoulders. Since we're in Jerusalem and visiting some of the Holy sites, Naomi dresses modestly in loose fitting clothes in neutral shades, appropriate for the warm weather – long dresses, skirts, tops with long sleeves and cotton wraps. At the Dome of the Rock, Naomi wears a dark head scarf. (ep. 1)

Similarly to the style of the main AD, all extended descriptions were written in a chatty and engaging way. Apart from the description of the visuals, most of them also provided information about historical context or cultural titbits:

> El llançador de flors, de l'artista Banksy, és un grafit d'un home amb mocador i gorra de beisbol que llança un ram de flors. L'home i l'embolcall del ram són en blanc i negre; les flors i les tiges que sobresurten de l'embolcall són de color. El grafit d'aquest artista sigil·lós és només un dels molts que han convertit el mur de separació israelià en un ampli llenç. Per molts turistes, són l'atractiu principal de Betlem.
>
> [back translation] "Flower Thrower", by the artist Bansky is a graffiti of a man wearing a kerchief and baseball cap throwing a bouquet of flowers. The man and the flower wrapping are in black and white; the flowers and the stems protruding from the wrapper are in colour. **The graffiti by this stealthy artist is just one of the many graffities that have turned the Israeli Separation Wall into a vast canvas. For many tourists, this is the main attraction of Bethlehem.** (ep. 3)

The AD was recorded by a professional studio with a male voice to differentiate the audio description track from the main narration read by a female voice. While the script of the Classic presentation mode was read with neutral intonation, AD in the Radio presentation mode was read aloud in a livelier and more engaging way. Additionally, some AD instances were read with non-standard prosodic features. For example, in a scene happening in a mosque filled with tourists, AD was read in a whispering voice. Furthermore, as sound effects often interact with dialogues and music to create a more vivid mental imagery in radio drama (Fryer 2010), AD was complemented by ambient sound effects. For example, when the story was taking place on narrow, winding streets of Jerusalem, AD was reinforced by ambient sound reminiscent of street noises and when the story moved to a windy desert, AD users could hear the blowing of the wind, which established a sense of place.

5.3.3. Procedure

Similarly to the preliminary study, an audio introduction was presented to participants before exposing them to the three episodes. Apart from familiarising participants with 360° videos, and providing a broader context to the series, the introduction included information on how to activate the extended audio description. Our assumption was that the Radio and Extended presentation modes would have a direct effect on the reception of the videos, as they give persons with sight loss a chance of a more engaging experience.

5.3.4. Results

The following subsections discuss the results obtained in the main reception study, both from the IPQ questionnaire and additional preference questions.

5.3.4.1. Results from the IPQ questionnaire

A paired samples t-test was used in order to compare the scores on each subscale across the different types of AD. None of these comparisons gave a statistically significant difference (all $p>.05$).

	General presence	**Spatial presence**	**Involvement**	**Experienced realism**
Classic	4.73	4.62	4.98	3.40
Radio	4.70	4.55	5.04	3.54
Extended	4.73	4.64	4.86	3.63

Table 7. Median scores for all participants in each subscale across AD type

A one-way ANOVA comparing the results of the blind and partially-sighted users shows that there are significant differences between these type of users in their scores on IPQ. Planned comparisons show significant differences in General presence across all types of AD and that the scores are higher for partially-sighted users (see Table 5). In addition, for

Classic AD, scores for spatial presence are also significantly higher for partially-sighted users. There is also a trend to significance for Radio and Extended presentation mode (p=.087). Table 8 shows statistics and p value for each comparison. Highlighted cells mark significant differences:

		General presence	**Spatial presence**	**Involvement**	**Experienced realism**
Classic	Blind	3.50	3.5500	4.6667	2.7292
	Partially-sighted	5.56	5.3333	5.1806	3.8472
	p=	0.004	0.004	0.332	0.075
Radio	Blind	3.83	3.9167	4.8750	3.1458
	Partially-sighted	5.28	4.9778	5.1528	3.8056
	p=	0.040	0.087	0.640	0.280
Extended	Blind	3.83	4.0333	4.5208	3.3750
	Partially-sighted	5.33	5.0444	5.0833	3.8056
	p=	0.036	0.087	0.313	0.529

Table 8. IPQ results from persons with sight loss

A one-way ANOVA comparing the results of the blind and partially-sighted users shows that there are significant differences in general presence across all types of AD and that the scores are higher for partially-sighted users. In addition, for Classic AD, scores for spatial presence are also significantly higher for partially-sighted users. There is also a trend to significance for Radio and Extended AD (p=.087).

5.3.4.2. Results from the preference questionnaire

In the preference questionnaire, 12 participants indicated Extended presentation mode as their preferred option, 10 participants selected Classic presentation mode and the 8 remaining participants stated a preference for the clips with the Radio presentation mode. Two participants commented in open questions that they appreciated every description.

	AD-C (Classic)	**AD-R (Radio)**	**AD-E (Extended)**
1 preferred mode	10 (33.33%)	8 (26.67%)	12 (40%)
2 preferred mode	10 (33.33%)	10 (33.33%)	10 (33.33%)
3 preferred mode	10 (33.33%)	12 (40%)	8 (26.67%)

Table 9. Results on preferences in the main study

Upon analysing data for the blind and partially-sighted participants separately, the results are following:

	AD-C (Classic)	**AD-R (Radio)**	**AD-E (Extended)**
Blind persons			
1 preferred mode	1 (3.33%)	2 (6.67%)	9 (30%)
2 preferred mode	4 (13.33%)	6 (20%)	2 (6.67%)
3 preferred mode	7 (23.33%)	4 (13.33%)	1 (3.33%)
Persons with partial sight loss			
1 preferred mode	9 (30%)	6 (20%)	3 (10%)
2 preferred mode	6 (20%)	4 (13.33%)	8 (26.67%)
3 preferred mode	3 (10%)	8 (26.67%)	7 (23.33%)

Table 10. Results on preferences in the main study

The most frequent comments from participants who selected Classic presentation mode in the first place suggest that this scripting style allowed them to create a more complete mental representation of the storyworld (1), had a more appropriate level of detail than another scripting style (3), and was more coherent with the style of the main narration.

The most frequent reason behind choosing the Radio presentation mode as the preferred option was the integration of ambient sounds (2), which made the experience more realistic. One comment that focuses on the relation between ambient sounds and presence seems particularly relevant:

> I would like for the ambient sounds and the sound effects to be more in the first line than the explanations. I would like to have audio description more in the background and the sound effects more underlined to feel that the presented world is real. The explanations were fast with strong intonation and this way of explaining is a bit tense. I would like it to be more natural, relaxed.

Similarly, another participant who chose Radio presentation mode as the preferred option commented that this type of scripting, which combines verbal description of visuals with evocative sound effects, seems more appropriate for this innovative media format and added that Classic presentation mode does not make the viewing experience different from watching standard films. Another participant commented positively on the engaging use of language in this presentation mode. Participants who least preferred this presentation mode pointed to the following reasons behind their choice: this presentation mode resembles a separate narration rather than AD (2), the level of detail is not sufficient.

The results obtained in the first and second questions also show a favourable reception of the Extended presentation mode. Out of 12 blind participants who took part in the study, 9 blind participants selected Extended descriptions, including 2 congenitally blind participants. What participants appreciated the most was the possibility to listen more at will (9), the possibility to interact with the content (1), being provided with more details (5), and feeling more immersed (2):

> I was feeling more present in the virtual world in AD-E, as it had more information and a better quality of information that made me less conscious of the real world.

Two participants placed Extended presentation mode as their third choice. The reasons for this can be related to the testing conditions, as participants were asked to (1) activate all extended tracks and (2) listen to them until the end.

In the third question, participants proposed some improvements that could enable them to have a more immersive viewing experience when watching content in this media format. Chief among these are the possibility of being guided towards the described places,

characters and objects (5 participants), and being able to listen to extended descriptions (7 participants). Interestingly, one participant commented that she/he would prefer to have extended descriptions in all three episodes watched during the study. Other suggestions include watching content with spatial sound (2 blind participants), integration of ambient sounds (2) and having the possibility to adjust the sound of AD independently from the video (1).

Regarding the improvements that could be made in the Extended presentation mode, the following comment suggests integrating music or other sound effects at the beginning and at the end of each extended description:

> The background of the extended audio description should convey the same background as the narration so that there is no interruption [upon activating extended AD]. I would like extended AD to be more integrated in the video. The interruption is very noticeable, extended AD starts abruptly. I would like extended AD to start with background music, then to have a description, and to finish with the same background music.

Among other improvements, two participants would prefer a more distinctive sound signalling of extended descriptions. Their comments suggest that the sound should be changed or first introduced in audio introduction in order to familiarise users with it. Additionally, one participant suggested that these additional tracks should contain only audio description and not titbits. This participant further specified that such information should be provided only in the video.

5.3.5. Discussion

The three presentation modes yielded similar levels of presence for all participants, which suggests that none of them can lead to a significantly higher immersion in the story, but it also shows that none of them expelled participants from their viewing experience. The differences in the presence scores between blind and partially-sighted persons obtained in two presence subscales suggest that blind persons need additional solutions in order to feel more present in the 360° storyworld. As two blind persons suggested the possibility of spatial sound to feel more immersed, this sound technology could be further tested.

The qualitative results from the preference questionnaire suggest the positive response of participants towards all presentation modes. Additionally, although it cannot be concluded

from this study that combining verbal description with ambient sounds may stimulate presence, additional feedback suggests that persons with sight loss may appreciate this solution. As for Extended descriptions, one participant put forward an interesting comment in the second question:

> AD-E gives a little more details, it would be interesting for me to have the possibility to activate more extended descriptions of different points of the video. AD-R is like listening to an audiobook, like any documentary that I can watch on TV. At the AD level, it is poorer. The second (AD-C) gives more visual information and its details allow me to imagine (the storyworld).

In this test, although Extended descriptions were not linked to a given point within the video, it would be interesting to test this option with users in future studies. In such a case, extended descriptions would not be linear descriptions activated by a click, but instead participants would be able to stop the video and trigger different descriptions by head-movements.

5.4. Conclusions

The aim of this article was to discuss the first exploratory approach for implementing AD in 360° videos. It follows previous studies in the AVT field which measured the impact AD may have on the quality of end-user experience in 2D content (Fryer and Freeman 2012b, 2014; Wilken and Kruger 2016; Wissmath and Weibel 2012; Walczak and Fryer 2017), but it takes the previous research further, testing presence levels in a media format in which users have a degree of control over their experience.

The results of the main reception study show the positive response of participants toward extended audio description. They also show some possible improvements that could be made in order to better respond to their needs. There are several advantages related to integrating such solution in this media format. Firstly, thanks to the possibility of pausing the main narration, users can be provided with more visual information than is present in the 360° storyworld and that could not fit in between the dialogues, or with additional titbits that are given in standard content in audio introductions. It also provides persons with sight loss with a possibility of interaction. This solution can also prove useful for describers, who

deemed content selection in this media format challenging in the focus groups carried out within the project.

This is a small-scale, exploratory study and its limitations must be acknowledged. First and foremost, one limitation is related to the reduced sample. To obtain more reliable results, future studies should be conducted with more users. However, as stated by Orero et al. (2018), smaller sample sizes are acceptable when conducting research with persons with sight loss. Secondly, as only self-report measures were used in this study, data could be triangulated by using objective measures. Also, the present pilot study testing the implementation of spatial sound could be replicated with other video clips that contain more pauses between the dialogues and whose original sound is recorded in spatial audio technology. Additionally, more attention should be given to the presence measures in future studies. Existing self-report presence measures are designed for sighted persons, and they include statements such as "I felt like I was just perceiving pictures" (IPQ), which may not be appropriate for persons with sight loss. This is why validation of presence measures with persons with sight loss would be needed to ensure the validity of results when conducting research.

Several possible research avenues emerge from this study for future research into AD in 360° videos, or other types of immersive environments, that will increasingly emerge on the market in the upcoming years and reshape the landscape of Audiovisual Translation. Firstly, taking into account the interest of users, the technology of spatial sound in AD in 360° content could be further explored, including the object-based sound (Simon, Torcoli and Paulus 2019), which allows sound technicians to place the sound exactly on the object. Secondly, the research on spatial sound could focus not only on AD, but also on audio subtitles (AST), as it could enable users with sight loss to locate the characters in 360° scene.

This article aims to contribute to an emerging line of research in AVT, and it is hoped that the preliminary results it presents will contribute towards a better understanding of AD users' needs in 360° content to ensure not only access to this type of content, but also a more captivating viewing experience.

Acknowledgements

We owe our warmest thanks to B1B2B3 Association for their support in finding volunteers for this study and to all the participants who took part in our experiment. We are also grateful to Ryot and Jaunt VR for letting us use “Holy Land” for our study.

ImAc has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the grant agreement 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). Blanca Arias-Badia’s participation in this research has been financially supported by the Spanish Research Agency, in the framework of the postdoctoral scholarship programme Juan de la Cierva Formación (FJCI-2017-32064). This article is part of Anita Fidyka’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, Belén. 2019. Nuevos retos para la TAV: subtitulación en medios inmersivos. Resultados de un análisis de corpus. Alicante: AIETI, 23–25 January 2019.

ALLEN, Catherine, TUCKER, Dan. 2018. Immersive content formats for future audiences [online]. Available at: https://assets.ctfassets.net/nubxhjiwc091/5WypbBxFEQqS8I6WaKUQgg/ca25eb800e9f88a740079dbca623ca2c/Immersive_Content_Formats_for_Future_Audiences.pdf

BRAUN, Sabine. 2008. Audio description research: State of the art and beyond. In Translation Studies in the New Millenium, vol. 6, pp. 14–30.

BRYMAN, Alan. 2008. Social Research Methods. Oxford, New York: Oxford University Press.

BROWN, Sarah, LADEIRA, Ilda, WINTERBOTTON, Cara et al. 2003. The effects of mediation in a storytelling virtual environment. In Lecture Notes in Computer Science, vol. 2879, pp. 102–111. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.7163&rep=rep1&type=pdf

EBU (EUROPEAN BROADCASTING UNION). 2017. Virtual Reality: How are public broadcasters using it? Report [online]. Available at: https://www.ebu.ch/publications/strategic/login_only/report/virtual-reality-how-are-public-broadcasters-using-it

EC (EUROPEAN COMMISSION). 2010. Audiovisual Media Services Directive (2010/13/EU). Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0013

EC (EUROPEAN COMMISSION). 2019. The European Accessibility Act (ACC). Available at: https://ec.europa.eu/social/main.jsp?catId=1202

FELS, Deborah, UDO, J. P., TING, P., Diamond, J. I. 2006. Odd Job Jack described - a first person narrative aproach to described video. In Journal of Universal Access in the Information Society, vol. 5, no. 3, pp. 73–88.

FIDYKA, Anita, MATAMALA, Anna. 2018. Audio description in 360° videos: Results from focus groups in Barcelona and Kraków. In Translation Spaces, vol 7, no. 2, pp. 285–303. doi: 10.1075/ts.18018.fid

FRYER, Louise. 2010. Audio description as audio drama – a practitioner's point of view. In Perspectives: Studies in Translatology, vol. 18, no. 3, pp. 205–213.

FRYER, Louise, FREEMAN, Jonathan. 2012a. Cinematic language and the description of film. Keeping AD users in the frame. In Perspectives: Studies in Translatology, vol. 21, no. 3, pp. 412–426.

FRYER, Louise, FREEMAN, Jonathan. 2012b. Presence in those with and without sight: implications for virtual reality and audio description. In Journal of Cybertherapy and Rehabilitation, vol. 5, no. 1, pp. 15–23.

FRYER, Louise, FREEMAN, Jonathan. 2014. Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In FELNHOFER, A. and KOTHGASSNER, O. D. (eds.), Challenging presence: Proceedings of the 15th International Conference on Presence, Wien: Facultas.wuv, pp. 99–107.

HANSON, Nick. 2019. Storytelling of the future. Available at: https://www.bbc.co.uk/rd/blog/2019-02-storytelling-interactive-digital-drama

HARTANTO, Dwi, KAMPMANN, Isabel. L., MORINA et al. 2014. Controlling social stress in Virtual Reality Environments. In PLoS ONE [online], vol. 9, no. 3. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0092804

HERRE, Jürgen, HILPERT, Johannes, KUNTZ, Achim et al. 2015. MPEG-H 3D audio—The new standard for coding of immersive spatial audio, In IEEE Journal Of Selected Topics In Signal Processing, vol. 9, no. 5, pp. 770–779. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7056445

IGROUP. n.d. Igroup presence questionnaire (IPQ) scale construction. Available at: http://www.igroup.org/pq/ipq/construction.php

ISO (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION). Guidance on audiodescriptions [online]. Available at: https://www.iso.org/obp/ui/#iso:std:iso-iec:ts:20071:-21:ed-1:v1:en

ITU (INTERNATIONAL TELECOMMUNICATION UNION). 2018. Subjective evaluation methods for gaming quality (ITU-T P.809). Available at: https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13626&lang=en

KINATEDER, Max, GROMER, Daniel, GAST, Philipp et al. 2015. The effect of dangerous goods transporters on hazard perception and evacuation behavior – a virtual reality experiment on tunnel emergencies. In Fire Safety Journal, vol. 78, pp. 24–30.

LESSITER, Jane, FREEMAN, Jonathan, KEOGH, Edmund, DAVIDOFF, Jules. 2001. A cross-media presence questionnaire: the ITC sense of presence inventory. In Presence: Teleoperators and Virtual Environments, vol. 10, no. 3, pp. 282–297.

LOMBARD, Matthew, DITTON, Theresa. 1997. At the heart of it all: the concept of presence. In Journal of Computer Mediated Communication, vol. 3, no. 2, pp. 1–39.

LÓPEZ, Mariana, KEARNEY, Gavin, HOFSTÄDTER, Krisztián. 2016. Enhancing audio description: sound design, spatialisation and accessibility in film and television. Southhampton: Reproduced Sound Conference, 15–17 November 2016.

OFCOM (OFFICE OF COMMUNICATIONS). 2000. ITC Guidance on standards for audio description [online]. Available at: http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf

ORERO, Pilar, DOHERTY, Stephen, KRUGER, Louis-Jan et al. 2018. Conducting experimental research in audiovisual translation (AVT). A position paper. In JoSTrans, no. 30, pp. 105–126.

PORTILLO, Cesar. Films unseen. 2018. York: Accessibility in Film, Television and Interactive Media Conference, 14–15 October 2018.

REGENBRECHT, Holger, SCHUBERT, Thomas. 2002. Real and illusory interaction enhance presence in virtual environments. In Presence: Teleoperators and Virtual Environments, vol. 11, no. 4, pp. 425–434.

REMAEL, Aline, REVIERS, Nina, VERCAUTEREN, Gert. 2015. Pictures painted in words. ADLAB audio description guidelines [online]. Trieste: Edizioni Università di Trieste. Available at: http://www.adlabproject.eu/Docs/adlab%20book/index.html

REVIERS, Nina. 2016. Audio description services in Europe: an update. In The journal of Specialized Translation, no. 26, pp. 232–247. Available at: https://www.jostrans.org/issue26/art_reviers.pdf

SCHUBERT, Thomas, FRIEDMANN, Frank, REGENBRECHT, Holger. 2001. The experience of presence: Factor analytic insights. In Presence: Teleoperators and Virtual Environments, vol. 10, no. 3, pp. 266–281.

SCHWIND, Valentin, KNIERIM, Pascal, HAAS, Nico, HENZE, Niels. 2019. Using presence questionnaires in Virtual Reality. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, New York: ACM. Available at: https://vali.de/wp-content/uploads/2019/01/2019-CHI-QuestionnairesInVR.pdf

SIMON, Christian, TORCOLI, Matteo, PAULUS, Jouni. 2019. MPEG-H audio for improving accessibility in broadcasting and streaming. Available at: https://arxiv.org/abs/1909.11549

SLATER, Mel, USOH, Martin. 1993. Representation systems, perceptual position, and presence in immersive virtual environments. In Presence, vol. 2, no. 3, pp. 221–233.

SZARKOWSKA, Agnieszka. 2013. Auteur description: from the director's creative vision to audio description. In Journal of Visual Impairment & Blindness, vol. 107, no. 5, pp. 383–387.

UDO, John Patrick, FELS, Deborah. 2009. Suit the action to the word, the word to the action. An unconventional approach to describing Shakespeare's Hamlet. In Journal of Visual Impairment and Blindness, vol. 103, no. 3, p. 178–183.

UDO, John Patrick, FELS, Deborah. 2010. Horatio audio-describes Shakespeare's Hamlet. Blind and low-vision theatre-goers evaluate an unconventional audio description strategy. In British Journal of Visual Impairment, vol. 28, no. 2, pp. 139–156.

UN (UNITED NATIONS). 2006. Convention on the Rights of Persons with Disabilities (CRPD). Available at: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html.

WALCZAK, Agnieszka. 2017. Creative description: audio describing artistic films for individuals with visual impairments. In Journal of Visual Impairment & Blindness, vol. 111, no. 4, pp. 387–391.

WALCZAK, Agnieszka, FRYER, Louise. 2017. Creative description: The impact of audio description style on presence in visually impaired audiences. In British Journal of Visual Impairment, vol. 35, no. 1, pp. 6–17.

WCAG (WEB CONTENT ACCESSIBILITY GUIDELINES). Extended audio description (prerecorded) [online]. Available at: https://www.w3.org/WAI/WCAG21/Understanding/extended-audio-description-prerecorded.html

WILKEN, Nicole, KRUGER, Jan-Louis. 2016. Putting the audience in the picture: Mise-en-shot and psychological immersion in audio described film. In Across Languages and Cultures, vol. 17, no. 2, pp. 251–270.

WISSMATH, Bartholomäus, WEIBEL, David. 2012. Translating movies and the sensation of 'being there'. In PEREGO, E. (ed.) Eye Tracking in Audiovisual Translation, Rome: Aracne Editrice, pp. 281–298.

Chapter 6. Summary

6. Summary

This thesis has explored the possible approaches to implementing AD in the content of 360° videos with three degrees of freedom (3DOF), one of the novel media formats to have entered the market in recent years. The studies presented within this thesis, conducted at an early stage of adoption of this technology, have obtained feedback from two groups of users, so as to cater for their needs. On the one hand, they centre on identifying the needs of professional audio describers when describing 360° videos. On the other, their focus is on AD users in order to develop solutions that enhance their viewing experience. Before the experimental studies for this thesis could be undertaken, allowing for the fulfilment of the main objectives, four specific aims had to be met. The first specific aim was to define 360° videos within immersive virtual environments. To meet this aim, a literature review was conducted as the first step in the thesis' development. In the course of the literature review, 360° videos were defined as a form of VR environments, which are considered to have a high degree of immersive capacity. It was further established that this media format provides its users with a degree of interaction. To meet the second specific aim, 2D video content had to be compared with 360° narrative videos. Again, the results were obtained in a comprehensive literature review. The results showed that although 360° narrative videos bear similarities with 2D video content as regards narrative construction, the differences reside in the lack of framing, a different use of guiding cues or scene transition techniques. During this analysis, the main challenges related to describing 360° narrative videos were also identified.

The results from the literature reviews were complemented by the need analysis, conducted in the form of focus groups with AD users and professional audio describers. The results of these qualitative studies suggested possible approaches to AD in 360° content, including the use of spatial sound and the elements of interaction, which contributed towards defining the presentation modes for the experimental testing. Regarding the feedback obtained from professional describers, the results suggested the need of additional guidance as far as content selection is concerned, confirming the results of the literature review.

With regard to the AD users, three presentation modes, differing in terms of the placement of spatial sound, were evaluated in an exploratory pilot study in order to determine whether they could be further tested with a larger sample. Their effect on users' presence was

assessed by means of an IPQ questionnaire. Additional questions gathered users' preferences. The results of the IPQ show that the effect of the presentation modes was not substantial for any of the presence subscales. These results were confirmed in the preference questions, which did not show significant differences between the three conditions. They also showed that preferential choice was made based on script characteristics, not on sound placement. The main study, conducted with a larger sample followed the same methodological design as the pilot study, except for the presentation modes, which were reconsidered due to the research limitations. Based on users' feedback obtained in the pilot study, the presentation modes differed in terms of scripting.

The quantitative results obtained by means of the IPQ questionnaire show significant differences between blind and partially-sighted persons in the 'general presence' subscale, the scores being higher for partially-sighted persons. The results of comparison between these two groups also show that statistically significant differences were found for the Classic presentation mode on the 'spatial presence' subscale (p=0.004). A trend for significance was also observed for other presentation modes. Upon analysing data for all participants, however, significance was not reached on any of the subscales. The qualitative data from the preference questionnaire showed that approximately 33% of participants selected the Classic presentation mode as their preferred option, more than 26% of participants opted for the Radio presentation mode and 40% of participants chose the Extended presentation mode. Out of 12 blind participants who took part in the study, 9 blind participants selected the Extended presentation mode. Additional feedback provided by participants also showed a favourable reception of this presentation mode.

In relation to the professional describers, a prototype of the online tool for editing audio description was assessed by means of an online questionnaire consisting of two parts: a sub-questionnaire measuring the usability of the tool and open questions for gathering users' preferences. The numerical results on usability, obtained by means of the SUS questionnaire, showed that the prototype of the tool obtained the score of 55.9, which, as expected, suggested that the tool requires further improvements before meeting the standard quality as demanded by the market. In the preference questions, participants provided specific points for improvement, which included, among others: including a waveform, customizing the shortcuts, previewing the final AD product by means of speech synthesis or previewing it by means of a HMD.

Chapter 7. Conclusions

7. Conclusions

Given the fact that the role immersive environments play in our lives is likely to increase in the coming years due to investment in technologies, it is important that potential approaches to implementing access services within immersive content be defined and tested. It is within this context that this thesis has investigated the implementation of AD in the content of 360° videos with three degrees of freedom (3DOF), a novel media format that allows users to have a degree of control over their experience.

With this in mind, two main objectives were set for this thesis. The first main objective was linked to the study with AD users and focused on defining and evaluating different AD presentation modes in 360° videos. The second main objective relating to professional describers was centred on identifying their needs when describing 360° videos.

This chapter presents and discusses the most relevant results of the experimental studies. It likewise analyses whether the main objectives set for this thesis were met, while also validating or refuting the hypotheses. The overall evaluation is presented in the next two sections, differentiating between the studies with AD users (section 7.1) and the studies with professional audio describers (section 7.2). Finally, it includes sections on the contributions of this research (section 7.3), its limitations (section 7.4) and pathways for future research (section 7.5).

7.1. Studies with AD users

The first main objective of this thesis was to define and evaluate different AD presentation modes in 360° videos with 3DOF. In order to attain this objective, four specific aims were set. As the first step in this thesis development, 360° videos had to be defined within immersive virtual environments. To meet this specific aim, a literature review was conducted. As discussed in Chapter 2, this research established the potential and the limitations of this media format as far as its immersive capacity and user interaction are concerned. During the literature review, 360° videos with 3DOF were defined as a form of VR environments. Such environments are considered to have a high degree of immersive capacity, as its users wear head-mounted displays. Such devices provide them only with computer-generated sensory input, and input from the real world is limited. Regarding user

interaction, 360° videos with 3DOF provide users with a degree of interactivity, as they can change their viewing direction within the 360° storyworld.

As the second step, 2D video content was compared with 360° narrative videos, meeting the second specific aim. Similarly to the previous specific aim, a comprehensive literature review was conducted. The results of this preliminary comparison show that although the users of 360° narrative videos follow a linear story, pre-constructed by a filmmaker in a similar way to standard films, there are some differences in narrative construction, such as lack of framing, the different use of guiding cues or scene transition techniques. From the results obtained in the two literature reviews, it became evident that, regarding AD users with sight loss, one of the challenges will be related to guiding this group towards audio described places, characters and actions within the storyworld.

The third specific aim was linked to obtaining feedback from AD target users in the early stages of the development of this thesis. A need analysis was conducted in a form of two focus groups, involving two groups of users. The results obtained from AD users suggested possible approaches to AD in 360° videos, including the use of spatial sound and elements of interaction (e.g. the option to pause the video to listen to additional AD tracks triggered by head movements). The results obtained in this qualitative study contributed towards defining the presentation modes for the experimental part of this thesis.

With regard to the experimental studies, the specific aim of comparing the levels of presence and preferences of AD users when exposed to different presentation modes was set. Linked to this aim, two hypotheses were put forward, the first one being linked to the pilot study and the second one to the main study with a larger sample:

H1: Users will report higher presence levels when exposed to Static and Dynamic sound presentation mode, compared to Classic mode. Static and Dynamic will be the preferred options.

H2: Users will report higher presence levels when exposed to Radio and Extended presentation mode, compared to Classic mode. Radio and Extended will be the preferred options.

To allow for the comparison, a methodological framework had to be established for both studies. Methodology for the evaluation of the presentation modes followed a mixed-

method approach and it was based on the subjective measures. As measures, presence and preferences were selected. The Igroup Presence Questionnaire (IPQ) was chosen as a scale for measuring the sense of presence experienced by the participants in VR. A separate questionnaire was designed to gather users' preferences. As for the participants, the sampling method and the inclusion criteria for the sampling population were identical in both studies, the only difference being sample size.

The aim of conducting the pilot study was to explore the implementation of spatial sound in the AD produced for the 360° videos with 3DOF and to gather preliminary users' feedback. It also aimed at testing the methodology before conducting the study with a larger sample. As for the results from the IPQ questionnaire, they indicate that the effect of none of the presentation modes was significant for presence. In all subscales, the three conditions returned similar median scores.

To complement the IPQ results, participants were explicitly asked about their preferred sound option. Similarly, the data obtained did not show a preference of the participants towards any of the presentation modes. It pointed, however, to the difficulties related to distinguishing between the three presentation modes and to the focus on script characteristics. The reasons behind it may be two-fold. Firstly, they may be related to the technology of spatial sound employed. Secondly, they may be due to the characteristics of the clips. The three videos used for the study, although comparable in terms of length and content, were not recorded in spatial sound. This sound technology was inserted only in audio descriptions provided between the dialogues. Furthermore, the duration of the audio descriptions could have impacted the results. As mentioned above, all the videos provided enough spaces between the dialogues for inserting AD. The content was, however, fast-paced, with rapid changes of locations. It is possible that a more-slowly paced content, with longer pauses between the dialogues, would allow participants to have more time to listen to audio descriptions, which would in turn allow them to perceive the differences in sound placement more clearly. Although the results obtained in this study did not lead to the validation of Hypothesis 1, the study allowed for testing the methodology before the main study.

Due to the limited availability of the clips that would meet test requirements and at the same time be recorded in a technology of spatial sound, the presentation modes for the study with a larger sample were reconsidered. The presentation modes tested in the main

study differed in terms of scripting, a subject studied in several previous studies in the AD field in relation to presence. Additionally, taking into account the results obtained during the literature review as well as the focus groups on the challenge related to the prioritization of information when describing the 360° storyworld (see Chapter 3), a decision was taken to research the reception of the Extended presentation mode, which included extended descriptions.

Two statistical analyses were carried out for the results obtained during the main study. In the first one, data from all participants, both the blind and partially-sighted, was analysed together. The second analysis separated the results for the blind and partially-sighted persons, which allowed for comparing the results from the two groups. The medians are given in Tables 7 and 8 in Chapter 5.

The analysis of the results obtained from all the participants did not reveal statistical differences between the presentation modes in all IPQ subscales. However, upon analyzing data from blind and partially-sighted persons separately, such differences were found in the 'general presence' subscale (p=0.004), the scores being higher for persons with partial sight loss. The results of a comparison between the two groups also shows statistically significant differences between the two groups on the 'spatial presence' subscale for the Classic presentation mode. A trend of significance was also observed for other presentation modes on the same subscale. The differences between the groups did not reach significance for either involvement or for realness. Thus, the results obtained in the IPQ questionnaire suggest that blind persons would need additional solutions in order to have a satisfactory viewing experience.

The qualitative data gathered in the preference questionnaire shows that 10 participants (33.33%) opted for the Classic mode, 8 participants (26.67%) chose the Radio mode and 12 participants (40%) selected the Extended presentation mode. Out of 12 blind participants who took part in the study, 9 blind participants selected the Extended mode, including 2 congenitally blind participants. The results show that the reasons behind choosing this presentation mode are linked to: having the possibility of listening to additional information at will or having the possibility of interacting with the content, being provided with more details and feeling immersed in the storyworld. Regarding the reasons behind choosing the remaining two options, the results show that participants selected the Classic presentation mode because of the adequate level of detail contained in this mode. The most frequent

reasons given behind choosing the Radio presentation mode were: the use of ambient sounds and engaging use of language.

When asked about possible improvements that could enable them to have a more immersive experience when watching content in this media format, 7 participants pointed to being able to listen to the extended descriptions. The results also show that the participants would like to be guided towards the described actions, places and characters within the storyworld. Other possible improvements include the integration of spatial sound, ambient sounds and having the possibility to adjust the sound of the AD independently from the video. The results of the main reception study also show some specific points for future improvement of the extended descriptions, such as integrating music or other sound effects at their beginning and end to make the transition between these tracks and the main video smooth, or linking extended descriptions to different points within the video. Thanks to this option, users could activate those tracks, linked to places, characters and objects upon stopping the video and listen to them by head-movements.

Besides gathering data on the evaluation of the presentation modes, this study contributed through more general feedback regarding the usage of immersive technologies and users' attitudes towards them. Such data was gathered by means of closed and open questions. As far as the usage of immersive technologies is concerned, the results show that most of the participants had never watched a VR content before taking part in the study. Similarly, only one participant reported using an HMD on a daily basis. When asked about the reasons behind not watching immersive content, the comments of twenty of the participants pointed towards the novelty of this technology. The second most frequent reason was the lack of accessibility of such content. Regarding the level of interest in this technology, the results show that most of the participants were strongly interested or interested (23), followed by a neutral attitude (6).

Comparing the presentation modes allowed us to gather feedback on the reception of the approaches to AD in 360° videos, which were defined at earlier stages of the thesis' development. Although it was initially hypothesized that the innovative presentation modes would yield higher presence levels, no evidence supported this expectation in Hypothesis 1, which leads to its refutation. Regarding Hypothesis 2, the results on the Extended mode obtained in the preference questionnaire lead to its partial validation.

All in all, the reception studies compared presence levels and preferences of participants with sight loss when exposed to the different presentation modes, which were defined using the results of the focus groups conducted in the early stages of this thesis development. Therefore, the last specific aim set for this user group was met, allowing for the fulfilment of the first main objective. The experimental studies also contributed with relevant comments on how AD could be improved in this media format to provide a better viewing experience for users, paving a way for further analyses.

7.2. Studies with professional describers

The second main objective of this thesis was to identify the needs of professional audio describers when describing 360° videos. Again, in order to attain this objective, two specific aims were set.

The first specific aim, set at an early stage of this thesis development, was to obtain qualitative feedback from professional audio describers about their needs regarding the production of AD in 360° videos. As discussed in the previous section, two focus groups were organized to meet this aim. Regarding professional describers, the results confirmed the findings of the literature review on storytelling and indicated that the content selection in this media format raises concerns for audio describers. Cooperation with content creators was also suggested as a possible solution to this challenge.

The second specific aim was related to testing a prototype version of an AD editor for describing videos with professional describers. In order to gather users' feedback, a methodological framework had to be established.

A methodology for the evaluation of the editing tool was developed based on subjective measures, combining usability with preferences. As a measure of usability, a SUS questionnaire was selected. To obtain users' preferences, an additional sub-questionnaire consisting of nine questions was designed. The analysis of data collected on usability shows that the prototype of the editor obtained a score of 55.9. As expected, this result shows that the prototype needed further improvements before meeting the demands of the market, as the score of 68 is considered an average.

Apart from numerical data, a preference questionnaire allowed to obtain qualitative data. The specific proposals made by describers taking part in the test allowed for defining the points for further improvement, such as including the waveform, reading AD for preview via speech synthesis, customizing the shortcuts or previewing the AD by means of an HMD. To compare, the second score, obtained in the next usability study which does not form part of this thesis (Annex 4.3) equals to 60.52, corresponding to a letter grade D and the percentile rank 15–34%. This result shows an improvement made in the editor based on the results obtained during the first usability test and the relevance of a user testing as tools are being developed.

Beside the qualitative data related to the tool, the preference questions allowed for gathering users' feedback on their needs when describing 360º content. Users' comments pointed to such relevant subjects as content selection and professional training, further confirming the results obtained in the literature review on storytelling and in the focus groups. In this regard, content selection in this media format was deemed challenging by participants. The obtained data points to the need for guidance on what to describe in form of best practice guidelines or specific training in order to work efficiently and provide a quality of service.

Additionally, relevant findings on the users' profile were found in the pre-questionnaire. When asked about the usage of the AD editors, the results show that not all audio describers use specific editing tools in their practice. Ten of the participants reported using only text editors and video players when writing AD. Although the results of the second iteration of the usability study do not form part of this thesis, it is worth mentioning that the results obtained in that study show that not all audio describers record their own descriptions (for details see Annex 4.3).
Similarly to the studies with AD users, professional audio describers were asked questions related to the usage of immersive technologies and the impact they may have on their practice.

Despite the overall positive attitude of the participants toward the immersive content, the results show that most of them do not watch VR on a daily basis. Indeed, 50% of the participants reported that they have not had the chance to use it. When asked how often they watch such content, 87.5% of the participants stated that they had never watched it on a smartphone plugged to a head-mounted display. 83.33% of the participants indicated that

they had never watched it in a HMD. This may be again due to the novelty of this technology.

One closed question in the preference questionnaire asked participants whether they consider that 360° videos will impact their work in the coming years. In the first iteration, 58.3% answered affirmatively to this question. In the second iteration, the results were more than 16 percentage points higher, reaching 75% of the participants. These numbers can be due to the gradual emergence of such technologies on the market and it is possible that they will further proliferate in the coming years.

To conclude this section, the two specific aims set for this user group were met, allowing for the fulfilment of the second main objective. Apart from assessing the tool in the experimental study, the thesis gathered valuable qualitative feedback on users' profile and the needs related to content selection in 360º content.

7.3. Contributions of this research

This thesis can be seen as one of the first steps in providing access to immersive content, an under-researched subject upon starting this investigation. The motivation behind this thesis was to address this knowledge gap and pave the way for future analyses related to virtual environments within Audiovisual Translation and Media Accessibility. The contributions of this thesis can be grouped into two broad categories: academic contribution and contribution to the AVT industry and broadcasters.

With regard to the former, this research presents one of the first studies in relation to virtual environments in the AD field. Its major academic contribution is related to the methodology employed. Inspired by the recent shift in Media Accessibility, this thesis is centred on the users, involving them at every stage of the research process. The novelty of the experimental studies conducted with the AD users reside in the presence measures employed. This thesis follows previous studies that employed presence scales as a measure of viewing experience in the content that can be categorized as "low-immersive on the scale of environment interactivity" (Fryer and Freeman 2013). This thesis adds to previous research, by measuring presence in the content which belongs to VR. Also, as this thesis compared 2D content with 360° videos in order to establish whether AD in 360° content

requires a specific approach to AD, it contributes to narratological studies within the AD field.

Another strength of this thesis is that this research is not country-specific, but instead the feedback has been gathered in more than one country when defining how AD should be implemented. As the focus groups conducted at the early stage of this thesis development were conducted both in Barcelona and in Poland, feedback from users belonging to two different AV traditions was gathered. Similarly, the design of the online questionnaires in English for the usability study allowed for gathering feedback from the users from different European countries, and outside Europe, including media accessibility experts for major internet media and technology companies (see Article 3, pp. 292–293).

This thesis, and particularly the conclusions of the Chapter 3, can be seen as preliminary feedback for audio describers on how to approach AD in this media format. Although the results reported in Chapter 4 show that the number of audio describers who have experience in writing audio description in 360° videos is currently very low, it is possible that the number of describers who work with immersive content will gradually grow in the next years. Similarly, the results show that during the development of this thesis, few audio describers considered that immersive media could impact on their professional practice. These results may be different if this question is to be repeated in the near-to-medium future. As the results obtained within this thesis show that audio describers may benefit from additional guidance regarding content selection in 360° videos, the conclusions reached in Chapter 3 of this thesis may serve as a first stepping stone towards the inclusion of immersive content in AD guidelines or as part of a training on AD.

Additionally, this thesis may benefit the AVT industry, as it proposes possible approaches for implementing AD in 360° videos. Another contribution for AVT is that the research has helped to define the requirements needed to develop an innovative editor, which can be used both for standard and 360° content. Chapter 5 also presents some feedback from AD users for content creators who may want to implement access services during the content production process, aiming at creating a high sense of presence for all.

7.4. Limitations

The limitations of this exploratory research have to be acknowledged before closing this thesis.

One of the main limitations is related to the research sample in all experimental studies conducted. Forty two professional audio describers (twenty four participants per test) took part in the usability studies. A total of thirty six AD users with sight loss participated in both the pilot and the main study. Increasing the number of participants in all the studies would be desirable in order to obtain a sounder statistical analysis of the results. However, it should be stressed that one of the most challenging aspects of conducting research with persons with sight loss is related to user recruitment (Chmiel and Mazur 2012). Because of this, sample sizes of more than twenty five persons with sight loss per group are considered acceptable when conducting statistical analyses, as has already been established in the field (Orero et al. 2018).

Regarding the experimental studies with the AD users, a chief limitation was linked to the lack availability of the 360° clips that would meet testing requirements. Due to the novelty of this media format, finding videos which would be comparable in terms of content and length and would also have enough spaces between the dialogues in order to insert audio descriptions was challenging. Likewise, the 360° videos available on the market are generally short, and their length is usually between 5–15 minutes (Allen and Tucker 2018; Agulló 2019). The duration of the videos used in the studies was, on average 3–4 minutes, and entire episodes were played to add to ecological validity. However, although there is no recommendation for an ideal stimuli length when measuring presence, a duration of 10–15 minutes is recommended in the gaming context (ITU 2018: 12), therefore, it would be worthwhile conducting similar studies with longer clips. Also, the experimental studies conducted for the purposes of this thesis used a travel documentary. Selecting clips belonging to fiction in order to compare the effect of the presentation modes on both genres would confer higher reliability on the results obtained.

Regarding the studies conducted with professional describers, one limitation pertains to the participants' profile in both usability studies. When designing the methodology, the research population for the study was controlled, allowing only for the participation of professional describers. Taking into account that a considerable number of participants

indicated a lack of previous experience with the AD software, it could have had an impact on the results obtained. Similarly, the fact that many participants indicated that they are only in charge of scripting descriptions and not recording them could have impacted on the results. On the one hand, this research design allowed for gathering feedback from all potential users of the editing tool. On the other, controlling the sample solely to the aforementioned groups would be worthwhile in order to compare both results, as factors such as the lack of familiarity with the new format of 360° videos and the new software, age, usage of AD editors and not being accustomed to record AD may have impacted the results.

Another limitation related to the usability study is linked to the fact that the 360° video clips used for the study were watched by the describers on the flat screens of laptops or PCs of participants taking part in the study, not in an HMD. This was due to the research design, which aimed to reach professional describers from different countries. However, watching the clips in a HMD before the study would have been beneficial, as it would have allowed participants to better understand the presented content.

Finally, more attention should be given to presence measures when conducting research with persons with sight loss in future studies, as the existing self-report measures are designed for sighted persons. An inclusive questionnaire, validated with persons with sight loss would add to presence research, and it would ensure the validity of the results obtained in the AVT field.

7.5. Pathways for future research

From the data obtained, numerous opportunities for future research become evident.

In light of the findings of this thesis, further research is needed on the implementation of spatial sound in AD. The present study on spatial sound could be replicated with the video clips that have original sound recorded in spatial sound technology and have more pauses between the dialogues for inserting audio description. A collaboration with content creators would be particularly valuable, as including AD already at the production stage would allow for the development of content with a high immersive capacity, from which everyone could benefit. Regarding spatial sound technology, the technology of object-based sound

(Simon, Torcoli and Paulus 2019) – which enables placement of the sound of AD exactly on the object – is of particular value.

In all of the experimental studies conducted for this thesis, data was gathered by means of subjective, post-hoc measures. It would be advisable to complement these with objective measures, which could include comprehension (Cabeza-Cáceres 2011; Chmiel and Mazur 2012, 2016; Walczak and Rubaj 2014), recall (Fresno 2016), eye-tracking, galvanic skin response (Kuniecki, Pilarczyk and Wichary 2015), heart variability (Kuniecki, Barry and Kaiser 2003) or cortisol secretion (Sgoifo et al. 2003) in future studies, in order to allow for a more detailed analysis.

Apart from AD, other modalities, for instance audio subtitling (AST), would also merit further research. Investigations into spatial sound could be focused not only on audio description, but also to audio subtitles. This subject should be given particular attention, as it would be particularly helpful for users to precisely localize the characters in a 360° storyworld. Elsewhere, more studies on AD and AST should be conducted in relation to videos with 6DOF or other types of immersive environments. Additionally, the effect of AD or AST voiced by means of speech synthesis could be studied in this media format, as this option is more cost effective.

It is hoped that the conclusions reached in this thesis, which are based on the results of extensive testing, will serve as a useful stepping stone towards a broader understanding of accessibility in virtual environments.

Bibliography

Bibliography

The reference lists in the articles that form this thesis follow different styles, conforming to the style guidelines established by each journal. To provide this thesis with a unified format, the APA 6th Referencing and Citation Style has been used in the list below.

Aenor. (2005). *Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguias*. Madrid: AENOR.

Agulló, B. (2019). *Nuevos retos para la TAV: subtitulación en medios inmersivos. Resultados de un análisis de corpus*. [PowerPoint slides]. Retrieved from https://ddd.uab.cat/record/201457

Agulló, B., Matamala, A., & Orero, P. (2018). From disabilities to capabilities: Testing subtitles in immersive environments with end users. *Hikma, 17,* 195–218. doi: 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. *Journal of Specialised Translation, 32,* 217–235.

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. doi:10.1017/S0890060419000362

Allen, C., & Tucker, D. (2018). *Immersive content formats for future audiences*. Retrieved from https://assets.ctfassets.net/nubxhjiwc091/5WypbBxFEQqS8I6WaKUQgg/ca25eb800e9f88a740079dbca623ca2c/Immersive_Content_Formats_for_Future_Audiences.pdf

American Council of the Blind. (2003). *ADI AD guidelines.* Retrieved from http://www.acb.org/adp/guidelines.html

American Council of the Blind. (2009). *Audio description standards*. Retrieved from http://www.acb.org/adp/docs/ADP_Standards.doc

Baker, M., & Saldanha, G. (Eds.). (2019). *Routledge Encyclopedia of Translation Studies.* London: Routledge.

Bal, M. (2009). *Narratology: Introduction to theory of narrative.* Toronto: University of Toronto Press.

Barbour, R. (2007). *Qualitative research kit: Doing focus groups.* London: Sage Publications.

Biocca, F. (1997). The cyborg's dilemma: Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication, 3*(2). doi: 10.1111/j.1083- 6101.1997.tb00070.x.

Bleumers, L., Van den Broeck, W., Lieven, B., & Pierson, J. (2013). Extending the field of view: A human-centered design perspective on 360° TV. *Behaviour & Information Technology, 33*(8), 800–814. doi: 10.1080/0144929X.2013.810780

Booton, J. (2013, October 11). By tricking the brain, Disney's bringing digital sight to the blind. *Foxbusiness.* Retrieved from https://www.foxbusiness.com/features/by-tricking-the-brain-disneys-bringing-digital-sight-to-the-blind

Bordwell, D., & Thompson, K. (1990). *Film art: An introduction*. New York: McGraw-Hill College.

Braun, S. (2007). Audio description from a discourse perspective: A socially relevant framework for research and training. *Linguistica Antverpiensia, 6*, 357–369.

Braun, S. (2008). Audio description research: State of art and beyond. *Translation Studies in a New Millenium 6,* 14–30.

Braun, S., & Orero, P. (2010). Audio description with audio subtitling: An emergent modality of audiovisual localization. *Perspectives: Studies in Translatology, 18*(3), 173–188. doi: 10.1080/0907676x.2010.485687

Brown, S., Ladeira, I., Winterbottom, C., & Blake, E. (2003). The effects of mediation in a storytelling Virtual Environment. In O. Balet, G. Subsol & P. Torguet (Eds.), *Virtual storytelling. Using Virtual Reality technologies for storytelling*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-40014-1_13

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

Bryman, A. (2008). *Social research methods*. Oxford, New York: Oxford University Press.

Burch, N. (1979). *To the distant observer. Form and meaning in the Japanese cinema*. London: Scolar Press.

Cabeza-Cáceres, C. (2011). *Intonation in AD: does it affect users' comprehension?* Paper presented at the 4th International Conference Media for All. Audiovisual Translation: Taking stock, London, Great Britain.

Chmiel, A., & Mazur, I. (2012). AD reception research. Some methodological considerations. In E. Perego (Ed.), *Emerging topics in translation: Audio description* (pp. 57–80). Trieste: Edizioni Università di Trieste.

Chmiel, A., & Mazur, I. (2014). Audiodeskrypcja [Audio description]. Poznań: Wydział Anglistyki UAM.

Chmiel, A., & Mazur, I. (2016). Researching preferences of audio description users. Limitations and solutions. *Across Languages and Cultures, 17*(2), 271–288. doi:10.1556/084.2016.17.2.7

Cho, H., Shen, L., & Wilson, K. (2014). Perceived realism: Dimensions and roles in narrative persuasion. *Communication Research, 41*(6), 828–851. doi:10.1177/0093650212450585

Colwell, C., Petrie, H., Kornbrot, D., Hardwick, A., & Furner, S. (1998). Haptic Virtual Reality for blind computer users. In M.M. Blattner & A.I. Karshner (Eds.), *Assets '98: Proceedings of the Third International ACM Conference on Assistive Technologies* (pp. 92–99). New York: ACM. doi: 10.1145/274497.274515

Damiani, J. (2017). Writing for VR: The definitive guide to VR storytelling. *VR Scout*. Retrieved from https://vrscout.com/news/writing-vr-definitive-guide-vr-storytelling/

Díaz-Cintas, J. (2005). Audiovisual translation today. A question of accessibility for all. *Translating Today Magazine, 4*, 3–5.

Di Giovanni, E. (2014). Visual and narrative priorities of the blind and non-blind: eye-tracking and audio description. *Perspectives: Studies in Translatology*, *22*, 136–153. doi: 10.1080/09076767X.2013.769610

Di Giovanni, E. (2018a). Audio description and reception-centred research. In E. Di Giovanni & Y. Gambier (Eds.), *Reception studies in Audiovisual Translation* (pp. 225–252). Amsterdam: Benjamins.

Di Giovanni, E. (2018b). Audio description for live performances and audience participation. *JosTrans: The Journal of Specialised Translation*, *29*, 189–211.

Directive 2010/13/EU. *Audiovisual Media Services Directive*. European Parliament, Council of the European Union. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0013

Directive (EU) 2019/882. *The European Accessibility Act*. European Parliament, Council of the European Union. https://eur-lex.europa.eu/legal-content/EN/LSU/?uri=CELEX%3A32019L0882

Dolan, D., & Parets, M. (2016). Redefining the axiom of story: The VR and 360 video complex. *Medium*. Retrieved from https://medium.com/@devon.michael/redefining-the-axiom-of-story-the-vr-and-360-video-complex-bee3c20d69df

Dooley, K. (2017). Storytelling with Virtual Reality in 360-degrees: a new screen grammar. *Studies in Australasian Cinema, 11*(3), 161–171. doi: 10.1080/17503175.2017.1387357

Ecorys. (2017). *Virtual Reality and its potential for Europe*. Retrieved from https://ec.europa.eu/futurium/en/system/files/ged/vr_ecosystem_eu_report_0.pdf

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

European Commission. (2015). *Europe 2020 strategy policy*. Retrieved from https://ec.europa.eu/digital-single-market/en/europe-2020-strategy

European Commission. (2017). *Virtual Reality at the service of healthcare.* Retrieved from https://ec.europa.eu/easme/en/news/virtual-reality-service-healthcare

Fels, D., Udo, J., Ting, P., & Diamond, J. (2006). Odd Job Jack described – a first person narrative approach to described video. *Journal of Universal Access in the Information Society, 5*(3), 73–88.

Fernández-Torné, A. (2016). Machine translation evaluation through post-editing measures in audio description. *inTRAlinea, 18*.

Fernández-Torné, A., & Matamala, A. (2015). Text-to-speech vs. human voiced audio descriptions: A reception study in films dubbed into Catalan. *The Journal of Specialised Translation, 24,* 61–88.

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

Fidyka, A., & Matamala, A. (2021). Retelling narrative in 360º videos: Implications for audio description. *Translation Studies*. doi: 10.1080/14781700.2021.1888783

Fresno, N. (2012). Experimenting with characters: An empirical approach to the audio description of fictional characters. In A. Remael, P. Orero & M. Caroll (Eds.), *Audiovisual Translation and Media Accessibility at the crossroads. Media for All 3* (pp. 147–171). Amsterdam: Rodopi.

Fresno, N. (2014). *La (re)construcción de los personajes fílmicos en la audiodescripción* (Doctoral dissertation, Universitat Autònoma de Barcelona, Barcelona, Spain). Retrieved from https://ddd.uab.cat/pub/tesis/2014/hdl_10803_285420/nfc1de1.pdf

Fresno, N., Castellà, J., & Soler, O. (2016). What should I say? Tentative criteria to prioritize information in the audio description of film characters. In A. Matamala & P. Orero (Eds.), *Researching audio description: New approaches* (pp. 143–167). London: Palgrave Macmillan. doi: 10.1057/978-1-137-56917-2_8.

Fryer, L. (2010). Audio description as audio drama – a practitioner's point of view. *Perspectives: Studies in Translatology, 18*(3), 205–213.

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

Fryer, L., & Freeman, J. (2012). Presence in those with and without sight: Audio description and its potential for Virtual Reality applications. *Journal of CyberTherapy & Rehabilitation, 5*(1), 15–23.

Fryer, L., & Freeman, J. (2013). Cinematic language and the description of film: Keeping AD users in the frame. *Perspectives: Studies in Translatology, 21*(3), 1–15. doi: 10.1080/0907676X.2012.693108

Fryer, L., & Freeman, J. (2014). Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In A. Felnhofer & O. Kothgassner (Eds.), *Challenging presence: Proceedings of the 15th International Conference on Presence* (pp. 99–107). Wien: Facultas.wuv.

Fryer, L., Pring, L., & Freeman, J. (2013). Audio drama and the imagination: The influence of sound effects on presence in people with and without sight. *Journal of Media Psychology: Theories, Methods, and Applications, 25*(2), 65–71. doi: 10.1027/1864-1105/a000084

Ghali, N., Soluiman, O., El-Bendary, N., Nassef, T., Ahmed, S., Elbarawy, Y. M., & Hassanien, A. E. (2012). Virtual Reality technology for blind and visually impaired people: Reviews and recent advances. In T. Gulrez & A. E. Hassanien (Eds.), *Advances in robotics and Virtual Reality, 26* (pp. 363–385). Berlin, Heidelberg: Springer.

Gleb, B. (n.d.). VR vs AR vs MR: Differences and real-life applications. *Ruby Garage Blog*. Retrieved from https://rubygarage.org/blog/difference-between-ar-vr-mr

Gödde, M., Gabler, F., Siegmund, D., & Braun, A. (2018). Cinematic narration in VR – Rethinking film conventions for 360 degrees. In J. Y. Chen (Ed.), *Virtual, augmented and mixed reality: Applications in health, cultural heritage and industry* (pp. 184–201). Cham: Springer International Publishing. doi: 10.1007/978-3-319-91584-5_15

Greco, G. M. (2016). On accessibility as a human right, with an application to Media Accessibility. In A. Matamala & P. Orero (Eds.), *Researching audio description: New approaches* (pp. 11–33). London: Palgrave Macmillan.

Greco, G. M. (2018). The Nature of Accessibility Studies. *Journal of Audiovisual Translation 1*, 205–232.

Greco, G. M. (2019). Fostering unity in diversity: The social value of European research on Media Accessibility, *EuroScientist*. Retrieved from https://www.euroscientist.com/fostering-unity-in-diversity-the-social-value-of-european-research-on-media-accessibility/

Greening, J. (2011). Do we need 3D audio description guidelines: Recommendations from focus group study report. *Royal National Institute of Blind People*. Retrieved from: http://audiodescription.co.uk/uploads/general/3D_film_and_tv_report_2.pdf

Hanson, N. (2019, March 4). Storytelling of the future. *BBC*. Retrieved from https://www.bbc.co.uk/rd/blog/2019-02-storytelling-interactive-digital-drama

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P., Neerincx, M., & Brinkman W. P. (2014). Controlling social stress in Virtual Reality environments. *Plos One, 9*(3). doi: 10.1371/journal.pone.0092804

Herman, D. (2002). *Story Logic: Problems and Possibilities of Narrative. Narrative*. Lincoln: University of Nebraska Press.

Herre, J., Hilpert, J., Kuntz, A., & Plogsties, J. (2015). MPEG-H 3D audio – The new standard for coding of immersive spatial audio. *IEEE Journal Of Selected Topics In Signal Processing, 9*(5), 770–779. doi: 10.1109/JSTSP.2015.2411578

Igroup. (n.d.) *Igroup presence questionnaire (IPQ) scale construction*. Retrieved from http://www.igroup.org/pq/ipq/construction.php

International Organization for Standardization. (2015). *Information technology – User interface component accessibility – Part 21*: *Guidance* on *audio descriptions (ISO/IEC TS 20071-21:2015)*. Retrieved from https://www.iso.org/standard/63061.html

International Telecommunication Union. (2018). *Subjective evaluation methods for gaming quality (ITU-T P.809)*. Retrieved from https://www.itu.int/rec/T-REC-P.809-201806-I/en

IRT Lab. (n.d.) Interactive Content. *Interactive Lab Website*. Retrieved from https://lab.irt.de/demos/object-based-audio/interactive/

Iturregui-Gallardo, G. & Méndez-Ulrich, J. L. (2019). Towards the creation of a tactile version of the self-assessment manikin (T-SAM) for the emotional assessment of visually impaired people, *International Journal of Disability, Development and Education, 67*(6), 1–18. doi: 10.1080/1034912X.2019.1626007

Johansson, M. (2019). VR for your ears: Dynamic 3D audio is coming soon. *IEEE Spectrum.* Retrieved from https://spectrum.ieee.org/consumer-electronics/audiovideo/vr-for-your-ears-dynamic-3d-audio-is-coming-soon

Jakobson, R., (2009). O językoznawczych aspektach przekładu [On linguistic aspects of translation]. In P. Bukowski & M. Heydel (Eds.), *Współczesne teorie przekładu. Antologia* (pp. 41–49). Kraków: Znak.

Jankowska, A. (2015). *Translating audio description scripts: Translation as a new strategy of creating audio description*. Frankfurt am Main: Peter Lang.

Jankowska, A. (2018). *Translation, crowdsourcing, collaboration and quality in audio description*. Paper presented at the UMAQ conference, Barcelona, Spain. doi: 10.13140/RG.2.2.12736.02564

Jankowska, A., Fidyka, A., Psiuk, A., & Stanisławska-Kloc, S. (2017a). *Audio description legally-speaking*. Paper presented at the MAPIC conference, Vigo, Spain.

Jankowska, A., Szarkowska, A., Krejtz, K., Fidyka, A., Kowalski, J., & Wichrowski, M. (2017b). Smartphone app as a museum guide. Testing the Open Art application with blind, deaf, and sighted users. *Rivista internazionale di tecnica della traduzione, 19*, 113–130.

Jerald, J. (2015). *The VR book: Human-centered design for Virtual Reality*. San Rafael, CA: Morgan & Claypool. doi: 10.1145/2792790.2792827

Johansson, M. (2019). *VR for your ears: Dynamic 3D audio is coming soon*. Retrieved from https://spectrum.ieee.org/consumer-electronics/audiovideo/vr-for-your-ears-dynamic-3d-audio-is-coming-soon

Jones, S. (2017). Disrupting the narrative: Immersive journalism in virtual reality. *Journal of Media Practice, 18*(2-3), 171–185. doi: 10.1080/14682753.2017.1374677

Kinateder, M., Gromer, D., Gast, P., Buld, S., Müller, M., Jost, M., Nehfischer, M., Mühlberger, A., & Pauli, P. (2015). The effect of dangerous goods transporters on hazard perception and evacuation behavior – a virtual reality experiment on tunnel emergencies. *Fire Safety Journal, 78*, 24–30. doi: 10.1016/j.firesaf.2015.07.002

Krijn, M., Emmelkamp, P. M., Biemond, R., de Wilde de Ligny, C., Schuemie, M. J., & van der Mast, C. A. (2004). Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy, 42*(2), 229–239. doi: 10.1016/S0005-7967(03)00139-6

Krueger, R. (1998). *Focus group kit: Moderating focus groups*. Thousand Oaks: SAGE Publications. doi: 10.4135/9781483328133

Kruger, J. L. (2010). Audio narration: Re-narrativising film. *Perspectives: Studies Translatology, 18*(3), 232–248. doi: 10.1080/0907676X.2010.485686

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

Kruger, J. L., Soto-Sanfiel, M., Doherty, S., & Ibrahim, R. (2016). Towards a cognitive audiovisual translatology: Subtitles and embodied cognition. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 171–194). doi: 10.1075/btl.128

Kruger, J. L., Doherty, S., & Ibrahim, R. (2017). Electroencephalographic beta coherence as an objective measure of psychological immersion in film. *International Journal of Translation, 19,* 99–111. doi:10.13137/2421-6763/17353

Kruger, J. L., Doherty, S., & Soto-Sanfiel, M. (2017). Original language subtitles: Their effects on the native and foreign viewer. *Comunicar, 50*, 23–32. doi: 10.3916/C50-2017-02.

Kuniecki, M., Barry, R., & Kaiser, J. (2003). The influence of emotionally relevant context on the evoked cardiac response triggered by an irrelevant stimulus. *Journal of Psychophysiology, 17*(2), 61–68.

Kuniecki, M., Pilarczyk, J., & Wichary, S. (2015). The colour red attracts attention in an emotional context. An ERP study. *Frontiers in Human Neuroscience, 9*, 1–14. doi: 10.3389/fnhum.2015.00212

Lee, N. (2017). Pixar's VR debut takes you inside the entrancing world of 'Coco'. *Engadget*. 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. doi: 10.1162/105474601300343612

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2). doi: 10.1111/j.1083-6101.1997.tb00072.x

López, M., Kearney, G., & Hofstädter, K. (2016). Enhancing audio description: sound design, spatialisation and accessibility in film and television. *Proceedings of the Institute of Acoustics, 38(2)*. Retrieved from http://enhancingaudiodescription.com/wp-content/uploads/2016/11/RS2016-paper-Lopez-et-al.pdf

Manjoo, F. (2014, April 3). If you like immersion, you'll love this reality. *The New York Times*. Retrieved from https://mobile.nytimes.com/2014/04/03/technology/personaltech/virtual-reality-perfect-for-an-immersive-society.html

Maszerowska, A. (2014). The verbalizations of the effects of light and contrast in audio description (Doctoral dissertation, Universitat Autònoma de Barcelona, Barcelona, Spain). Retrieved from https://ddd.uab.cat/record/129133

Maszerowska, A., Matamala, A., & Orero, P. (Eds.). (2014). *Audio description: New perspectives illustrated*. Amsterdam: John Benjamins.

Matamala, A. (2014). Audio describing text on screen. In A. Maszerowska, A. Matamala & P. Orero (Eds.), *Audio description. New perspectives illustrated* (pp. 103–120). Amsterdam: John Benjamins. doi: 10.1075/btl.112.07mat

Matamala, A., & Remael, A. (2015). Audio description reloaded. An analysis of visual scenes in 2012 and Hero. *Translation Studies, 8*(1), 63–81. doi: 10.1080/14781700.2014.943678

Mazur, I. (2014). Gestures and facial expressions in audio description. In A. Matamala & P. Orero (Eds.), *New perspectives illustrated* (pp. 179–197). Amsterdam: Benjamins.

Mazur, I., & Chmiel, A. (2016). Should audio description reflect the way sighted viewers look at films? Combining eye-tracking and reception study data. In A. Matamala & P. Orero (Eds.), *Researching audio description. New approaches* (pp. 97–121). London: Palgrave Macmillan. doi: 10.1057/978-1-137-56917-2_6

Mazur, I., & Kruger, J. L. (2012). Pear Stories and audio description: Language, perception and cognition across cultures, *Perspectives, 20*(1), 1–3.
doi: 10.1080/0907676X.2012.633769

Neves, J. (2012). Multisensory approaches to (audio)describing the visual arts. *MonTI, 14*(13), 277–293. doi: 10.6035/*MonTI*.2012.4.12

Nielsen, L. T., Møller, M. B., Hartmeyer, S. D., Ljung, T. C. M., Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Missing the point: An exploration of how to guide users' attention during cinematic virtual reality. *Proceedings of the 22nd ACM Symposium on Virtual Reality Software and Technology*. New York: Association for Computing Machinery. doi: 10.1145/2993369.2993405

Office of Communications. 2000. *ITC Guidance on standards for audio description.* Retrieved from
http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf

Orero, P. (2007). Pioneering audio description: An interview with Jorge Arandes. *Journal of Specialized Translation, 7,* 179–189.

Orero, P. (2011). Audio description for children: Once upon a time there was a different audio description for characters. In E. Di Giovanni (Ed.), *Entre texto y receptor: Accesibilidad, doblaje y traducción* [Between text and receiver: Translation and accessibility] (pp. 169–184). Peter Lang: Frankfurt am Main.

Orero, P., & Vilaró, A. (2012). Eye tracking analysis of minor details in films for audio description. *Multidisciplinarity in Audiovisual Translation*, *4*, (295–319). doi: 10.6035/MonTI.

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. *JoSTrans, 30*, 105–126.

Ortiz-Boix, C., & Matamala, A. (2013). Post-editing wildlife documentary films: a new possible scenario? *The Journal of Specialised Translation, 26*. 187–210.

Perego, E. (2014). Film language and tools. In A. Maszerowska, A. Matamala & P. Orero (Eds.), *Audio description. New perspectives illustrated* (pp. 81–101). Amsterdam: Benjamins.

Perego, E. (2016). History, developments, challenges and opportunities of empirical research in Audiovisual Translation. *Across Languages and Cultures, 17*(2), 155–162. doi: 10.1556/084.2016.17.2.1

Picinali, L., Feakes, C., Etherington, A., & Lloyd, T. (2011). VR interactive environments for the blind: Preliminary comparative studies. In K. Helin & M. D'Cruz (Eds.), *Proceedings of the Joint Virtual Reality Conference of euroVR and EGVE* (pp. 113–115). Helsinki: Julkaisija-Utgivare.

Portillo, C. (2018). Films unseen: Approaching sound design alternatives to enhance perception and mental imagery models among audio description audiences on sci-fi films. *Proceedings from Accessibility in Film, Television and Interactive Media Conference*. Retrieved from https://www.dropbox.com/sh/08009l5g2g88i97/AAAI3b_yRkrrljNuyiPeRvP7a?dl=0

Rai, S. (2009). *Bollywood for all: The demand for audio described Bollywood films.* London: Royal National Institute for the Blind.

Ramos, M. (2015). The emotional experience of films: does Audio Description make a difference? *The Translator, 21*(1), 68–94. doi: 10.1080/13556509.2014.994853

Ramos Caro, M. (2016). Testing audio narration: the emotional impact of language in audio description. *Perspectives: Studies in Translatology*, *24*(4), 1–29. doi: 10.1080/0907676X.2015.1120760

Ramos Caro, M., & Rojo López, A. (2014). 'Feeling' audio description: Exploring the impact of AD on emotional response. *Translation Spaces, 3,* 133–150. doi: 10.1075/ts.3.06ram

Regenbrecht, H., & Schubert, T. (2002). Real and illusory interaction enhance presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *11*(4), 425–434. doi: 10.1162/105474602760204318

Remael, A. (2012). For the use of sound. Film sound analysis for audio description: Some key issues. *MonTI – Monographs in Translation and Interpreting, 4*, 255–276.

Remael, A., Orero, P., & Caroll, M. (Eds.) (2012). *Audiovisual Translation and Media Accessibility at the crossroads. Media for All 3*. Leiden: Rodopi.

Remael, A., Reviers, N., & Vercauteren, G. (Eds). (2015). *Pictures painted in words. ADLAB audio description guidelines*. Trieste: Edizioni Università di Trieste. Retrieved from http://www.adlabproject.eu/Docs/adlab%20book/index.html

Reviers, N. (2016). Audio description services in Europe: An update. *The journal of Specialized Translation*, *26,* 232–247.
Retrieved from https://www.jostrans.org/issue26/art_reviers.pdf

Reviers, N., & Remael, A. (2015). Recreating multimodal cohesion in audio description: A case study of audio subtitling in Dutch multilingual films. *New Voices in Translation Studies, 13*(1), 50–78.

Reyes, M. C. (2017). *Screenwriting framework for an interactive virtual reality film*. Paper presented at the 3rd Immersive Research Network Conference iLRN, Coimbra, Portugal. Retrieved from:
https://openlib.tugraz.at/download.php?id=5e6b52c99fbfa&location=browse

Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between Audiovisual Translation, Accessibility and Filmmaking. *JosTrans. The Journal of Specialised Translation, 20*, 201–223.

Romero-Fresco. P., & Fryer, L. (2013). Could audio described films benefit from audio introductions? An audience response study. *Journal of Visual Impairment and Blindness, 107*(4), 287–285.

Rosakranse, C., & Oh, S. Y. (2014). Measuring presence: The use trends of five canonical presence questionnaires from 1998-2012. In A. Felnhofer & O. D. Kothgassner (Eds.), *Challenging Presence: Proceedings of the 15th International Conference on Presence* (pp. 25–30). Wien: facultas.wuv.

Rothe, S., Hußmann, H., & Allary, M. (2017). Diegetic cues for guiding the viewer in cinematic virtual reality. *VRST'17 – Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. New York: ACM. doi: 10.1145/3139131.3143421

Rothe, S., Tran, K., & Hußmann, H. (2018). Dynamic subtitles in cinematic virtual reality. *TVX'18 – Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video.* doi: 10.1145/3210825.3213556

Saldanha, G., & O'Brien, S. (2013). *Research methodologies in Translation Studies*. Manchester: St Jerome Publishing.

Sauro, J., & Lewis, J. (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*(2), 69–71. doi: 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* (pp. 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. doi: 10.1162/105474601300343603

Schwind, V., Knierim, P., Haas, N., & Henze, N. (2019). Using presence questionnaires in virtual reality. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York: ACM. Retrieved from: https://vali.de/wp-content/uploads/2019/01/2019-CHI-QuestionnairesInVR.pdf

Sgoifo, A., Braglia, F., Costoli, T., Musso, E., Meerlo, P., Ceresini, G., & Troisi, A. (2003). Cardiac autonomic reactivity and salivary cortisol in men and women exposed to social stressors: Relationship with individual ethological profile. *Neuroscience and Biobehavioural Reviews, 27*, 179–188.

Sherman, W., & Craig, A. (2003). *Understanding virtual reality: Interface, application, and design*. Amsterdam: Morgan Kaufmann Publishers.

Slater, M., & Usoh, M. (1993). Representations systems, perceptual position, and presence in immersive virtual environments. *Presence. Teleoperators and Virtual Environments, 2*(3), 221–233.

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

Simon, C, Torcolli, M., & Paulus, J. (2019). MPEG-H audio for improving accessibility in broadcasting and streaming. Retrieved from: https://arxiv.org/abs/1909.11549

Snyder, J. (2008). Audio description: The visual made verbal. In Díaz-Cintas, J. (Ed.), *The didactics of Audiovisual Translation* (pp. 191–198). Amsterdam/Philadelphia: John Benjamins Publishing. doi: 10.1075/btl.77.18sny

Steuer, J. (1992). Defining Virtual Reality: Dimensions determining telepresence. *Journal of Communication, 42*(4), 73–93.

Suojanen, T., Koskinen, K., & Tuominen, T. (2015). *User-centered translation*. Abingdon, NY: Routledge.

Syrett, H., Calvi, L., & Van Gisbergen, M. (2017). The culus Rift film experience: a case study on understanding films in a head-mounted display. In R. Poppe, J. J. Meyer, R. Veltkamp & M. Dastani (Eds.), *Intelligent technologies for interactive entertainment* (pp. 197–208). Berlin: Springer. doi: 10.1007/978-3-319-49616-0_19

Szarkowska, A. (2013). Auteur description: from the director's creative vision to audio description. *Journal of Visual Impairment & Blindness, 107*(5), 383–387.

Szarkowska, A., & Jankowska, A. (2012). Text-to-speech audio description of voiced-over films. A case study of audio described *Volver* in Polish. In E. Perego (Ed.), *Emerging topics in translation: Audio description* (pp. 81–94). Trieste: Edizioni Università Di Trieste.

Szarkowska, A., Krejtz, I., Krejtz, K., & Duchowski, A. (2013). Harnessing the potential of eye-tracking for Media Accessibility. In S. Grucza, M. Pluzycka & J. Alnajjar (Eds.), *Translation Studies and eye-tracking analysis* (pp. 153–183). Frankfurt am Main: Peter Lang. doi: 10.3726/978-3-653-02932-1

Szarkowska, A., & Wasylczyk, P. (2014). Audiodeskrypcja autorska [Auteur description]. *Przekładaniec, 28,* 48–62. doi: 10.4467/16891864PC.14.004.1711

Udo, J. P., & Fels, D. (2009). Suit the action to the word, the word to the action. An unconventional approach to describing Shakespeare's Hamlet. *Journal of Visual Impairment and Blindness, 103*(3), 178–183.

Udo, J. P., & Fels, D. (2010a). The rogue poster-children of universal design: Closed captioning and audio description. *Journal of Engineering Design, 21*(2-3), 207–221. doi: 10.1080/09544820903310691

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

United Nations. (1948). Universal Declaration of Human Rights. Retrieved from http://www.un.org/en/universal-declaration-human-rights

United Nations. (2006). *Convention on the Rights of Persons with Disabilities (A/RES/61/106 2006).* Retrieved from https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html

Vandaele, J. (2012). What meets the eye. Cognitive narratology for audio description. *Perspectives: Studies in Translatology, 20*(1), 87–102. doi: 10.1080/0907676X.2011.632683

Vercauteren, G. (2012). A Narratological Approach to Content Selection in Audio Description. A Strategy for the Description of Narratological Time. *MonTI, 4,* 207–231.

Vercauteren, G. (2016). A narratological approach to content selection in audio description. (Doctoral dissertation, University of Antwerp, Antwerp, Belgium).

Vercauteren, G., & Orero, P. (2013). Describing facial expressions: Much more than meets the eye. *Quaderns. Revista de Traducció, 20,* 187–199.

Vercauteren, G., & Remael, A. (2014). Spatio-temporal settings. In A. Matamala (Ed.), *Perspectives Illustrated* (pp. 61–80). Amsterdam: Benjamins.

Viswanathan, L. N., McDaniel, T., Krishna, S., & Panchanathan, S. (2010). Haptics in audio described movies. Paper presented at the *IEEE International Symposium on Haptic Audio Visual Environments and Games*, Phoenix, USA. doi: 10.1109/HAVE.2010.5623958

Walczak, A. (2017a). Creative description: Audio describing artistic films for individuals with visual impairments. *Journal of Visual Impairment & Blindness, 111*(4), 387–391.

Walczak, A. (2017b). (Doctoral dissertation, Universitat Autònoma de Barcelona, Barcelona, Spain).

Walczak, A., & Rubaj, M. (2014). Audio Description in History, Biology and Physics Classes for Visually Impaired Learners. *Przekładaniec, 28,* 63–79.

Walczak, A., & Fryer, L. (2017). Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment, 35*(1), 6–17.

Web Content Accessibility Guidelines 2.1. (2020). *1.2.7. Extended audio description (Prerecorded)*. Retrieved from https://www.w3.org/WAI/WCAG21/Understanding/extended-audio-description-prerecorded.html

Whitehead, J. (2005). What is audio description? In S. Jones (Ed.), *Proceedings of Vision 2005, International Congress Series, 1282* (pp. 960–963). doi: 10.1016/j.ics.2005.05.194

Wilken, N., & Kruger, J. L. (2016). Putting the audience in the picture. Mise-en-shot and psychological immersion in audio described film. *Across Languages and Cultures, 17*(2), 251–270. doi: 10.1556/084.2016.17.2.6

Wissmath, B., Weibel, D., & Groner, R. (2009). Dubbing or subtitling? Effects on spatial presence, transportation, flow, and enjoyment. *Journal of Media Psychology: Theories, Methods, and Applications, 21*, 114–125. doi: 10.1027/1864-1105.21.3.114.

Wissmath, B., & Weibel, D. (2012). Translating movies and the sensation of 'being there'. In E. Perego (Ed.), *Eye Tracking in Audiovisual Translation* (pp. 281-298). Rome: Aracne Editrice.

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

Zhao, Y., Cutrell, E., Holz, C., Morris, M., Ofek, E., & Wilson, A. (2019). Seeing VR: a set of tools to make virtual reality more accessible to people with low vision. Paper presented at the CHI 2019 Conference. Glagow, Great Britain. doi: 10.1145/3290605.3300341

Annexes

Annex 1. Articles as published

Annex 1.1. Article 3

Fidyka, A., & Matamala, A. (2019). Production of access services in immersive content: Understanding the needs of audio describers. *Hikma*, 18(2), 277–300. doi: 10.21071/hikma.v18i2.11683

ISSN: 1579-9794

Production of access services in immersive content: understanding the needs of audio describers

ANITA FIDYKA,
ANNA MATAMALA
anita.fidyka@uab.cat
anna.matamala@uab.cat
Universitat Autònoma de Barcelona

Fecha de recepción: 20 de febrero de 2019
Fecha de aceptación: 26 de julio de 2019

Resumen: Los contenidos en 360°, ofrecidos cada vez con más frecuencia por diversas emisoras e instituciones culturales, deberían satisfacer las necesidades de todos los miembros de nuestra sociedad, incluidas las personas con pérdida de visión. Sin embargo, hasta ahora se ha investigado poco sobre la audiodescripción (AD) de dichos contenidos. Este artículo presenta los resultados de un estudio de usabilidad del prototipo de editor de AD que se ha desarrollado en el proyecto Immersive Accessibility (ImAc). También pone de relieve las necesidades de los audiodescriptores profesionales cuando trabajan con contenidos de 360°. El editor es una herramienta en línea que permite a los audiodescriptores elegir diferentes tipos de sonido para la AD y colocar los segmentos de la AD en la esfera de 360°. El estudio se llevó a cabo en línea y los datos se recopilaron a través de un precuestionario demográfico y un poscuestionario que incluía el SUS (System Usability Scale) y preguntas sobre preferencias. Los resultados obtenidos proporcionan información valiosa sobre cómo mejorar las funcionalidades de la herramienta. También detectan la necesidad de disponer de directrices sobre la selección de información en este formato, lo que sugiere que la AD en contenido inmersivo podría integrarse en estándares o recomendaciones de AD o en cursos de formación. Los resultados de esta investigación son solo un punto de partida en el campo de la accesibilidad inmersiva, de ahí la recomendación de seguir investigando sobre accesibilidad en este formato.

Palabras clave: Traducción audiovisual, Accesibilidad en los medios, audiodescripción, vídeos de 360°, usabilidad

Abstract: 360° content, offered more and more frequently by various broadcasters and culture institutions, should cater for the needs of all members of our society, including persons with sight loss. So far, however,

the question of providing audio description (AD) in such content has been under researched. This study aims to report the results of the usability study of the prototype AD editor developed within the Immersive Accessibility (ImAc) project, which allowed us to gain insights into the needs of professional audio describers when working with 360° content. The editor is an online tool which allows describers to choose an appropriate sound type for AD, and place AD segments in the 360° sphere. The study was conducted online and data was collected by means of a demographic pre-questionnaire and a post-questionnaire, consisting of a System Usability Scale and additional preference questions. The results obtained provide valuable feedback on how to improve the functionality of the tool to meet the needs of its users. They also indicate the need for guidance when selecting content to be described in this media format, which suggests that AD in immersive content could be integrated into AD guidelines or specific courses offered by training institutions. The results of this study are just a starting point in the field of immersive accessibility, hence the recommendation for further research on the subject of accessibility in this media format.

Keywords: Audiovisual Translation, Media Accessibility, audio description, 360° videos, usability

INTRODUCTION

Audio description (AD) is an access service that is considered as a way of retelling a story: it translates the visual channel to a verbal mode (Maszerowska, Matamala, & Orero, 2014; Fryer, 2016; Snyder, 2008). It is used widely to render standard, two-dimensional audiovisual (AVT) products, such as films and TV programs, accessible. It is also used to describe, for instance, artworks in museums or live events, rendering cultural property accessible for those who cannot access the visuals, and to other groups at risk of social exclusion (Greco, 2016). So far, AD has been mainly provided for two-dimensional content (Fidyka and Matamala, 2018a). Although more interactive forms of AD exist, such as AD in the theatrical environment (Fryer and Freeman, 15) or AD in planetariums – in which content is displayed in a dome, surrounding the viewer –, research on AD in immersive media is practically non-existent nowadays (Fidyka and Matamala, 2018a). The few studies conducted to date include audio description in 3D cinema (Greening, 2011), or the integration of haptics in audio description (Viswanathan, McDaniel, Krishna, & Panchanathan, 2010). Regarding other access services, the implementation of subtitles has been researched (Agulló, Matamala, & Orero, 2018; Agulló, Matamala, & Orero, In press; Rothe, Tran, & Hußmann, 2018) and some solutions on the implementation have been proposed by

major distributors (Brown, 2017). However, the issue of implementing AD in 360° content has not been addressed so far.

All new technologies appearing on the market should be accessible to ensure that all members of our society have access to culture and arts, as specified in documents on Human Rights, such as the Universal Declaration of Human Rights and the UN Convention on the Rights of Persons with Disabilities. To ensure that such highly visual environments are accessible for persons with sight loss, devices used to consume immersive content (eg. glasses) should be made accessible and access services such as audio description should be provided.

Although immersive media are still emerging on the market, they already have a wide array of creative formats (Allen and Tucker, 2018). One of them is 360° videos, belonging to Virtual Reality (VR). These videos are typically between 5 and 15 minutes long, and they are «the most tightly authored among all VR formats» (Allen and Tucker, 2018:17) that have particular market potential. It means that the story – which is told with a central protagonist's «rise, fall and resolution» (ibid.) – is linear, often driven by dialogue and pre-scripted by a director or content creator (ibid.).

In 360° videos, also referred to as omnidirectional or spherical videos, the main challenge lies in storytelling techniques, which are still being defined. Sighted users can access 360° content by means of head-mounted displays, i.e. special glasses to access the Virtual Reality content, which allow them to feel as if they are inside a sphere, while the linear story pre-scripted by a director unfolds all around them (Fidyka and Matamala, 2018b; Dooley, 2017). As different events can occur at different angles of the visual scene, guiding the sighted users inside the virtual world towards the main action is considered challenging (Rothe, Hußmann, & Allary, 2017; Jones, 2017; Syrett, Calvi, & van Gisbergen, 2017), and various storytelling techniques are being researched (Gödde, Gabler, Siegmund, & Braun, 2018; Jerald, 2015). In any case, sighted audiences can look around the sphere, following the main story or ignoring it.

The media format is mostly visually-driven, with images being triggered by head movements. Finding a way to guide persons with sight loss effectively inside the sphere increases the challenge (Fidyka and Matamala, 2018a, 2018b). The technology of spatial sound, which gives audiences a three-dimensional soundscape of the elements presented in the content may prove of value, as well as different scripting styles. The implementation of spatial sound, including ambisonics (Johansson, 2019) in AD provided for 360° videos is being researched within the Immersive Accessibility (ImAc) project.

Its application in AD is also being investigated by other researchers (López, Kearney, & Hofstädter, 2016; Portillo, 2018).

ImAc is an H2020 project based on a user-centred methodology (Suojanen, Koskinen & Tuominen, 2015). Thanks to this methodology, end users and professional audio describers collaborate with project partners at every stage of the project. In its early stages, feedback from users was gathered through a series of studies based on focus groups (Fidyka and Matamala 2018a, 2018b). These qualitative studies allowed researchers to define the implementation of access services and editing tools, and their results suggest that spatial sound can serve as a tool to facilitate orientation within the sphere. As far as the production of AD in immersive content is concerned, the development of an editing tool for this access service is one of the aims of the project.

To respond to the challenge of guiding persons with sight loss in this new medium, a new method of producing audio description needed to be proposed prior to the development of the AD web editor. Based on the early feedback from end users, it was decided that when providing AD by means of the web editor, audio describers would be able to choose between three different audio description modes, associated with specific sound and scripting features, as discussed in detail in the next section. This choice will allow describers to choose the most appropriate style for each content (Allen & Tucker, 2018). This new approach also needed to be tested with professional describers, as it impacts on their current workflow.

This article will discuss the methodology and results of the usability tests of the AD web editor in its first prototype version, which allowed us to understand the needs of professional audio describers when producing AD in this media format. To contextualize the study, the main features of the prototype editor, developed by Anglatècnic, including the functions specific to immersive environments, will be described in the next section. Section 2 explains the methodology of the test, and section 3 offers the discussion of its results. Although the article reports on the results of a usability test, the interest lies in the interesting insights provided by professionals on the specific challenges the immersive content poses when audio describing, and how this can impact their current practices.

1. FEATURES OF THE WEB EDITING TOOL

The following section describes different features of the first version of the AD editing tool. Firstly, standard features that can also be found in other existing editors are discussed. In the second part, features specific to the medium under discussion are explained.

1.1. Standard features

The AD editor is an online tool, and is comprised of different sections. In the central section of the editor, audio describers can edit the AD script, add timecodes and preview the video with AD, as shown on Figure 1.

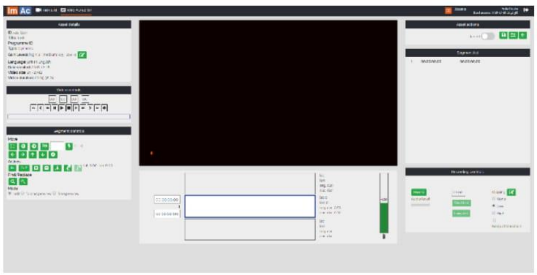


Figure 1 General view of the AD web editor

Another section, «Asset details», displays the basic information about the video, such as its name, size and language. Also, specific messages appear in that section in case of errors. Figure 2 shows the «Video controls», which allow audio describers to play, pause or stop the video, and navigate through it. Video controls have their own shortcuts, which were pre-set for the test, but will be customizable in future versions of the AD editor.

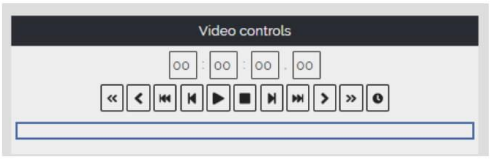


Figure 2 Video controls

The next section, «Segment controls», is related to audio description. It allows audio describers to add, remove and navigate through AD segments as well as add or remove timecodes (Figure 3). It also allows audio describers to set the angle, as described in section 1.2. Similarly to «Video controls», «Segment controls» also have their own shortcuts.

Figure 3 Segment controls section

As already mentioned, audio describers edit the script below the video player (Figure 4). To the left of the script, audio describers can set timecodes for a given AD segment. The number of the segment is also displayed in this section. To the right of the script editing area, the longitude and latitude of the current segment can be found. Also, the duration of the segment is displayed below. In the corner of the right side of this section, reading speed is displayed. When the reading speed is appropriate, the colour is green. When the number of characters per second is too high, it becomes red.

Figure 4 Script editing area

On the right side of the editor, there are three sections, named respectively: «Asset action», «Segment list» and «Recording controls» (Figure 5). The section «Asset action» allows to save audio files, and go back to the main page of the editing interface. «Segment list» contains the AD script with timecodes and a segment number. When AD for a given segment is recorded, the colour of this segment changes to green. When all segments are green, it means that all segments are recorded correctly. AD segments can be recorded by pressing the «Record» button. A countdown is provided for the recording to show audio describers how much time is left according to the timecodes set by them. Below the «Record» button, the audio level of the recording can be checked. After the AD is recorded, the recording can be

previewed in two tests: one starting two seconds before the timecode, and another one starting 5 seconds before the timecode.

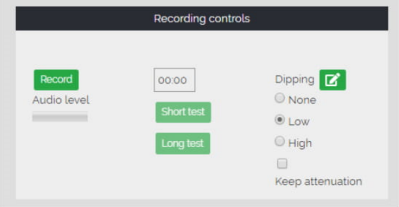


Figure 5 Recording controls

Additionally, fading can be chosen by audio describers. It refers to the decrease in the volume of the video when the AD is playing. In this regard, audio describers can choose between: none, low or high attenuation. The «None» option does not reduce the volume of the video. Conversely, when the option «High» is chosen, it will result in the significant reduction of the volume of the video. When the box «Keep fading» is checked, the volume of the main audio will be lowered until the next AD segment.

1.2. Features specific to immersive content

The AD web editor allows professional users to produce, preview and record audio description for 360° videos. As such, it includes functions which are specific to immersive content, which happens all around the 360° sphere.

Because of the characteristics of this media format, three types of AD can be created for each 360° video clip when working with the editor. They differ regarding the placement of spatial sound, and it is also recommended that audio describers use a different way of scripting for each of them. For the purposes of this project, these three AD types are provisionally referred to as: «Classic», «Static» and «Dynamic» (Figure 6). When professional users access the web editor, they choose one of them. The first sound option allows users to hear the AD as if it was coming from above their heads. It is the one that more closely mimics what is usually provided in standard, 2D, audio described content, and such way of scripting is recommended. The second one, «Static», is heard from the left or right side, as if someone was standing beside the user. As the AD is heard as if coming from a short distance from the user's location, it is recommended to write the AD in a non-standard way, with a stronger involvement of the describer. The third AD type, «Dynamic» allows audio describers to place the descriptions at different angles of the sphere, and it can be used to locate a character, object or event. In other

words, when a user plays a video with this type of AD, they will hear the AD coming from a specific point in the space. We believe that this option can guide users effectively within this highly visual medium, and will also allow them to feel more present in the content. This is an innovation in the field of audio description, as it can provide additional information on the location of the visuals being described by means of audio cues. As far as the scripting is concerned, a minimum audio description is recommended to allow users to be guided by the sound of the video.

Figure 6 AD types

Although many of the AD web editor's features are similar to the features of others subtitling or audio description editors, there are some new features that needed to be implemented, taking into account the specificity of the environment. One of such functions is «Set angle», a function implemented within the «Dynamic» AD type. This function allows users to assign a given AD segment to a specific angle within the sphere, specified by latitude and longitude. It means that the user, when watching the content with AD, will hear an AD of a given object, or event from the direction set by an audio describer, as it stays 'tied' to that part of the 360º sphere. This function, which can help users of access services to orient themselves inside the sphere, was implemented based on requirements defined in the focus groups (Fidyka and Matamala, 2018a). When an audio describer wants to create a «Dynamic» AD type, they need to look inside the sphere for a desired angle, and set it by clicking on a special button (Figure 7), or by a combination of keyboard shortcuts (customizable in the future version of the editor).

This new option will change the current workflow of producing AD. So far, audio describers needed to write AD instances between the dialogues, and mark their timecodes. Those audio describers whose task was also to voice the AD, needed to record it and preview it. When producing AD with a «Dynamic» AD type, besides following all the aforementioned steps, audio describers need to set an angle for every AD instance. In other words, they need to decide from which angle the linear AD will be heard by the end users.

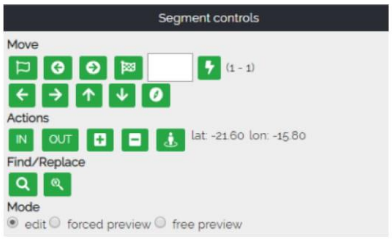


Figure 7 Set angle function and preview modes

As far as the preview of AD is concerned, two modes can be used for verification in the «Dynamic» AD type, namely «Forced preview» and «Free preview». In the first verification mode, AD and angle are bound to the video, which means that the video changes the angle during playback in the AD web editor. In other words, audio describers do not need to navigate through the video to find the angles they have set, as the software does it automatically. Thanks to this preview mode, audio describers are able to check whether the angles they set are correct. In the «Free preview», however, only the AD segments are bound to the timecodes. It means that audio describers can freely navigate the video during the playback, as during the editing.

As the functionalities described above are new, further testing is required to ascertain that they meet the needs of actual audio describers. This is why usability tests were conducted on the «Dynamic» AD type, in which they are implemented. The overall goal of our study was to assess the progress made in the development of the audio description editor. In particular, we wanted to learn whether the tool meets the needs of its actual users, so that its future version can cater for them more effectively. Additionally, our aim was to gather feedback on the work of the describer in relation to immersive environments and learn how such content will impact on their workflow.

2. METHOD

The usability test of the AD editor was performed online from the 24th of September to the 19th of October 2018. It aimed at receiving feedback from audio describers from different countries. The test instructions were given in English, and participants were asked to fill in the questionnaires in the same language. They were, however, requested to provide audio description in their native languages, which allowed us to gather feedback from users from

diverse countries, with different AD traditions. The response rate was 70.59%. The study was conducted voluntarily by professionals in accordance with ethical procedures approved by the Ethics Committee at Universitat Autònoma de Barcelona (UAB). Participants were informed about the aim and context of the study, and gave their consent before the test. Data are confidential, and the privacy of participants is ensured.

2.1. Participants

24 participants completed the test, 15 females and 8 males, plus one participant who preferred not to reply to this question. Their ages ranged from 25 to 64 years. The mean age was 36.71 years old. Their main languages were Catalan (1 respondent), Spanish (6 respondents), both languages (2 respondents), Bosnian (1 respondent), English (6 respondents), Dutch (2 respondents), Polish (3 respondents), German (2 respondents), and Swedish (1 respondent). They were mainly AVT translators, audio describers, AD and Media Accessibility supervisors, researchers and project managers. 21 participants had completed university studies, one participant had further education, and two participants, had secondary education.

2.2. Materials

Measures included usability and preferences. To obtain demographic data, two online questionnaires were prepared: a pre-questionnaire, and a post-questionnaire. They were sent to participants through separate e-mails upon agreeing to participation, as explained in section 2.3.

The post-questionnaire consisted of two parts: the System Usability Scale (SUS) questionnaire (Sauro & Lewis, 2016) and a preference questionnaire, with both closed and open questions which aimed at gathering additional user feedback. The SUS questionnaire chosen as a measure of usability contains 10 items, each with five steps: from strongly disagree to strongly agree (Sauro & Lewis, 2016, p. 198). It was available in English.

Regarding preferences, a specific questionnaire containing nine items was developed. It asked what participants appreciated or disliked about the editor, the possible improvements and missing functionalities, the level of difficulty related to the «Set angle» function, and the usefulness of preview modes. It also gathered the opinions of participants regarding whether or not describing 360º content is more time-consuming than that of standard content, and whether 360º videos will impact on their work in the coming years. The last question provided space for additional comments.

Each participant was assigned with one clip to be audio described with a «Dynamic» AD type. The video chosen was an initial 1-minute excerpt of a

fictional story suitable for audio description: there was enough space to produce audio description between the music and other sounds and, most importantly, the story developed at different angles of the sphere, which allowed us to test the «Set angle» function. The video was played in low resolution (720s) to avoid overloading the server, and to make the audio describing task smoother.

Other materials used in the study included a document with detailed instructions on how to perform the tasks, and a step-by-step user guide with screenshots on how to use the web editor. They were provided to participants by e-mail, as discussed in the next section.

2.3. Procedure

Prior to the study, a small-scale pilot test was performed with two users to evaluate the experimental protocol to be used on a larger scale. The participants met the criteria for inclusion in the sample, as they were professional audio describers. However, no changes were made to the final methodology because no problems arose in the pilot test development.

The study was developed online, and consisted of the subsequent steps. Participants were contacted through social media and personal contacts. Upon agreeing to participation, they received separate e-mails from the researchers, outlining the broader context and the procedure of the study. These e-mails included: (1) access to the editor, (2) a link to instructions on how to complete the test, (2) a link to a user guide, (4) links to the online questionnaires. Participants were informed about the exact order in which they should complete the tasks and fill in the questionnaires.

Participants completed the study in the following order. First they gave their informed consent to take part in the study. Then, the information that their data would be kept confidential was provided by the researchers. They were then asked to fill in the demographic questionnaire, perform a series of tasks in the editor, and fill in the post-questionnaire.

Participants were asked to complete the following tasks: (1) log in to the editor and open the assigned video, (2) audio describe the video excerpt in the user's native language, (3) preview the video in forced preview mode, (4) preview the video in free preview mode, and (5) save the AD and go back to the main window. Task 2 mentioned above («Audio describe the video excerpt») consisted of a series of sub-tasks: (1) add AD instances with correct timecodes, (2) set the angle for each AD instance, (3) record the AD segments produced, (4) insert one AD segment between two existing ones, and (5) delete two AD segments.

Once the tasks were completed, participants were asked to fill in the post-questionnaire, consisting of SUS and preference questions. At the end

of the study, participants were thanked, and information on how to obtain feedback was provided.

3. RESULTS AND DISCUSSION

Below we present results obtained from pre- and post-questionnaire as well as their discussion.

3.1. Results from pre-questionnaire: user profile

The pre-questionnaire focused on gathering socio-demographic information. Regarding the previous experience of participants in the AD field, most of the participants had received training in AD (91.67%). Nine participants (37.5%) had produced more than 300 hours of AD content, four participants (16.67%) had produced between 151 and 300 hours of AD content, four participants (16.67%) had produced between 51 and 150 hours, and seven participants (29.2%) had produced less than 50 hours of audio description. However, as little as four participants had had previous experience in audio describing a 360º video.

When asked about the usage of the AD editors, results show that not all audio describers use specific editing tools on a daily basis, with 10 of the participants reporting using only text processors and video players when producing AD.

The pre-questionnaire also aimed at determining cyber potential of research participants. When asked which devices they use on a daily basis, almost all participants (23) confirmed using mobile phones (95.83%); 21 participants use laptops (87.5%); 15 participants use TVs (62.5%), 14 participants use PCs (58.33%), eight of them use tablets (33.33%), one uses HMD (4.17%), and one participant chose the option «Other» (4.17%).

Unsurprisingly, most of participants do not use VR on a daily basis. To establish how frequently participants use VR, we asked them: 'How often do you watch Virtual Reality content (for instance, 360º videos)?' The results show that most of the participants (21, 87.5%) have never watched such content on a smartphone plugged into a head-mounted display or in a head-mounted display (20, 87.5%). Only one participant declared occasionally consuming Virtual Reality content on their smartphone, one participant occasionally uses a tablet to consume VR content and, regarding PCs, two participants (8.33%) use this device occasionally.

The next question asked about the reasons behind not using VR content or using it only occasionally. In this question, six participants (25%)

replied that they are not interested, three participants (12.5%) replied that it is not accessible, 12 participants (50%) replied that they have not had the chance to use it, two participants (8.33%) chose the option «Other reasons», and one participant (4.17%) did not provide any answer to this question. One of the participants provided an additional comment: «I don't normally access this content, I thought there were just a few, although I was surprised when accessing the project.»

Research participants were also asked to state their level of agreement with the statement «I am interested in Virtual Reality content (such as 360º videos)». The results show that four participants strongly agree with the statement (16.67%), eight participants agree (33.33%), nine participants neither agree nor disagree (37.5%), one participant disagrees (4.17%), and two participants strongly disagree (8.33%). Finally, when asked if they own any device with which to access Virtual Reality content, 10 participants replied that they do not (41.67%), five replied that they do not know or prefer not to reply (20.83%), and eight replied that they do (33.33%).

3.2. Results on usability

Regarding the results on usability, the score obtained in the SUS questionnaire is 55.9, which is considered below average, with a score of 68 or more considered as average. The obtained score corresponds to the percentile rank of 19%, and when converted to the letter grades, the obtained mark is D (Sauro & Lewis, 2016, pp. 203–204). This shows that the prototype web AD tool still has a lot of potential for improvement, and makes user testing at this stage even more relevant, as users are contributing to the definition of requirements as the tool is being developed.

The second part of the post-questionnaire focused on gathering data on users' preferences, and the results will be discussed question by question in the following section.

3.3. Results from preference questions

As far as the first question is concerned («What did you like the most about the AD editor?»), two of the participants (8.33%) highlighted the fact that the whole process of producing AD takes place in one piece of software. Four (16.66%) of the comments pointed to the fact that the most important functions (video, AD segments and recording) are displayed on one page, which facilitates the production of AD. In this regard, one of the participants commented: «It is quite easy, it has shortcuts and everything is visible and easily accessible on one page». Nine participants (37.5%) commented positively on the interface and its layout. They used the following words to describe it: «very clear», «simple», and «easy to understand». One of the

participants (4.17%) also positively assessed that the software is available online.

In the second and third questions, we asked participants which functions are the least useful, and how they could be improved. Nine audio describers (37.5%) pointed to the second question («What did you like less about the AD editor?») the problems encountered in the recording and preview modules. Also in the third question («What do you think could be improved, and how?»), problems with the video and buttons which would freeze or play with delay, were mentioned by five describers (20.83%). One comment (4.17%) in the third question suggested that a better video quality would be needed in order to describe all details. In this case, this was due to the testing conditions, as a higher quality video could overload the editor, and slow down the AD production process. Additionally, five (20.83%) participants suggested in both questions that they would prefer a different, more intuitive configuration of the shortcuts – or that they would prefer to customize the shortcuts themselves. They reported that this change is essential, as using the shortcuts they are accustomed to would allow them to work more efficiently. As already mentioned, although the shortcuts were not customizable in the prototype version of the editor, audio describers will be able to adapt them to their needs in the final version. Importantly, one comment (4.17%) in the second question pointed out that it would be necessary to preview the produced AD not only in the web editor, but also in a head-mounted display. We also believe that this option would be worth implementing, as it would allow audio describers to preview their work in the way in which end users will be consuming the content.

The following question asked participants about other functionalities that could be implemented. In this regard, participants suggested many improvements that are technically feasible. Among them, a waveform that would indicate music and other sounds, was suggested by two participants (8.33%) as a way to time audio description more precisely. This question was put forward in one comment (4.17%) as follows: «Perhaps one [function] in which you could see the lines corresponding to sound (…) is very useful to ensure that AD does not interfere with dialogues, or other sounds in the film.» Regarding combining AD with the other sounds of the video, one of the participants suggested that more options for fading the audio of the video would be needed, as there are currently only three options which are possible: «none», «low» and «high fading». This comment seems particularly relevant, as the audio volume of the AD track is a critical aspect when ensuring that descriptions can be heard clearly throughout the video or film.

An important improvement was suggested by another participant, who considers that a synchrony between AD segments and video would be needed

to facilitate the work. It means that if someone clicks on an AD segment on the «Segment list», the video should move to its timecode.

Other suggested improvements focused on both standard functionalities, and functions specific to the medium under discussion. They included: moving back or forward 5–10 frames at a time, a map in which all the locations with ADs assigned to them would be located, being able to merge or separate segments, setting a minimum separation between AD segments and dialogues or other sounds, and seeing the actual reading speed of the AD. Another participant suggested that using earphones during recording would be needed.

Importantly, an option to export the script to a text file for a professional recording was suggested in one comment (4.17%), as not all audio describers record AD themselves. For example, in Poland, audio describers send their script to a professional responsible for recording (Chmiel & Mazur, 2014). Audio describers who do not record the AD themselves, mark parts of the dialogue in the script or sounds that proceed, or follow a given AD segment. They also indicate in brackets how the given instance should be read, for example slowly or rapidly. The need to mark the part of dialogue in the script intended for recording by a professional was put forward in the comment of one of the participants: «I am used to using sound and dialogue cues rather than 'In times' only. I find it helpful to see the dialogue cue that leads into a description».

Regarding the innovative «Set angle» function, the results suggest that most of the participants (75%) found it easy to use, despite one participant reporting technical problems with their laptop. Another response suggests that one participant would like to set the angle for only some events, and not for all AD segments: «(...) I would like to have more freedom. The tutorial tells me we need an angle for each segment. I would like to have an angle only for very important situations». Therefore, it would be worth exploring a combination of the so-called «Dynamic» and «Classic» audio descriptions. Moreover, it would be worth exploring other spatial sound possibilities that allow one not only to set the angle from which the audio description will be heard, but also a specific position, ideally linked to an object.

As far as the preview modes – crucial to check if the set angles are correct – are concerned, five participants (20.83%) did not encounter any problems while using them, e.g. «(...) one allows you to move, the other one makes you see your fixed angles», but for 11 participants (approximately 45.8%) one or both preview modes did not work properly, and delay and visualisation problems were reported in the prototype version. Five participants (20.83%) did not provide an answer to this question.

Beyond the specific questions linked to the editor, it was interesting to gather general feed-back on the experience of audio describing 360º content. When asked about whether it takes longer to describe 360º videos than standard content, most participants (79.2%) replied positively, adding that setting an angle for each AD segment needs additional time.

Although not entirely within the scope of this article, some participants addressed the question of content selection in this emerging medium, the question discussed in depth in Fidyka and Matamala (2018b). Similarly to the results of the focus groups organized to provide a first insight into this question, it was deemed challenging, as explained in one of the comments: «technically, I consider it easy to use, the problem is which angle is the most important to describe.» It was added in other comments that audio describers will need special guidance when deciding which elements are the most relevant and should be audio described. This comment is similar to the comments put forward in the focus group organized in Poland, during which participants suggested that cooperation with content creators should be sought when audio describing 360º videos (Ibid.). In this regard, guidelines that would provide audio describers with guidance on what to describe, when to describe it and how would prove particularly useful, but they have not yet been created.

The last question in the post-questionnaire asked participants whether they consider that 360º videos would impact on their AD practice in the coming years. In this regard, participants expressed varying opinions. 58.3% of the participants answered affirmatively to this question. One of the comments further explained that the application for this medium is vast, and as more content is created, it should be provided with access services. Another comment pointed to the fact that this medium changes the current approach for producing AD, which means that specific training, or guidelines should be offered to audio describers producing AD in this medium. As explained by one of the participants: «I wouldn't feel comfortable taking on a job like this without proper instructions from the client and/or relevant training». This opens an opportunity for training institutions, which should accept the challenge of training professionals not only in the traditional modes, but also in more innovative media.

Additionally, one of the participants provided a detailed report after completing the test. This participant's overall assessment was positive, but they suggested some technical changes. Most of the comments pointed to improvements that could be made in the standard functions of the prototype editor (such as including a more responsive video control bar, improving the navigation by timecode, moving between AD segments), but one comment focused on the functions specific to immersive environments: audio describers

producing AD in this medium would need a globe («Mercator map projection») on which all the set locations could be seen on one map. Moreover, echoing the opinions voiced in the focus group in Kraków, discussed in Fidyka and Matamala (2018b), this participant suggested that there should be the opportunity to read the audio description text via speech synthesis during the preview. Importantly, previewing in a full-screen mode should be possible.

All in all, the results of both parts of the questionnaire have provided valuable feed-back to improve the preliminary version of the tool, thanks to a user-centric methodology in which users are asked for feedback during the development process. Although quantitative data regarding usability in these stages generally falls short of what is expected from a market tool, it provides a benchmark for comparison in future releases, and contributes to qualitative feedback.

CONCLUSIONS

The usability test of the AD web editor, developed so as to respond to the characteristics of 360º content, allowed us not only to assess the progress made in the development of the tool, but also to better understand the habits and needs of professional audio describers.

The results of the usability test show that although many of the features of the AD web editor have been assessed positively by the respondents taking part in the study, there is still room for considerable improvement in order to meet the needs of professional users. Thanks to participants' suggestions, we found several aspects of the editor which could be improved. These include, for instance, customizing the shortcuts, previewing AD with speech synthesis or previewing it by means of HMD. The technical feasibility of some other proposals, such as a map on which AD segments could be seen, needs to be explored.

Additionally, the results confirm the findings of the previous studies conducted in the form of focus groups (Fidyka and Matamala, 2018a, 2018b): content selection in this media format is considered challenging by audio describers who need guidance on how to describe. Beyond the specific analysis of the tool, the test has shown some attitudes of current audio describers towards new immersive media: in this regard, they state that this new production system will be more time-consuming, and request specific guidelines and training.

The limitations of this study need to be acknowledged. One limitation is the sample size of 24 participants. The future usability study testing the final version of the tool could be conducted with a higher number of participants to

obtain more reliable results. Secondly, the fact that the 360º video used as a stimulus in the study was watched on the flat screens of laptops or PCs with which the audio describers worked might have had impact on the results. Although this was due to the fact that the study was conducted online to reach professional audio describers from different backgrounds, watching the stimulus first in the HMD would be beneficial for audio describers to better understand the presented content. Another limitation could be related to the audio describers' lack of previous experience with AD software, which could have an impact on the obtained results.

Regarding future studies, more research will be needed on the implementation of spatial audio in AD in this media format to better respond to the needs of the end users of the access services. Additionally, it should be explored how spatial sound could enhance user experience when used in audio subtitles (AST), which serve audiences who cannot access the written subtitles and cannot understand the language of the original (Braun & Orero, 2010; Reviers & Remael, 2015). Also, more studies would be needed on AD in more complex VR types. Finally, an additional set of empirical tests will need to be conducted with end-users in order to assess whether the access services produced cater to their needs, allowing them to understand, enjoy, orient themselves, and feel present in the 360º content.

ACKNOWLEDGEMENTS

ImAC has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement 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 Anita Fidyka'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., Matamala, A., & Orero, P. (2018). From Disabilities to Capabilities: Testing Subtitles in Immersive Environments with End Users. Hikma, 17, pp. 195–220. doi: 10.21071/hikma.v17i0.11167

Agulló, B., Matamala, A., & Orero, P. (In press). The Challenge of Subtitling for the Deaf and Hard-of-hearing in Immersive Environments: Results from a Focus Group. Journal of Specialised Translation, 32.

Allen, C., & Tucker, D. (2018). Immersive Content Formats for Future Audiences. Retrieved from Digital Catapult on February 14, 2019: https://assets.ctfassets.net/nubxhjiwc091/5WypbBxFEQqS8I6WaKUQgg/ca25eb800e9f88a740079dbca623ca2c/Immersive_Content_Formats_for_Future_Audiences.pdf

Braun, S., & Orero, P. (2010). Audio Description with Audio Subtitling: An Emergent Modality of Audiovisual Localization. Studies in Translatology, 18(3), pp. 173–188. doi: https://doi.org/10.1080/0907676x.2010.485687

Brown, A. (2017). User Testing Subtitles for 360º Content. Retrieved from BBC (Research & Development) on February 14, 2019: https://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr

Chmiel, A. & Mazur, I. (2014). Audiodeskrypcja [Audio description]. Poznań: Wydział Anglistyki UAM.

Dooley, K. (2017). Storytelling with Virtual Reality in 360-degrees: A New Screen Grammar. Studies in Australasian Cinema, 11(3), pp. 161–171. doi: 10.1080/17503175.2017.1387357

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

Fidyka, A., & Matamala, A. (2018b). Retelling Narrative in 360º Videos: Implications for Audio Description. Manuscript submitted for publication.

Fryer, L. (2016). An Introduction to Audio Description. A Practical Guide. London: Routledge.

Fryer, L., & Johnatan, F. (2013). Cinematic Language and the Description of Film: Keeping AD Users in the Frame. Perspectives: Studies in Translatology, 21(3), pp. 1–15. doi: 10.1080/0907676X.2012.693108

Gödde, M., Gabler, F., Siegmund, D., & Braun, A. (2018). Cinematic Narration in VR – Rethinking Film Conventions for 360 Degrees. Virtual, Augmented and Mixed Reality. 10th International Conference, VAMR 2018. (pp. 184–201). Las Vegas: Springer. doi: 10.1007/978-3-319-91584-5_15

Greco, G. M. (2016). On Accessibility as a Human Right, with an Application to Media Accessibility. In A. Matamala & P. Orero (Eds.), Researching Audio Description: New Approaches (pp. 11–34). London: Palgrave.

Greening, J. (2011). Do we Need 3D Audio Description Guidelines: Recommendations from Focus Group Study Report. Retrieved from (February 20, 2019): http://audiodescription.co.uk/uploads/general/3D_film_and_tv_report_2.pdf

Jerald, J. (2015). The VR Book: Human-centered Design for Virtual Reality. Sant Rafael, CA: Morgan & Claypool. doi: 10.1145/2792790.2792827

Johansson, M. (2019). VR for Your Ears: Dynamic 3D Audio Is Coming Soon. Retrieved from IEEE Spetrum on February 19, 2019: https://spectrum.ieee.org/consumer-electronics/audiovideo/vr-for-your-ears-dynamic-3d-audio-is-coming-soon

Jones, S. (2017). Disrupting the Narrative: Immersive Journalism in Virtual Reality. Journal of Media Practice, 18(2–3), pp. 171–185. doi: 10.1080/14682753.2017.1374677

López, M., Kearney, G., & Hofstädter, K. (2016). Enhancing Audio Description: Sound Design, Spatialisation and Accessibility in Film and Television. Proceedings of the Institute of Acoustics, 38(2). Retrieved from (December 16, 2018): http://enhancingaudiodescription.com/wp-content/uploads/2016/11/RS2016-paper-Lopez-et-al.pdf

Maszerowska, A., Matamala, A., & Orero, P. (Eds.) (2014). Audio Description: New Perspectives Illustrated. Amsterdam: John Benjamins.

Nielsen, L. T., Møller, M. B., Hartmeyer, S. D., Ljung, T., Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Missing the Point: An Exploration of how to Guide Users' Attention during Cinematic Virtual Reality. Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology (pp. 229–232). New York: ACM. doi: 10.1145/2993369.2993405

Portillo, C. (2018). Films Unseen: Approaching Sound Design Alternatives to Enhance Perception and Mental Imagery Models among Audio Description Audiences on Sci-fi Films. Paper presented at Accessibility in Film, Television and Interactive Media Conference. Retrieved from (December 17, 2018): http://enhancingaudiodescription.com/conference/

Reviers, N., & Remael, A. (2015). Recreating Multimodal Cohesion in Audio Description: A Case Study of Audio Subtitling in Dutch Multilingual Films. New Voices in Translation Studies, 13(1), pp. 50–78. Retrieved from (February 20, 2019): https://www.iatis.org/images/stories/publications/new-voices/Issue13-2015/Articles/Reviers_New_voices_PUBL.pdf

Rothe, S., Hußmann, H., & Allary, M. (2017). Diegetic Cues for Guiding the Viewer in the Cinematic Virtual Reality. Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. New York: ACM. doi: 10.1145/3139131.3143421

Rothe, S., Tran, K., & Hußmann, H. (2018). Dynamic Subtitles in Cinematic Virtual Reality. Paper presented at ACM TVX 2018. Seoul. doi: https://dl.acm.org/citation.cfm?id=3213556

Sauro, J., & Lewis, J. (2016). Quantifying the User Experience: Practical Statistics for User Research. Amsterdam: Morgan Kaufmann.

Snyder, J. (2008). Audio description: The Visual Made Verbal. In J. Díaz-Cintas (Eds.), The Didactics of Audiovisual (pp. 191–198). Amsterdam/Philadelphia: John Benjamins. doi: 10.1075/btl.77.18sny

Streit, M., & Victoria, J. M. (2017). Guiding Viewer's Attention in 360 video and VR through Audio. Paper presented at the 4th International Conference on Spatial Audio.

Suojanen, T., Koskinen, K., & Tuominen, T. (2015). User-centered translation. London: Routledge.

Syrett, H., Calvi, L., & van Gisbergen, M. (2017). The Oculus Rift Film Experience: A Case Study on Understanding Films in a Head-mounted Display. In R. Poppe, J.J. Meyer, R. Veltkamp, & M. Dastani (Eds.). Springer. doi: 10.1007/978-3-319-49616-0_19

Viswanathan, L. N., McDaniel, T., Krishna, S., & Panchanathan, S. (2010). Haptics in Audio Described Movies. IEEE International Symposium on Haptic Audio Visual Environments and Games, (pp. 1–2). Phoenix. doi: https://doi.org/10.1109/have.2010.5623958

APPENDICES:

Appendix 1: Pre-questionnaire

1. Sex:
    a. Female

   b. Male
   c. Other
   d. I prefer not to reply
2. Age:
3. Main language: (open question)
4. Please, describe your current job: (open question)
5. Have you ever audio described a 360º video? Yes / No
6. For how long have you been working in the field of audio description? (open question)
7. How many hours of audio description have you produced in your professional life?
   a. Less than 50 hours
   b. 51-150 hours
   c. 151-300 hours
   d. More than 300 hours
8. In what language or languages do you normally audio describe?
9. What software do you normally use?
10. Please indicate your level of studies.
   a. Primary education
   b. Secondary education
   c. Further education. Please specify (open field)
   d. University. Please specify (open field)
11. If you replied "Further education" or "University" in the previous question, please specify. (open question)
12. If you have received specific training on audio description, please indicate it here.
13. What devices do you use on a daily basis? Multiple replies are possible.
   a. TV
   b. PC
   c. Laptop
   d. Mobile phone
   e. Tablet
   f. HMD
   g. Other: (open field)
14. How often do you watch Virtual Reality content (for instance, 360º videos)?

	Never	Occasionally	At least once a month	At least once a week	Every day
On a smartphone					

On a tablet					
On a PC					
On a smartphone plugged to HMD					
In HMD					

15. If you have never used Virtual Reality content such as 360º videos or only occasionally, please indicate why. Multiple answers are possible.
    a. Because I am not interested.
    b. Because it is not accessible.
    c. Because I have not had the chance to use it.
    d. Other reasons. Please explain: (open field)
16. Please state your level of agreement with the following statement: "I am interested in Virtual Reality content (such as 360º videos)."
    a. I strongly agree
    b. I agree
    c. Neither agree nor disagree
    d. Disagree
    e. Strongly disagree
17. Do you own any device to access Virtual Reality content?
    a. Yes (If yes, which one? _____)
    b. No
    c. I don't know or I don't want to reply
18. If you replied "yes" to the previous question, please specify which device(s). (open question)

Appendix 2: Post-questionnaire

System Usability Scale

1 – strongly disagree, 5 – strongly agree

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found that the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Preference questions

Now please reply to the following questions in your own words.

11. What did you like most about the AD editor?
12. What did you like less about the AD editor?
13. What do you think could be improved, and how?
14. Did you miss any functionality? If yes, can you tell us which?
15. Do you find the feature for setting the angle for the AD 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 audio describe videos in 360º? Why?
18. Do you think 360º videos will impact on your work as an audio describer?
19. Other comments: (open field)

Annex 1.2. Article 4

Fidyka, A., Matamala, A., Soler Vilageliu, O., & Arias-Badia, B. (2021). Audio description in 360° content: Results from a reception study. *Skase*, 14(1), 14–32.

Audio description in 360° content: results from a reception study

Anita Fidyka, Anna Matamala, Olga Soler Vilageliu, Blanca Arias-Badia

Abstract

The ImAc project was the first European initiative aiming to propose and test the model of implementing access services in 360° videos, paving the way for future studies in the under-researched field of immersive accessibility. This article reports on the methodology and results of a pilot study and a small-scale reception study, conducted in the last months of the project. The results show a favourable reception of extended audio descriptions by AD users. They also indicate interest in the implementation of spatial sound in AD provided for 360° content, which could be tested in future reception studies.

Keywords: *Audiovisual Translation, Media Accessibility, audio description, 360° videos, reception study, presence, extended AD.*

1. Introduction

The media landscape is being reshaped by new technologies and media formats, which are becoming more and more personalised and interactive (Allen and Tucker 2018). Likewise, new user-created content and interactive ways of storytelling, such as object-based media (Hanson 2019) are increasingly available. One emerging media format is 360° videos. They have become a new outlet for journalists, artists and videomakers, offering an interactive way of conveying the story to the viewer. They belong to Virtual Reality (VR), characterised by a high level of immersiveness (Slater and Usoh 1993: 221). Users can access 360° content, sometimes referred to as omnidirectional or spherical videos by means of headsets called head-mounted displays, as well as on smartphones equipped with gyroscopes, PCs, and tablets. 360° videos include both videos with six degrees of freedom and videos with three degrees of freedom (Fidyka and Matamala 2018), the latter being the focus of this article. 360° videos with three degrees of freedom allow users for certain, yet not complete, interactivity; when watching 360° content with three degrees of freedom, users stand in one physical position, triggering images with their head movements, but they are constrained to a fixed viewpoint. In other words, users remain at the centre of the action and have an impression of being surrounded by the storyworld. As they are immersed in the story, the concept of presence, defined as the "perceptual illusion of non-mediation" (Lombard and Ditton 1997: 9) is essential in this media format and it serves as a quality metric employed to evaluate virtual environment content (Lessiter et al. 2001: 282).

In line with an effective legal framework (CRPD 2006) and European directives, namely the Audiovisual Media Services Directive (2010/13/EU) and the European Accessibility Act (EAA), audiovisual media products should be made accessible for all European citizens. In recent years, audio description (AD), a translation form that conveys the visual code of audiovisual productions in words (Braun 2008: 14), has been researched extensively in various European training institutions (Reviers 2016). Nevertheless, with the emergence of immersive content, the need to propose and test a model for implementing AD in 360° content has arisen. Such model should ensure that the viewing experience of AD

users is more interactive than the experience of watching regular content on TV or in the cinema. In other words, as presence is crucial to a satisfactory user experience in this media format, AD should not only grant access to the visual content, but it should also engage its users on a more immersive level.

The first model of implementing AD in immersive environments was proposed within the EU-funded ImAc project. The project started in 2017, together with the early adoption of 360° videos by European broadcasters (EBU 2017: 9). As the project followed a user-centric approach, in its early stages a series of focus groups was organised in order to involve AD users and learn about their needs and preferences. The results of these qualitative studies, discussed in Fidyka and Matamala (2018), show the interest of the actual users in the integration of spatial sound, an audio technology already researched in the AD field (López, Kearney and Hofstädter 2016; Portillo 2018) and the elements of interaction in AD in this media format.

Based on the obtained feedback, a pilot reception study was designed and carried out in the next stages of the project, testing different presentation modes of spatial sound (see section 2.3). However, as the results from the pilot were inconclusive (see section 2.5), the methodology for the actual reception study was reconsidered, testing both a non-standard approach to AD scripting and extended audio description, which offers users a possibility of interaction (see section 3.2).

In order to evaluate the experience of participants when consuming audio described 360º content, presence measures were used in both studies, as they proved effective for assessing the experience of AD users (Fryer and Freeman 2012b; Walczak and Fryer 2017). What follows is a brief summary of the methodology and results of the pilot study (section 2) and the actual reception study (section 3). Finally, suggestions for further research are discussed in section 4.

2. Pilot study

This section discusses the methodology and results of the pilot study, conducted in Barcelona between 11–13 June 2019 in the form of individual testing. The aim of this study was to test the implementation of spatial sound in AD produced for 360° videos and to test the methodology with a reduced sample before the main study.

2.1. Participants

Six participants aged between 23 and 34 (2 blind, 4 partially-sighted) took part in the preliminary test. Only one participant was blind from birth (1) and other participants reported the beginning of their sight loss between the ages of 0-4 (1 participant), 5-12 (1 participant), and 13-20 (3 participants). All participants were frequent users of technological devices, such as smartphones, laptops and tablets, but only two participants reported watching 360º content occasionally on a smartphone and one participant (16.67%) by means of a head-mounted display. When asked about the reasons behind it, two participants pointed to the novelty of this media format (“I have not had the chance to use it”), two participants to the lack of access services (“It is not accessible”), and the remaining two to the lack of interest in immersive content (“I am not interested”). Similarly, none of the participants reported having a device on which to watch immersive content. All participants were familiar with AD and

50% of participants reported using this access service daily (two participants for 2-3 hours a day and one participant for less than 1 hour). Regarding accessing online content, three participants reported using screen readers, one participant using magnifiers, one participant both devices, and another one none of these tools.

2.2. Measures

Three online questionnaires were developed for this study in order to measure presence and preferences. These measures were chosen as presence and spatial realism are two important goals in the field of spatial audio research (Herre et al. 2015: 770). Presence was measured by means of the Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, Regenbrecht 2001), which includes 14 items on a 7-point scale. It consists of four components: (1) spatial presence – the sense of being physically present in a virtual environment, (2) involvement – attention devoted to the virtual environment, (3) experienced realism, defined as the subjective experience of realism in a virtual environment (Igroup n.d.; Regenbrecht and Schubert, 2002), and (4) a last component related to the general definition of the sense of presence: "I had a sense of being in the virtual environment" (Slater and Usoh 1993). IPQ was chosen as a measurement of user experience, as it has been used in previous studies on presence in virtual environments (Regenbrecht and Schubert 2002; Brown et al. 2003; Krijn et al. 2004; Hartanto et al. 2014; Kinateder et al. 2015), and it is recommended as a measure of presence because of its high reliability (Schwind et al. 2019).

The preference questionnaire, administered at the end of the study, included four questions. It asked participants to (1) rank the AD modes in order of preference, (2) explain the reasons behind their choice, and (3) suggest ways of improving AD. The last question provided space for additional comments. Ethical approval for the study was given by the Universitat Autònoma de Barcelona (UAB). Consent forms and coded questionnaires will be securely stored at the Universitat Autònoma de Barcelona for three years after the completion of the project.

2.3. Materials

Because of the novelty of this media format, one challenge related to this study was the limited availability of 360° videos that would meet testing requirements. Three initial episodes of the series "Holy Land" by Jaunt Ryot were chosen as a stimuli, as they were stand-alone narrative pieces, comparable in length. In this travel documentary, viewers are transported to various cultural sites in Israel, guided by the main narrator. Each episode chosen provided enough time to insert AD within the constraints of dialogue and made testing spatial sound possible, as the action develops at various angles of the 360° scene. A Catalan voiced-over version was created for the test, using a professional voice talent.

Another challenge related to the selection of stimuli was related to their length, as although there is no recommendation for an ideal stimuli length when measuring presence, a duration of 10-15 minutes is recommended in gaming context (ITU 2018:12). The 360° videos currently available on the market are shorter, those with a linear narrative oscillating around 5-15 minutes (Allen and Tucker 2018; Agulló 2019). Because of this, the three complete episodes were chosen, as clips should have a duration of typical 360° videos to reach ecological validity (Bryman 2008: 48).

For each episode, three AD modes were created (Classic, Static, Dynamic), referred to for the test purposes as AD-C, AD-S and AD-D.

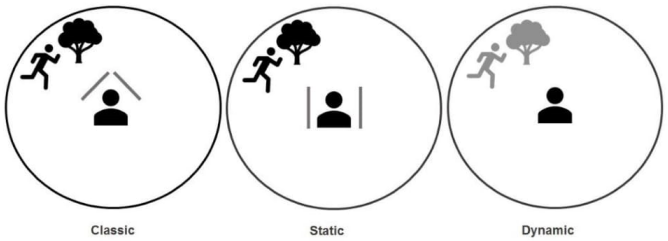


Figure 1 *AD presentation modes used in the preliminary study*

In the first presentation mode, the AD sound was placed above the user's head, while in the second presentation mode it was located on the user's side, as if someone was standing or sitting close to them, telling the story. In the Dynamic presentation mode, the AD sound was placed at different angles of the scene, depending on where the main action or other visual elements relevant to the plot were located. As the last presentation mode allows users to locate the events within the storyworld, our assumption was that it could guide viewers effectively within the storyworld and have a better viewing experience, which in turn would be reflected in higher presence scores.

The AD script for three episodes was originally written in English using the ImAc project editor, following existing AD guidelines (Ofcom 2000; Remael, Reviers and Vercauteren 2015), the reason being that the same test was conducted in the UK. This English AD was then translated into Catalan and voiced by a Catalan professional voice talent (female). An audio introduction was also created and voiced by a professional.

2.4. Procedure

The test was developed with a head-mounted display, and it was administered by the main researcher and a research assistant. Firstly, participants were welcomed, then presented with the ImAc project and the aim of the test. Secondly, they were assigned an individual participant's code and asked to sign informed consent forms. In the next step, participants replied to the questions in the demographic pre-questionnaire and listened to one general audio introduction (sound only). The actual AD test was comprised of: watching three video clips with randomised audio presentation modes, and replying to the IPQ questionnaire after each of them. At the end of the study, participants answered a preference questionnaire. All participants had the questionnaires read aloud and their responses were recorded by researchers on written online forms. For the purposes of this paper, participants' responses were translated into English.

2.5. Results

In the next sections, results from presence and preference questions will be presented, followed by a discussion.

2.5.1. Results from the IPQ questionnaire

Regarding presence, the median values for IPQ in the three conditions for each subscale are shown in Table 1 below:

	General presence	**Spatial presence**	**Involvement**	**Experienced realism**
Classic	3.00	3.40	3.00	1.75
Dynamic	4.00	3.70	3.00	2.13
Static	4.00	2.70	4.00	2.50

Table 1 *IPQ scores – preliminary test*

Non-parametric Friedman tests reveal no statistically significant differences between the scores of any subscale between conditions: general presence (Chi-Square(2)=.200; N=6; *p*=.905); spatial presence (Chi-Square(2)=.087, N=6, *p*=.957); involvement (Chi-Square(2)=1.810, N=6, *p*= .405); experienced realism (Chi-Square(2)=1.00, N=6, *p*=.607).

2.5.2. Results from the preference questionnaire

Regarding the preference questionnaire, the results from questions 1 and 2, which asked participants about their preferred sound option and the reasons behind their choice, show that participants based their choice on script characteristics, rather than audio presentation. This is confirmed in comments such as the following: "It is hard for me to distinguish the three videos. [...] The criterion has been the videos that I have enjoyed the most" (participant 1), "There were details that captivated me more" (participant 2), "AD of places was better, and you could hear the noise of the environment, streets, music" (participant 5), "I liked it more because it had more details than other [videos]. I have noticed things that I would not have noticed otherwise" (participant 6). Similarly, one answer in additional comments (participant 5) pointed to the difficulties in differentiating between the three sound modes: "There are no differences in the three videos, between the types of sound. You only notice changes in the content of the audio description which could be improved by adding more details."

In questions 3 and 4, participants made several suggestions on how to improve AD. Firstly, the following comments suggest that two participants prefer more detailed descriptions to create a more complete mental image of the storyworld:

(1) "[...] perhaps by creating a more specific description of the places. I understand that there may be not enough time to describe more things. But everything is very general. I missed more elements, a more specific way of explaining [...]" (participant 5).
(2) "By adding more. It gave me the feeling that there was little description, and at certain times I did not know what was happening on the screen" (participant 1).

Secondly, related to the question of how AD could be improved to allow participants to be more immersed in the story, two comments pointed to the interest in listening to the original music of the video or background sounds: “I would also like to hear more music or more ambient sound from the scene” (participant 2), “The ambient sound is very important” (participant 5).

Thirdly, some comments focused on the improvements which could be implemented to make 360° content more immersive. Although these comments are beyond the scope of this article, as they focus on storytelling techniques in 360° content, they can, however, serve as a recommendation for future content creators who wish to integrate access services already at the production stage. In this regard, participants suggested more slowly-paced content, with less sudden shifts of location, and more hearable ambient sounds.

2.6. Discussion

The results from the IPQ and preference questionnaires are inconclusive, as they demonstrate that participants were not able to clearly perceive the differences among the three audio treatments. Regarding the Dynamic presentation mode, none of the users noticed that the sound of AD was placed at different locations, depending on where the action took place.. The reasons for this could be content-related. It is possible that participants could not perceive the differences in the AD correctly because the original videos were not recorded in spatial sound. It is also possible that the AD instances in these episodes were too short and the differences in audio would be more perceivable in content with longer pauses between the dialogues. In spite of the inconclusive results as regards preference for audio treatments, qualitative feedback on users’ needs was gathered thanks to the adopted methodology. Based on participants’ preferences, the AD presentation modes were reconsidered for the main reception study.

3. Main study

The main reception study followed the same methodology as the pilot test. To respond to users’ preferences, a solution had to be found regarding the need of a more detailed AD. This posed a challenge; although the 360° storyworld is larger than standard content and it can contain more narratologically-relevant elements, AD is time-constrained, as it needs to fit in between the dialogues. Therefore, it was decided to test the Extended presentation mode which included additional descriptions, activated at the user’s will (ISO/IEC 20071−11, WCAG 2017).

Secondly, it was decided to test an unconventional approach to AD scripting in order to see if it could have a positive impact on AD users’ presence. Previous studies in the AD field have researched non-standard approaches to AD, including first-person AD (Fels et al. 2006), AD with elements of film language (Fryer and Freeman 2012a), or AD based on the production’s screenplay (Szarkowska 2013; Walczak 2017). Previous reception studies have shown that unconventional AD scripting can increase a sense of immersion in the presented story for persons with sight loss compared to standard AD (Walczak and Fryer 2017). However, this question has not yet been tested in relation to more immersive content.

3.1. Participants

30 participants took part in the main reception study, with ages ranging from 22 to 78. Most participants had a university degree at an undergraduate (20) or postgraduate (7) level. 18 participants defined themselves as partially-sighted and 12 participants as blind. 10 participants taking part in the study reported onset of sight loss from birth. All participants reported using mobile phones on a daily basis, followed by television (20), laptop (14), PC (12) and tablet (9), which suggests that participants are frequent users of technological devices. As far as immersive technologies are concerned, only one participant reported using a head-mounted display on a daily basis, and most of the participants had never watched Virtual Reality content before. Similarly, only two participants reported having a device to access VR content (PS4 and PlayStation VR). The most frequent reason behind not watching such content was not having had the chance to use it (20), followed by a lack of accessibility (5) or lack of interest (2). Regarding the level of interest in immersive content, most participants were strongly interested (12) or interested (11), followed by a neutral attitude ("neither agree nor disagree" – 6), and only one participant chose the option "strongly disagree". All participants were frequent AD users. Regarding the usage of assistive technologies, most participants reported using screen readers (13), two participants reported using magnifiers, eight participants chose the option "both" and seven chose the option "none".

3.2. Materials

The same comparable clips were used as in the pilot test, with new presentation modes, referred to as Classic (AD-C), Radio (AD-R) and Extended (AD-E). The original English script was written by RNIB (the Royal National Institute for the Blind) and the Catalan translation was prepared by the Universitat Autònoma de Barcelona. The text was rephrased and adapted, where necessary, to fit in the time slots between the dialogues. While the first presentation mode was the conventional AD, the second presentation mode followed an unconventional scripting style, which may be seen as a combination of the first-person narration introduced by Udo and Fels (2006) and audio drama (Fryer 2010). Audio description in Radio presentation mode was presented by a guide who accompanied the viewer, pointing to the most relevant visual elements of the storyworld. The viewer was addressed directly, in a conversational manner, and with a use of colloquialisms, nominal phrases, and discourse markers of casual speech (Table 2):

Classic	**Radio**
Una botiga ven articles de pell. En una altra, una dona mira collarets de granadura. [back translation] 'One shop sells leather goods. In another, a woman looks at beaded necklaces.' (ep. 1)	Practiqueu la cara de pòquer i el regateig abans d'entrar en aquest mercat. Prepareu-vos! [back translation] ' 'Practice your poker face and haggling skills before entering this market. Get ready!' (ep. 1)

Més pelegrins esperen en una llarga cua a l'entrada d'un santuari. [back translation] 'More pilgrims wait in a long line outside the entrance to a shrine.' (ep. 1)	Veniu d'hora si voleu calma i no haver de fer cua. L'any passat van venir aquí 4 milions de turistes, i cada cop en són més. [back translation] Get here early if you want to be calm and avoid the queues. Last year, 4 million tourists travelled here, and each year the number is going up. (ep. 1)

Table 2 *Classic and Radio scripting style*

The rationale for choosing this scripting approach was two-fold. Firstly, while the first presentation mode describes the events, characters, and surroundings following the existing AD guidelines, the second presentation mode positions the viewer as an observer inside the presented world, as shown in Table 3 (discussed elements in bold):

Una dona de mitjana edat amb un mocador blanc al cap i una faldilla llarga i grisa contempla les pintures religioses de les parets del passadís del costat. La gent va amunt i avall. Alguns estan asseguts, esperant el seu torn. Una dona puja pels esglaons de la gruta, que són molt alts, mentre un home s'agenolla i pressiona el front contra l'esglaó superior. Un altre home es descalça en senyal de respecte. [back translation] A middle-aged woman in a white headscarf and long grey skirt gazes up at the religious paintings on the walls in the adjoining corridor. People go up and down. Some sit on chairs to await their turn. A woman climbs the steep steps out of the grotto, which are very steep, while a man kneels and presses his forehead to the top step. Another man takes off his shoes as a mark of respect. (ep. 3)	Per entrar a la gruta, us heu d'ajupir i passar per una petita porta amb un nom ben adequat: Porta de la Humilitat. A dins, està molt decorat: llums brillants, frescos i cortines de vellut. De seguida s'omple; haureu de fer cua. Fora, Betlem és una ciutat animada, però no heu d'anar gaire lluny per recordar la història de Maria i Josep. Si ets el típic turista, hi ha molt per fer. Hi ha un antic basar ple de vida. [back translation] **To go into the grotto, you have to duck and go through a tiny door aptly called the Door of Humility.** Inside, it's very ornate: sparkling lanterns, frescoes, velvet curtains. But it gets busy here so you have to wait your turn. Outside, Bethlehem is a pulsing city, but you don't have to go far to be reminded of the story of Mary and Joseph. Still, there is plenty to do if you're a regular tourist. There's a lively Old bazaar. (ep. 3)

Table 3 *Classic and Radio scripting style*

Secondly, this approach includes sentences that convey the atmosphere of the presented scenes. Examples of such sentences include the following:

> Com he dit: increïble! Jerusalem és nova i antiga: edificis moderns, carreteres transitades, palmeres tropicals. Tot això més enllà dels merlets de la ciutat vella i la dona que toca l'arpa a la porta de Jaffa. Una estranya harmonia.
>
> [back translation] Like I said: unreal! Jerusalem is both old and new: modern buildings, busy roads, tropical palm trees. All this beyond battlements of the old city and the woman who plays the harp at Jaffa Gate. **There's a strange harmony**. (ep. 2)

> El monestir de Sant Jordi és una meravella! Penja entre penya-segats al desert de Judea. Us semblarà un lloc perfecte per trobar una mica de pau.
>
> [back translation] The monastery of Saint George is a marvel! It is perched on the cliff in the Judean desert. **It will seem like a perfect place to find some peace**. (ep. 3)

Regarding the Extended presentation mode, it consists of the main script written in the style of Radio presentation mode and additional descriptions. These descriptions can be optionally activated by the user after hearing a special bell sound, informing about the availability of an additional commentary. This option was suggested in focus groups organised at the beginning of the ImAc project (Fidyka and Matamala 2018). After hearing such an audio cue, participants have a 5-second interval to play the AD by clicking on the controller of the head-mounted display. Upon activation, the main video is paused and the extended track is played until the end. In non-testing conditions, the activation of the extended description is optional, and it can also be activated by voice interaction with devices such as Amazon Echo Dot. However, for testing purposes, participants were asked to activate every additional track with a controller. Additionally, in non-testing conditions, the return to the main AD would be possible earlier, without having to listen to the extended track until the end.

The extended descriptions were inserted whenever a new landscape or an architectural object relevant to the plot were introduced in the episodes, allowing for a description of the visual elements that would not be described otherwise due to time constraints. For example, the first extended track in the first episode included the description of the presenter, Naomi Darg, who was also present in the remaining two episodes. Including such detailed description that allows persons with sight loss to visualise the character would not be possible without pausing the main narration:

> Hola, gràcies per treure el cap darrere l'escena! Aquí és on us donem tota la informació addicional. Comencem per la protagonista. Naomi Darg, en pantalla, presenta Holy Land. Diria que té entre 30 i 40 anys, és de complexió normal i té els cabells arrissats. Li arriben fins a les espatlles. Com que som a Jerusalem i visitem llocs sagrats, duu roba còmoda i més aviat discreta, de colors neutres, adequada per al clima càlid: vestits llargs, faldilles, samarretes de màniga llarga i mocadors de cotó. A la Cúpula de la Roca, porta un mocador fosc que li cobreix el cap.
>
> [back translation] Hi, thanks for joining me behind the scenes! This is where we give you all the extra bits of information. Let's start with the on-screen talent. On screen, Holy Land is led by Naomi Darg. I'd say, she is in her 30s, average build, with curly hair that falls to her shoulders. Since we're in Jerusalem and visiting some of the Holy sites, Naomi dresses modestly in loose fitting clothes in neutral shades, appropriate for the warm weather – long dresses, skirts, tops with long sleeves and cotton wraps. At the Dome of the Rock, Naomi wears a dark head scarf. (ep. 1)

Similarly to the style of the main AD, all extended descriptions were written in a chatty and engaging way. Apart from the description of the visuals, most of them also provided information about historical context or cultural titbits:

> El llançador de flors, de l'artista Banksy, és un grafit d'un home amb mocador i gorra de beisbol que llança un ram de flors. L'home i l'embolcall del ram són en blanc i negre; les flors i les tiges que sobresurten de l'embolcall són de color. **El grafit**

> **d'aquest artista sigil·lós és només un dels molts que han convertit el mur de separació israelià en un ampli llenç. Per molts turistes, són l'atractiu principal de Betlem.**
>
> [back translation] "Flower Thrower", by the artist Bansky is a graffiti of a man wearing a kerchief and baseball cap throwing a bouquet of flowers. The man and the flower wrapping are in black and white; the flowers and the stems protruding from the wrapper are in colour. **The graffiti by this stealthy artist is just one of the many graffities that have turned the Israeli Separation Wall into a vast canvas. For many tourists, this is the main attraction of Bethlehem.** (ep. 3)

The AD was recorded by a professional studio with a male voice to differentiate the audio description track from the main narration read by a female voice. While the script of the Classic presentation mode was read with neutral intonation, AD in the Radio presentation mode was read aloud in a livelier and more engaging way. Additionally, some AD instances were read with non-standard prosodic features. For example, in a scene happening in a mosque filled with tourists, AD was read in a whispering voice. Furthermore, as sound effects often interact with dialogues and music to create a more vivid mental imagery in radio drama (Fryer 2010), AD was complemented by ambient sound effects. For example, when the story was taking place on narrow, winding streets of Jerusalem, AD was reinforced by ambient sound reminiscent of street noises and when the story moved to a windy desert, AD users could hear the blowing of the wind, which established a sense of place.

3.3. Procedure

Similarly to the preliminary study, an audio introduction was presented to participants before exposing them to the three episodes. Apart from familiarising participants with 360° videos, and providing a broader context to the series, the introduction included information on how to activate the extended audio description. Our assumption was that the Radio and Extended presentation modes would have a direct effect on the reception of the videos, as they give persons with sight loss a chance of a more engaging experience.

3.4. Results

The following subsections discuss the results obtained in the main reception study, both from the IPQ questionnaire and additional preference questions.

3.4.1. Results from the IPQ questionnaire

A paired samples t-test was used in order to compare the scores on each subscale across the different types of AD. None of these comparisons gave a statistically significant difference (all p>.05).

	General presence	**Spatial presence**	**Involvement**	**Experienced realism**
Classic	4.73	4.62	4.98	3.40
Radio	4.70	4.55	5.04	3.54
Extended	4.73	4.64	4.86	3.63

Table 4 *Median scores for all participants (N=30) in each subscale across AD type*

A one-way ANOVA comparing the results of the blind and partially-sighted users shows that there are significant differences between these type of users in their scores on IPQ. Planned comparisons show significant differences in General presence across all types of AD and that the scores are higher for partially-sighted users (see Table 5). In addition, for Classic AD, scores for spatial presence are also significantly higher for partially-sighted users. There is also a trend to significance for Radio and Extended presentation mode (p=.087). Table 5 shows statistics and p value for each comparison. Highlighted cells mark significant differences:

		General presence	**Spatial presence**	**Involvement**	**Experienced realism**
Classic	**Blind**	3.50	3.5500	4.6667	2.7292
	Partially-sighted	5.56	5.3333	5.1806	3.8472
	***p*=**	0.004	0.004	0.332	0.075
Radio	**Blind**	3.83	3.9167	4.8750	3.1458
	Partially-sighted	5.28	4.9778	5.1528	3.8056
	***p*=**	0.040	0.087	0.640	0.280
Extended	**Blind**	3.83	4.0333	4.5208	3.3750
	Partially-sighted	5.33	5.0444	5.0833	3.8056
	***p*=**	0.036	0.087	0.313	0.529

Table 5 *IPQ results from persons with sight loss*

A one-way ANOVA comparing the results of the blind and partially-sighted users shows that there are significant differences in general presence across all types of AD and that the scores are higher for partially-sighted users. In addition, for Classic AD, scores for spatial presence are also significantly higher for partially-sighted users. There is also a trend to significance for Radio and Extended AD (p=.087).

3.4.2. Results from the preference questionnaire

In the preference questionnaire, 12 participants indicated Extended presentation mode as their preferred option, 10 participants selected Classic presentation mode and the 8 remaining participants stated a preference for the clips with the Radio presentation mode. Two participants commented in open questions that they appreciated every description.

	AD-C (Classic)	**AD-R (Radio)**	**AD-E (Extended)**
1 preferred mode	10 (33.33%)	8 (26.67%)	12 (40%)
2 preferred mode	10 (33.33%)	10 (33.33%)	10 (33.33%)
3 preferred mode	10 (33.33%)	12 (40%)	8 (26.67%)

Table 6 *Results on preferences in the main study*

Upon analysing data for the blind and partially-sighted participants separately, the results are following:

	AD-C (Classic)	**AD-R (Radio)**	**AD-E (Extended)**
Blind persons			
1 preferred mode	1 (3.33%)	2 (6.67%)	9 (30%)
2 preferred mode	4 (13.33%)	6 (20%)	2 (6.67%)
3 preferred mode	7 (23.33%)	4 (13.33%)	1 (3.33%)
Persons with partial sight loss			
1 preferred mode	9 (30%)	6 (20%)	3 (10%)
2 preferred mode	6 (20%)	4 (13.33%)	8 (26.67%)
3 preferred mode	3 (10%)	8 (26.67%)	7 (23.33%)

Table 7 *Results on preferences in the main study*

The most frequent comments from participants who selected Classic presentation mode in the first place suggest that this scripting style allowed them to create a more complete mental representation of the storyworld (1), had a more appropriate level of detail than another scripting style (3), and was more coherent with the style of the main narration.

The most frequent reason behind choosing the Radio presentation mode as the preferred option was the integration of ambient sounds (2), which made the experience more realistic. One comment that focuses on the relation between ambient sounds and presence seems particularly relevant:

> I would like for the ambient sounds and the sound effects to be more in the first line than the explanations. I would like to have audio description more in the background and the sound effects more underlined to feel that the presented world is real. The explanations were fast with strong intonation and this way of explaining is a bit tense. I would like it to be more natural, relaxed.

Similarly, another participant who chose Radio presentation mode as the preferred option commented that this type of scripting, which combines verbal description of visuals with evocative sound effects, seems more appropriate for this innovative media format and added that Classic presentation mode does not make the viewing experience different from watching standard films. Another participant commented positively on the engaging use of language in this presentation mode. Participants who least preferred this presentation mode pointed to the following reasons behind their choice: this presentation mode resembles a separate narration rather than AD (2), the level of detail is not sufficient.

The results obtained in the first and second questions also show a favourable reception of the Extended presentation mode. Out of 12 blind participants who took part in the study, 9 blind participants selected Extended descriptions, including 2 congenitally blind participants**.** What participants appreciated the most was the possibility to listen more at will (9), the possibility to interact with the content (1), being provided with more details (5), and feeling more immersed (2):

> I was feeling more present in the virtual world in AD-E, as it had more information and a better quality of information that made me less conscious of the real world.

Two participants placed Extended presentation mode as their third choice. The reasons for this can be related to the testing conditions, as participants were asked to (1) activate all extended tracks and (2) listen to them until the end.

In the third question, participants proposed some improvements that could enable them to have a more immersive viewing experience when watching content in this media format. Chief among these are the possibility of being guided towards the described places, characters and objects (5 participants), and being able to listen to extended descriptions (7 participants). Interestingly, one participant commented that she/he would prefer to have extended descriptions in all three episodes watched during the study. Other suggestions include watching content with spatial sound (2 blind participants), integration of ambient sounds (2) and having the possibility to adjust the sound of AD independently from the video (1).

Regarding the improvements that could be made in the Extended presentation mode, the following comment suggests integrating music or other sound effects at the beginning and at the end of each extended description:

> The background of the extended audio description should convey the same background as the narration so that there is no interruption [upon activating extended AD]. I would like extended AD to be more integrated in the video. The interruption is very noticeable, extended AD starts abruptly. I would like extended AD to start with background music, then to have a description, and to finish with the same background music.

Among other improvements, two participants would prefer a more distinctive sound signalling of extended descriptions. Their comments suggest that the sound should be changed or first introduced in audio introduction in order to familiarise users with it. Additionally, one participant suggested that these additional tracks should contain only audio description and not titbits. This participant further specified that such information should be provided only in the video.

3.5. Discussion

The three presentation modes yielded similar levels of presence for all participants, which suggests that none of them can lead to a significantly higher immersion in the story, but it also shows that none of them expelled participants from their viewing experience. The differences in the presence scores between blind and partially-sighted persons obtained in two presence subscales suggest that blind persons need additional solutions in order to feel more present in the 360° storyworld. As two blind persons suggested the possibility of spatial sound to feel more immersed, this sound technology could be further tested.

The qualitative results from the preference questionnaire suggest the positive response of participants towards all presentation modes. Additionally, although it cannot be concluded from this study that combining verbal description with ambient sounds may stimulate presence, additional feedback suggests that persons with sight loss may appreciate this solution. As for Extended descriptions, one participant put forward an interesting comment in the second question:

> AD-E gives a little more details, it would be interesting for me to have the possibility to activate more extended descriptions of different points of the video. AD-R is like listening to an audiobook, like any documentary that I can watch on TV. At the AD level, it is poorer. The second (AD-C) gives more visual information and its details allow me to imagine (the storyworld).

In this test, although Extended descriptions were not linked to a given point within the video, it would be interesting to test this option with users in future studies. In such a case, extended descriptions would not be linear descriptions activated by a click, but instead participants would be able to stop the video and trigger different descriptions by head-movements.

4. Conclusions

The aim of this article was to discuss the first exploratory approach for implementing AD in 360° videos. It follows previous studies in the AVT field which measured the impact AD may have on the quality of end-user experience in 2D content (Fryer and Freeman 2012b, 2014; Wilken and Kruger 2016; Wissmath and Weibel 2012; Walczak and Fryer 2017), but it takes the previous research further, testing presence levels in a media format in which users have a degree of control over their experience.

The results of the main reception study show the positive response of participants toward extended audio description. They also show some possible improvements that could be made in order to better respond to their needs. There are several advantages related to integrating such solutions in this media format. Firstly, thanks to the possibility of pausing the main narration, users can be provided with more visual information than is present in the 360° storyworld and that could not fit in between the dialogues, or with additional titbits that are given in standard content in audio introductions. It also provides persons with sight loss with a possibility of interaction. This solution can also prove useful for describers, who deemed content selection in this media format challenging in the focus groups carried out within the project.

This is a small-scale, exploratory study and its limitations must be acknowledged. First and foremost, one limitation is related to the reduced sample. To obtain more reliable results, future studies should be conducted with more users. However, as stated by Orero et al. (2018), smaller sample sizes are acceptable when conducting research with persons with sight loss. Secondly, as only self-report measures were used in this study, data could be triangulated by using objective measures. Also, the present pilot study testing the implementation of spatial sound could be replicated with other video clips that contain more pauses between the dialogues and whose original sound is recorded in spatial audio technology. Additionally, more attention should be given to the presence measures in future studies. Existing self-report presence measures are designed for sighted persons, and they include statements such as “I felt like I was just perceiving pictures” (IPQ), which may not be appropriate for persons with sight loss. This is why validation of presence measures with persons with sight loss would be needed to ensure the validity of results when conducting research.

Several possible research avenues emerge from this study for future research into AD in 360° videos, or other types of immersive environments, that will increasingly emerge on the market in the upcoming years and reshape the landscape of Audiovisual Translation.

Firstly, taking into account the interest of users, the technology of spatial sound in AD in 360° content could be further explored, including the object-based sound (Simon, Torcoli and Paulus 2019), which allows sound technicians to place the sound exactly on the object. Secondly, the research on spatial sound could focus not only on AD, but also on audio subtitles (AST), as it could enable users with sight loss to locate the characters in 360° scene.

This article aims to contribute to an emerging line of research in AVT, and it is hoped that the preliminary results it presents will contribute towards a better understanding of AD users' needs in 360° content to ensure not only access to this type of content, but also a more captivating viewing experience.

Acknowledgements

We owe our warmest thanks to B1B2B3 Association for their support in finding volunteers for this study and to all the participants who took part in our experiment. We are also grateful to Ryot and Jaunt VR for letting us use "Holy Land" for our study.

ImAc has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement 761974. The authors are members of TransMedia Catalonia, an SGR research group funded bySecretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2017SGR113). Blanca Arias-Badia's participation in this research has been financially supported by the Spanish Research Agency, in the framework of the postdoctoral scholarship programme Juan de la Cierva Formación (FJCI-2017-32064). This article is part of Anita Fidyka'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, Belén. 2019. Nuevos retos para la TAV: subtitulación en medios inmersivos. Resultados de un análisis de corpus. Alicante: AIETI, 23-25 January 2019.

ALLEN, Catherine, TUCKER, Dan. 2018. *Immersive content formats for future audiences* [online]. Available at: <https://assets.ctfassets.net/nubxhjiwc091/5WypbBxFEQqS8I6WaKUQgg/ca25eb800e9f88a740079dbca623ca2c/Immersive_Content_Formats_for_Future_Audiences.pdf>.

BRAUN, Sabine. 2008. Audio description research: State of the art and beyond. In *Translation Studies in the New Millenium*, vol. 6, pp. 14-30.

BRYMAN, Alan. 2008. *Social Research Methods*. Oxford, New York: Oxford University Press.

BROWN, Sarah, LADEIRA, Ilda, WINTERBOTTON, Cara et al. 2003. The effects of mediation in a storytelling virtual environment. In *Lecture Notes in Computer Science*, vol. 2879, pp. 102-111. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.7163&rep=rep1&type=pdf>.

EBU (EUROPEAN BROADCASTING UNION). 2017. Virtual Reality: How are public broadcasters using it? Report [online]. Available at: <https://www.ebu.ch/publications/strategic/login_only/report/virtual-reality-how-are-public-broadcasters-using-it>.

EC (EUROPEAN COMMISSION). 2010. Audiovisual Media Services Directive (2010/13/EU). Available at: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0013>.

EC (EUROPEAN COMMISSION). 2019. The European Accessibility Act (ACC). Available at: <https://ec.europa.eu/social/main.jsp?catId=1202> .

FELS, Deborah, UDO, J. P., TING, P., Diamond, J. I. 2006. Odd Job Jack described - a first person narrative aproach to described video. In *Journal of Universal Access in the Information Society*, vol. 5, no. 3, pp. 73-88.

FIDYKA, Anita, MATAMALA, Anna. 2018. Audio description in 360° videos: Results from focus groups in Barcelona and Kraków. In *Translation Spaces,* vol 7, no. 2, pp. 285-303. doi: 10.1075/ts.18018.fid

FRYER, Louise. 2010. Audio description as audio drama - a practitioner's point of view. In *Perspectives: Studies in Translatology*, vol. 18, no. 3, pp. 205-213.

FRYER, Louise, FREEMAN, Jonathan. 2012a. Cinematic language and the description of film. Keeping AD users in the frame. In *Perspectives: Studies in Translatology*, vol. 21, no. 3, pp. 412-426.

FRYER, Louise, FREEMAN, Jonathan. 2012b. Presence in those with and without sight: implications for virtual reality and audio description. In *Journal of Cybertherapy and Rehabilitation*, vol. 5, no. 1, pp. 15-23.

FRYER, Louise, FREEMAN, Jonathan. 2014. Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In FELNHOFER, A. and KOTHGASSNER, O. D. (eds.), *Challenging presence: Proceedings of the 15th International Conference on Presence*, Wien: Facultas.wuv, pp. 99-107.

HANSON, Nick. 2019. Storytelling of the future. Available at: <https://www.bbc.co.uk/rd/blog/2019-02-storytelling-interactive-digital-drama>.

HARTANTO, Dwi, KAMPMANN, Isabel. L., MORINA et al. 2014. Controlling social stress in Virtual Reality Environments. In *PLoS ONE* [online], vol. 9, no. 3. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0092804>.

HERRE, Jürgen, HILPERT, Johannes, KUNTZ, Achim et al. 2015. MPEG-H 3D audio—The new standard for coding of immersive spatial audio, In *IEEE Journal Of Selected Topics In Signal Processing*, vol. 9, no. 5, pp. 770-779. Available at: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7056445>.

IGROUP. n.d. Igroup presence questionnaire (IPQ) scale construction. Available at: <http://www.igroup.org/pq/ipq/construction.php>.

ISO (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION). Guidance on audiodescriptions [online]. Available at: <https://www.iso.org/obp/ui/#iso:std:iso-iec:ts:20071:-21:ed-1:v1:en>.

ITU (INTERNATIONAL TELECOMMUNICATION UNION). 2018. Subjective evaluation methods for gaming quality (ITU-T P.809). Available at: <https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13626&lang=en>.

KINATEDER, Max, GROMER, Daniel, GAST, Philipp et al. 2015. The effect of dangerous goods transporters on hazard perception and evacuation behavior – A virtual reality experiment on tunnel emergencies. In *Fire Safety Journal*, vol. 78, pp. 24-30.

LESSITER, Jane, FREEMAN, Jonathan, KEOGH, Edmund, DAVIDOFF, Jules. 2001. A cross-media presence questionnaire: the ITC sense of presence inventory. In *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 282-297.

LOMBARD, Matthew, DITTON, Theresa. 1997. At the heart of it all: the concept of presence. In *Journal of Computer Mediated Communication*, vol. 3, no. 2, pp. 1-39.

LÓPEZ, Mariana, KEARNEY, Gavin, HOFSTÄDTER, Krisztián. 2016. Enhancing audio description: sound design, spatialisation and accessibility in film and television. Southhampton: Reproduced Sound Conference, 15-17 November 2016.

OFCOM (OFFICE OF COMMUNICATIONS). 2000. ITC Guidance on standards for audio description [online]. Available at: <http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf>.

ORERO, Pilar, DOHERTY, Stephen, KRUGER, Louis-Jan et al. 2018. Conducting experimental research in audiovisual translation (AVT). A position paper. In *JoSTrans*, no. 30, pp. 105-126.

PORTILLO, Cesar. Films unseen. 2018. York: Accessibility in Film, Television and Interactive Media Conference, 14-15 October 2018.

REGENBRECHT, Holger, SCHUBERT, Thomas. 2002. Real and illusory interaction enhance presence in virtual environments. In *Presence: Teleoperators and Virtual Environments*, vol. 11, no. 4, pp. 425-434.

REMAEL, Aline, REVIERS, Nina, VERCAUTEREN, Gert. 2015. *Pictures painted in words. ADLAB audio description guidelines* [online]. Trieste: Edizioni Università di Trieste. Available at: <http://www.adlabproject.eu/Docs/adlab%20book/index.html>.

REVIERS, Nina. 2016. Audio description services in Europe: an update. In *The journal of Specialized Translation*, no. 26, pp. 232-247. Available at: <https://www.jostrans.org/issue26/art_reviers.pdf>.

SCHUBERT, Thomas, FRIEDMANN, Frank, REGENBRECHT, Holger. 2001. The experience of presence: Factor analytic insights. In *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 266-281.

SCHWIND, Valentin, KNIERIM, Pascal, HAAS, Nico, HENZE, Niels. 2019. Using presence questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in*

Computing Systems, New York: ACM. Available at: <https://vali.de/wp-content/uploads/2019/01/2019-CHI-QuestionnairesInVR.pdf>.

SIMON, Christian, TORCOLI, Matteo, PAULUS, Jouni. 2019. MPEG-H audio for improving accessibility in broadcasting and streaming. Available at: <https://arxiv.org/abs/1909.11549>.

SLATER, Mel, USOH, Martin. 1993. Representation systems, perceptual position, and presence in immersive virtual environments. In *Presence*, vol. 2, no. 3, pp. 221-233.

SZARKOWSKA, Agnieszka. 2013. Auteur description: from the director's creative vision to audio description. In *Journal of Visual Impairment & Blindness*, vol. 107, no. 5, pp. 383-387.

UDO, John Patrick, FELS, Deborah. 2009. Suit the action to the word, the word to the action. An unconventional approach to describing Shakespeare's *Hamlet*. In *Journal of Visual Impairment and Blindness*, vol. 103, no. 3, p. 178-183.

UDO, John Patrick, FELS, Deborah. 2010. Horatio audio-describes Shakespeare's *Hamlet*. Blind and low-vision theatre-goers evaluate an unconventional audio description strategy. In *British Journal of Visual Impairment*, vol. 28, no. 2, pp. 139-156.

UN (UNITED NATIONS). 2006. Convention on the Rights of Persons with Disabilities (CRPD). Available at: <https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html>.

WALCZAK, Agnieszka. 2017. Creative description: audio describing artistic films for individuals with visual impairments. In *Journal of Visual Impairment & Blindness*, vol. 111, no. 4, pp. 387-391.

WALCZAK, Agnieszka, FRYER, Louise. 2017. Creative description: The impact of audio description style on presence in visually impaired audiences. In *British Journal of Visual Impairment*, vol. 35, no. 1, pp. 6-17.

WCAG (WEB CONTENT ACCESSIBILITY GUIDELINES). Extended audio description (prerecorded) [online]. Available at: <https://www.w3.org/WAI/WCAG21/Understanding/extended-audio-description-prerecorded.html>.

WILKEN, Nicole, KRUGER, Jan-Louis. 2016. Putting the audience in the picture: Mise-en-shot and psychological immersion in audio described film. In *Across Languages and Cultures*, vol. 17, no. 2, pp. 251-270.

WISSMATH, Bartholomäus, WEIBEL, David. 2012. Translating movies and the sensation of 'being there'. In PEREGO, E. (ed.) *Eye Tracking in Audiovisual Translation*, Rome: Aracne Editrice, pp. 281-298.

Anita Fidyka
Department of Translation, Interpreting and East Asian Studies
Universitat Autònoma de Barcelona
08193 Bellaterra, Spain
*e-mail: anita.fidyka@gmail.com**

Anna Matamala
Department of Translation, Interpreting and East Asian Studies
Universitat Autònoma de Barcelona
08193 Bellaterra, Spain
e-mail: anna.matamala@uab.cat

Olga Soler Vilageliu
Department of Basic, Developmental and Educational Psychology
Universitat Autònoma de Barcelona
08193 Bellaterra, Spain
e-mail: olga.soler@uab.cat

Blanca Arias-Badia
Department of Translation, Interpreting and East Asian Studies
Universitat Autònoma de Barcelona
08193 Bellaterra, Spain
e-mail: blanca.arias@uab.cat

* corresponding author

In SKASE Journal of Translation and Interpretation [online]. 2021, vol. 14, no. 1 [cit. 2021-04-26]. Available online at http://www.skase.sk/Volumes/JTI20/pdf_doc/02.pdf. ISSN 1336- 7811.

Annex 2: Ethical considerations

Annex 2.1. Information sheet

[English version]

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.

[Spanish version]

INFORMACIÓN

Proyecto: ImAc (Immersive Accessibility, en castellano, Accesibilidad Inmersiva)

Investigador principal: Sergi Fernández (i2Cat)

Asesora sobre aspectos éticos: 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 los experimentos, que se pueden presentar en diversos formatos (experimento con cuestionario, "*focus group*" o grupo focal, entrevistas, etc.), se le pedirá que proporcione datos demográficos. A continuación, se le pedirá que visualice un estímulo determinado, realice una tarea o exprese su opinión sobre diversos aspectos. Si es necesario, se registrarán datos objetivos durante la sesión. El/La investigador/a le proporcionará más detalles sobre el experimento específico en el que está participando y los métodos de recopilación de datos. Si tiene alguna duda sobre el procedimiento, puede realizar las preguntas que estime necesarias.

La realidad virtual puede provocar algunas molestias como mareos cuando se visualiza contenido en realidad virtual. Por tanto, se proporcionará información y se tomarán las medidas adecuadas para garantizar la seguridad y el bienestar de los participantes. Los entornos inmersivos no son recomendables para personas que sufran claustrofobia, enfermedades cardiacas, enfermedades dorsales, antecedentes de ataques epilépticos y/o sensibilidad a las luces parpadeantes. Asimismo, los participantes no serán admitidos si se considera que puedan estar inestables o bajo los efectos de las drogas o el alcohol.

En caso de utilizar algún aparato de recopilación de datos fisiológicos o movimientos oculares ("*eye-tracking*"), no sufrirá ningún tipo de molestia, ya que los aparatos que se utilizarán son de última generación y de ningún modo invasivos.

Proceda a la lectura del consentimiento informado.

[Catalan version]

INFORMACIÓ

Projecte: ImAc (Immersive Accessibility, en català, Accessibilitat Immersiva)

Investigador principal: Sergi Fernández (i2Cat)

Assessora sobre aspects ètics: Pilar Orero

L'objectiu d'aquests experiments és recopilar informació sobre com es poden implementar els serveis d'accessibilitat en mitjans immersius. Això permetrà identificar les necessitats dels diferents tipus de públic i investigar com es pot millorar la qualitat de l'experiència d'usuari i dels serveis oferts.

Durant els experiments, que es poden presentar en diversos formats (experiment amb qüestionari, "focus group" o discussió de grup, entrevistes, etc.), se li demanarà que proporcioni dades demogràfiques. A continuació, se li demanarà que visualitzi un estímul determinat, realitzi una tasca o expressi la seva opinió sobre diversos aspectes. Si és necessari, es registraran dades objectives durant la sessió. El/La investigador/a li proporcionarà més detalls sobre l'experiment específic en el qual està participant i els mètodes de recollida de dades. Si té cap dubte sobre el procediment, pot fer les preguntes que consideri necessàries.

La realitat virtual pot provocar molèsties com marejos quan es visualitza contingut en realitat virtual. Per tant, es proporcionarà informació i es prendran les mesures adequades per garantir la seguretat i el benestar dels participants. Els entorns immersius no són recomanables per a persones que tenen claustrofòbia, malalties cardíaques, malalties de l'esquena, antecedents d'atacs epilèptics i/o sensibilitat als flaixos. Així mateix, no s'admetran participants si es considera que poden estar inestables o sota els efectes de les drogues o l'alcohol.

En cas d'utilitzar algun aparell de recopilació de dades fisiològiques o moviments oculars, no tindrà cap tipus de molèstia, ja que els aparells que s'utilitzaran són d'última generació i de cap manera invasius.

Ara llegeixi, sisplau, el consentiment informat.

Annex 2.2. Consent forms

[English version]

CONSENT FORM (written version)

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 (NAME and SURNAME).

If you are willing to participate, please confirm the following statements by signing at the end of this document.

— I have read and understood the information given for this research or have had the information read to me.

— I have had the opportunity to ask questions about the research.

— I consent to take part in the research sessions.

Name of the participant Date Signature

Name of the researcher Date Signature

Signed by Pilar Orero (UAB IP ImAc)

[Spanish version]

CONSENTIMIENTO (versión escrita)

Proyecto: ImAC (Immersive Accessibility, en castellano, Accesibilidad Inmersiva)

Su participación en los experimentos es absolutamente voluntaria.

Puede interrumpir su participación en el estudio en cualquier momento sin necesidad de justificarlo previamente. En caso de hacerlo, no existirá ninguna repercusión o consecuencia negativa de ningún tipo.

En caso de utilizar algún aparato de recopilación de datos fisiológicos o movimientos oculares ("*eye-tracking*"), no sufrirá ningún tipo de molestia, ya que los aparatos que se utilizarán son de última generación y de ningún modo invasivos.

La realidad virtual puede provocar algunas molestias como mareos cuando se visualiza contenido en realidad virtual. Por tanto, se proporcionará información y se tomarán las medidas adecuadas para garantizar la seguridad y el bienestar de los participantes. Los entornos inmersivos no son recomendables para personas que sufran claustrofobia, enfermedades cardiacas, enfermedades dorsales, antecedentes de ataques epilépticos y/o sensibilidad a las luces parpadeantes. Asimismo, los participantes no serán admitidos si se considera que puedan estar inestables o bajo los efectos de las drogas o el alcohol.

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 aspectos 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/La investigador/a 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

Firmado por Pilar Orero (UAB IP ImAc)

[Catalan version]

CONSENTIMENT (versió escrita)

Projecte: ImAC (Immersive Accessibility, en català, Accessibilitat Immersiva)

La seva participació en els experiments és totalment voluntària.

Pot interrompre la seva participació en l'estudi a qualsevol moment sense necessitat de justificar-ho prèviament. En cas de fer-ho, no tindrà cap repercussió o conseqüència negativa de cap mena.

En cas d'utilitzar algun aparell de recopilació de dades fisiològiques o moviments oculars ("*eye-tracking*"), no tindrà cap tipus de molèstia, ja que els aparells que s'utilitzaran són d'última generació i de cap manera invasius.

La realitat virtual pot provocar molèsties com marejos quan es visualitza contingut en realitat virtual. Per tant, es proporcionarà informació i es prendran les mesures adequades per garantir la seguretat i el benestar dels participants. Els entorns immersius no són recomanables per a persones que sofreixin claustrofòbia, malalties cardíaques, malalties de l'esquena, antecedents d'atacs epilèptics i/o sensibilitat als flaixos. Així mateix, no s'admetran participants si es considera que poden estar inestables o sota els efectes de les drogues o l'alcohol.

Tota la informació que proporcioni s'utilitzarà en el projecte de forma anònima.

ImAc és un projecte europeu liderat per Sergi Fernández, de l'empresa i2Cat. Pilar Orero és l'assessora sobre aspectes ètics i és responsable dels procediments ètics. Pot contactar amb Pilar Orero enviant un correu a pilar.orero@uab.cat per demanar més informació sobre el projecte o els resultats.

L'/La investigador/a que durà a terme aquest experiment és_____ (nom i cognoms)

Si vol participar en l'estudi, signi al final del document mostrant la seva conformitat amb les següents afirmacions.

— He llegit i entès la informació proporcionada per a aquesta recerca o m'han llegit aquesta informació.
— He tingut la possibilitat de preguntar sobre la recerca.
— Dono el consentiment per formar part de les sessions de recerca.

Nom del participant Data Signatura

Nom de l'investigador Data Signatura

Signat per Pilar Orero (UAB IP ImAc)

Annex 3: Documents related to the focus groups

Annex 3.1. Information sheet and consent form in Polish

ARKUSZ INFORMACYJNY

Projekt: ImAc (Immersive Accessibility, w języku polskim Dostępność Immersyjna)

Kierownik projektu: Sergi Fernández (i2Cat)

Doradca ds. Etyki: Pilar Orero

Celem niniejszych badań jest uzyskanie opinii uczestników na temat sposobów wdrożenia usług dostępnościowych do mediów immersyjnych. Pozwoli nam to określić potrzeby różnych grup odbiorców oraz zbadać, jak można poprawić jakość doświadczenia użytkownika oraz jakość dostarczanych usług.

Podczas badania, które może zostać przeprowadzone w różnych formach (np. badania z wykorzystaniem kwestionariusza, w postaci grupy fokusowej, czy wywiadów pogłębionych), uczestnicy zostaną poproszeni o podanie potrzebnych danych demograficznych. Następnie uczestnicy zostaną poproszeni o zapoznanie się z krótkim materiałem video lub materiałem dźwiękowym, wykonanie zadania lub wyrażenie swojej opinii na temat różnych aspektów przedstawionego materiału. W razie potrzeby również obiektywne dane zostaną zarejestrowane podczas badania. Osoba przeprowadzająca badanie przedstawi Pani/Panu więcej szczegółów dotyczących konkretnego badania, w którym Pani/Pan weźmie udział oraz przedstawi metody zbierania danych użyte w badaniu. Osoba przeprowadzająca badanie odpowie na wszystkie pytania uczestników w celu pełnego wyjaśnienia procedury badania.

Zapoznawanie się z treściami rzeczywistości wirtualnej może być powodem pewnego rodzaju dyskomfortu, w tym choroby symulatorowej. Z tego powodu osoba przeprowadzająca badanie przedstawi Państwu wszelkie potrzebne informacje oraz wszelkie środki zostaną podjęte, aby zapewnić bezpieczeństwo oraz komfort uczestników. Korzystanie z środowisk immersyjnych nie jest zalecane osobom cierpiącym na klaustrofobię, choroby serca, problem z kręgosłupem, atakami drgawek lub epilepsji oraz/lub wrażliwość na migotanie światła. Do wzięcia udziału w badaniu nie zostaną także

dopuszczone osoby cierpiące na zaburzenia psychiczne, znajdujące się pod wpływem substancji odurzających lub alkoholu.

W przypadku wykorzystania w badaniach urządzeń służących do pomiarów fizjologicznych lub okulograficznych, uczestnicy nie odczują żadnego dyskomfortu, jako że urządzenia wykorzystane w badaniu są najnowszej generacji oraz są nieinwazyjne.

Prosimy o zapoznanie się z formularzem zgody.

FORMULARZ ZGODY (wersja pisemna)

Projekt: ImAC (Immersive Accessibility, w języku polskim Dostępność Imersyjna)

Udział w badaniach jest całkowicie dobrowolny.

Uczestnik może zadecydować o zakończeniu uczestniczenia w badaniu w dowolnym momencie oraz bez podania przyczyny odmowy. Nie będzie to wiązało się z żadnymi negatywnymi konsekwencjami.

W przypadku wykorzystania w badaniach urządzeń do pomiarów fizjologicznych lub okulograficznych uczestnik w żadnym stopniu nie odczuje związanego z tym dyskomfortu, jako że wykorzystywane w badaniach urządzenia są najnowszej generacji oraz są nieinwazyjne.

Zapoznawanie się z treściami rzeczywistości wirtualnej może być przyczyną pewnego rodzaju dyskomfortu, w tym choroby symulatorowej. Z tego powodu osoba prowadząca badanie przedstawi Państwu wszelkie potrzebne informacje oraz wszelkie środki zostaną podjęte, aby zapewnić bezpieczeństwo oraz komfort uczestników. Korzystanie z środowisk immersyjnych nie jest zalecane osobom cierpiącym na klaustrofobię, choroby serca, problemy z kręgosłupem, atakami drgawek lub epilepsji oraz/lub wrażliwość na migotanie światła. Do wzięcia udziału w badaniu nie zostaną także dopuszczone osoby cierpiące na zaburzenia psychiczne, znajdujące się pod wpływem substancji odurzających lub alkoholu.

Wszelkie informacje uzyskane w badaniu od uczestników zostaną wykorzystane w projekcie anonimowo.

ImAc to Europejski projekt kierowany przez Sergi Fernándeza z firmy i2Cat. Pilar Orero jest doradcą ds. Etyki, odpowiedzialnym za procedury etyczne. W razie chęci uzyskania dodatkowych informacji o projekcie i jego wynikach, prosimy o kontakt z Pilar Orero: pilar.orero@uab.cat

Badanie zostanie przeprowadzone przez Anitę Fidykę.

W celu wyrażenia zgody na udział w badaniu, prosimy o złożenie podpisu pod poniższymi oświadczeniami:

— Zapoznałam/em się z informacjami dotyczącymi badania lub informacje te zostały mi przeczytane.

— Miałam/em okazję zadać pytania na temat badania.

— Wyrażam zgodę na udział w badaniu.

Imię i nazwisko uczestnika Data Podpis

Imię i nazwisko badacza Data Podpis

Podpisano przez Pilar Orero (UAB IP ImAc)

Annex 3.2. Demographic questionnaire

[English version]

General questions

1. **Sex:**
    a) Female
    b) Male
    c) Other
    d) I prefer not to reply

2. **Age: _____**

3. **Main language: _____**

4. **Level of finished studies:**
    a) No studies
    b) Primary education
    c) Secondary education
    d) Further education
    e) University

5. **I define myself as a…**
    a) Blind person
    b) Low vision person
    c) Deaf person
    d) Hearing impaired person
    e) Deaf-blind person

6. **Age in which your disability began:**
   a) From birth
   b) 0–4
   c) 5–12
   d) 13–20
   e) 21–40
   f) 41–60
   g) More than 60

7. **What technology do you use on a daily basis? You can select more than one.**
   a) TV
   b) PC
   c) Laptop
   d) Mobile phone
   e) Tablet

8. **Do you have any device to access virtual reality content?**
   a) Yes (If yes, which one? _____)
   b) No
   c) I don’t know or I don’t want to reply

9. **Which of the following is your preferred device for watching online video content (i.e., Youtube, Vimeo, Netflix, Amazon Prime, broadcast catch up service etc.)? You can select more than one.**
   a) PC
   b) Laptop
   c) Smartphone
   d) Tablet
   e) I don’t watch online video content.
   f) Others (if so, please specify: _____)

10. **(Only for visually impaired users) Which of the following do you use on your connected devices to access the above content?**
    a) Magnification (i.e. Zoomtext)
    b) Screen readers (i.e., JAWS, VoiceOver, TalkBack)
    c) Both
    d) None

11. **(Only for visually impaired users) Which of the following controls would you like to use with your screen reader /magnification tool when watching content online?**
    a) Browse content library
    b) Identify content
    c) Functions such as play, stop, pause, forward, rewind
    d) Switch AD/AS on and off

[Catalan version]

Preguntes generals

1. **Sexe:**
   a) Dona
   b) Home
   c) Other
   d) I prefer not to reply

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

11. (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

[Spanish version]

Preguntas generals

1. **Sexo:**
    a) Mujer
    b) Hombre
    c) Otro
    d) Prefiero no responder

2. **Edad: _____**

3. **Lengua principal: _____**

4. **Nivel de estudios finalizados:**
    a) Sin estudios
    b) Primaria
    c) Secundaria
    d) Estudios superiores
    e) Universidad

5. **Me defino como una persona…**
    a) ciega
    b) con baja visión
    c) sorda
    d) con pérdida de audición
    e) sordociega

6. **Edad de inicio de la discapacidad:**
    a) De nacimiento
    b) 0–4 años
    c) 5–12 años
    d) 13–20 años
    e) 21–40 años
    f) 41–60 años
    g) Más de 60 años

7. **Qué tecnologías utiliza en su día a día? Puede seleccionar más de una opción:**
    a) TV
    b) Ordenador de mesa
    c) Ordenador portátil
    d) Teléfono móvil
    e) Tableta

8. **¿Dispone de 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 contestar

9. **¿Cuál es su dispositivo preferido para ver vídeos en línea (por ejemplo, en Youtube, Vimeo, Netflix, Amazon Prime, vídeos a la carta, etc.)?**
    a) Ordenador de mesa
    b) Ordenador portátil
    c) Teléfono móvil
    d) Tableta
    e) No accedo a contenido en línea
    f) Otros (especificar)

10. **(Solo para personas ciegas o con baja visión) ¿Qué usa para acceder a los contenidos en línea?**
    a) Magnificador (por ejemplo ZoomText)
    b) Lector de pantalla (por ejemplo JAWS, VoiceOver, TalkBack)
    c) Ambos
    d) Ninguno

11. **(Solo para personas ciegas o con baja visión) ¿Cuál de los siguientes controles le gustaría usar con su magnificador o lector de pantalla al ver contenidos en línea?**
    a) Explorar la biblioteca de contenidos
    b) Identificar contenido
    c) Funciones como reproducir (“play”), pausa (“stop”), avanzar, retroceder
    d) Activar y desactivar la audiodescripción y los audiosubtítulos

[Polish version]

Pytania ogólne

1. **Płeć:**
   a) Kobieta
   b) Mężczyzna
   c) Inna
   d) Wolę nie podawać informacji o płci

2. **Wiek: _____**

3. **Główny język: _____**

4. **Wykształcenie**
   a) Brak ukończonej szkoły
   b) Wykształcenie podstawowe
   c) Wykształcenie średnie
   d) Wykształcenie zawodowe
   e) Wykształcenie wyższe

5. **Określam siebie jako…**
   a) Osobę niewidomą
   b) Osobę niedowidzącą
   c) Osobę niesłyszącą
   d) Osobę niedosłyszącą
   e) Osobę z zaburzeniami widzenia i słuchu

6. **Wiek rozpoczęcia się zaburzeń wzroku lub słuchu:**
   a) Od urodzenia
   b) 0–4
   c) 5–12
   d) 13–20
   e) 21–40
   f) 41–60
   g) Ponad 60

7. **Z jakiej technologii korzysta Pan/Pani na co dzień? Można wybrać więcej niż jedną odpowiedź:**
   a) Telewizja
   b) Komputer stacjonarny
   c) Laptop
   d) Telefon komórkowy
   e) Tablet

8. **Czy posiada Pan/Pani urządzenie do odtwarzania treści rzeczywistości wirtualnej?**
   a) Tak (Jeżeli tak, proszę podać jakie)
   b) Nie
   c) Nie wiem lub nie chcę podać odpowiedzi na to pytanie

9. **Które z poniższych urządzeń jest Pani/a ulubionym podczas oglądania materiałów wideo online (np. Youtube, Vimao, Netflix, Amazon Prime, usługi VOD)? Można wybrać więcej niż jedną odpowiedź:**
   a) Komputer stacjonarny
   b) Laptop
   c) Smartfon
   d) Tablet
   e) Nie oglądam materiału wideo online
   f) Inne (proszę określić jakie)

10. (Pytanie dla niewidomych i niedowidzących użytkowników) Które z poniższych oprogramowań używa Pan/Pani na podłączonych do sieci urządzeniach, aby uzyskać dostęp do powyższych treści?

a) Programy służące do powiększenia ekranu (i.e. Zoomtext)
b) Czytniki ekranu (np. JAWS, VoiceOver, TalkBack)
c) Oba
d) Żadne z powyższych

11. (Pytanie dla niewidomych i niedowidzących użytkowników) Które z poniższych możliwości chciałaby Pani/chciałby Pan używać korzystając z czytnika ekranu lub programu służącego do powiększenia ekranu podczas oglądania treści online?

a) Przeglądanie biblioteki zawartości
b) Identyfikacja zawartości
c) Funkcje takie jak odtwarzanie, zatrzymywanie, pauza, przewijanie do przodu i przewijanie do tyłu
d) Włączenie i wyłączenie audiodeskrypcji oraz audionapisów

Annex 4. Documents related to the usability studies

Annex 4.1. Demographic questionnaire

Some questions about yourself

Please reply to these general questions about yourself.

Please enter your user code (provided in the e-mail):

1. **Sex: _____**
   a) Female
   b) Male
   c) Other
   d) Prefer not to reply

2. **Age: _____**

3. **Main language: _____**

4. **Please, describe your current job: _____**

5. **Have you ever described a 360° video?**
   a) Yes
   b) No

6. **For how long have you been working in the field of AD?**

7. **How many hours of audio description have you produced in your professional life?**
   a) Less than 50 hours
   b) 51–150 hours
   c) 151–300 hours
   d) More than 300 hours

8. **In what language or languages do you normally audio describe?**

9. **What software do you normally use?**

10. **Please, indicate your level of studies:**
    a) Primary education
    b) Secondary education
    c) Further education
    d) University

11. **If you replied 'Further education' or 'University' in the previous question, please specify: _____**

12. **If you have received specific training on audio description, please indicate it here: _____**

13. **What devices do you use on a daily basis? Multiple replies are possible:**
    a) TV
    b) PC
    c) Laptop
    d) Mobile phone
    e) Tablet
    f) Head-mounted display
    g) Other

14. How often do you watch virtual reality content (for instance, 360° videos)?

	Never	**Occassionally**	**At least once a month**	**At least once a week**	**Every day**
In a smartphone					
On a tablet					
On a PC					
In smartphone plugged to HMD					
In HMD					

15. If you have never used virtual reality content such as 360° videos or only occasionally, please indicate why. Multiple answers are possible:

a) Because I am not interested
b) Because it is not accessible
c) Because I have not had the chance to use it
d) Other:

16. Please state your level of agreement with the following statement: 'I am interested in virtual reality content (such as 360° videos):

a) I strongly agree
b) I agree
c) Neither agree or disagree
d) I disagree
e) I strongly disagree

17. Do you own any device to access virtual reality content?

a) Yes
b) No
c) I don't know or I don't want to reply

18. If you replied 'yes' to the previous question, please specify which device(s):

_____

Annex 4.2. Post-questionnaire

Please, provide some feedback about the editor.

Please score the following statements.

Please, enter your user code (provided in the e-mail): _____

(1 – strongly disagree, 5 – strongly agree)

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

More questions…

Please reply the open questions with your own words. The aim of these questions is to gather feedback to improve the AD editor.

1. What did you like the most about the AD editor?
2. What did you like less about the AD 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 AD 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 audio describe videos in 360°? Why?
8. Do you think 360° videos will impact your work as an audio describer?
9. Other comments:

Annex 4.3. Second iteration of the usability study

Participants' profile

24 participants completed the test (18 female – 75%, 5 male – 20.83%, 1 participant preferred not to reply to this question – 4.17%), with ages ranging 24–61. 7 participants (29.2%) completed the test in the previous version of the editor.

The main languages of the participants were Spanish (8), English (3), German (3) and Catalan (3). Their professions were mostly: researcher/professor, audio describer, PhD candidate, translator, project manager and accessibility consultant.

Only 2 participants (8.33%) taking part in the study audio described a 360º video before. When asked how long they have worked in the field of AD, participants provided varying replies: from 1 year to 19 years. More than half of participants produced less than 50 hours of AD in their professional life (13=54.17%) and 5 participants (20.83%) produced more than 300 hours of audio description. Participants provide AD in the following languages: Catalan and Spanish (6), English (4), Spanish (4), German (3), Russian (2), Slovak (2), Dutch (2) and Cantonese (1).

14 participants (58.33%) reported using a software when producing AD. 10 participants use text editors when producing AD (41.67%).

7 participants (29.17%) record their ADs and 13 participants (54.17%) do not record it themselves (4 participants – 16.67% ask other professionals to record their ADs).

22 participants (91.67%) had university level education (mostly in Translation Studies) and 2 participants (8.33%) reported having further education. 22 participants (91.67%) received a specific training on AD.

The responses of participants confirm their technological potential: 13 participants (54.17%) use TV, 17 participants (70.83%) PC, 21 participants (87.5%) laptop, 22 participants (91.67%) mobile phone, 10 participants (41.67%) tablet. None of participants uses a head-mounted display on a daily basis.

When asked about how often they watch virtual reality content, most of participants (22=91.67%) reported never watching VR content is a smartphone plugged to HMD or in

HMD. 9 participants (37.5%) consume VR content in smartphone at least once a month or occasionally. 4 participants (16.67%) consume VR content occasionally on a tablet and 7 on PC (29.17%).

When asked for reasons behind not using VR content or using it only occasionally, most of the participants (18=75%) replied that they have not had the chance to use it, 5 participants (20.83%) pointed to the lack of interest in such content and 1 participant (4.17%) pointed to lack of accessibility of such content.

When asked to state their level of agreement with the statement "I am interested in virtual reality content (such as 360º videos)", 2 participants (8.33%) strongly agree, 10 participants agree (41.67%), 9 participants neither agree or disagree (37.5%) and 3 participants disagree (12.5%).

Finally, when asked if they own any device to access VR content, only 2 participants (8.33%) provided an affirmative answer to this question, further specifying that they use Buro VR glasses and mobile phones/iPads.

Results on usability

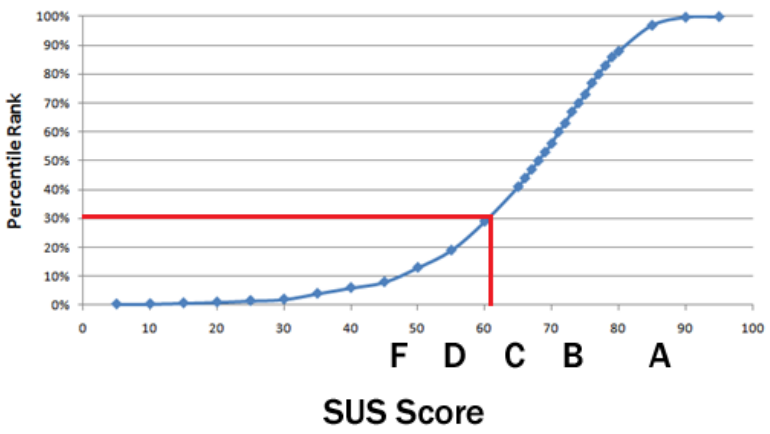
The SUS average score is **60.52.**

The graph below shows how the SUS scores associate with the percentile ranks and letter grades[8] and the red line specifies where the ImAc AD editor is at this moment.

The letter grade is D, and the obtained score corresponds to the percentile rank: 15–34%[9].

[8] Sauro, J. 2011. Measuring usability with the System Usability Scale (SUS). Retrieved from http://www.measuringu.com/sus.php

[9] Sauro, J. & Lewis, J. R. 2016. *Quantifying the user experience: Practical statistics for user research*. Amsterdam: Morgan Kaufmann, p. 203-204.



Feedback provided in the post-questionnaire

Responses obtained in the first question indicate that participants appreciated the most the following features of the editor: visual design and the layout of the user interface (P5, P10, P37, P4, P46, P8), the fact that the buttons for each function are well-placed and easy to identify (P10, P37, P52), being able to complete the whole process of producing AD in one single piece of software (P28, P31, P32, P33), being able to customize the shortcuts (P24, P52), the 'setting angle' option (P24, P33), sound wave (P15), speed thermometer (P2), the blue/yellow colours which help to see how the recorded AD fit in the video (P31).

Some participants accessed the usability of the editor positively in their comments, for example:

'The program was really easy to use' (P1);

'It's complete and accurate, and it allows writing and voice recording in an efficient way' (P13);

'Es sencillo de utilizar y muy práctico si quieres hacer el trabajo completo de escribir el guion y locutarlo (P36) [it is easy to use and very practical if you want to complete the whole process of writing the script and recording it];

'It is easy to use' (P34);

"I think it's quite intuitive to use which I liked very much" (P46).

Regarding the features/functions that the participants appreciated less and which could be improved, the following were mentioned in comments: the text of AD segments is not fully displayed and needed to be scrolled down for recording (P12, P15, P34); adjustable text and windows sizes (P10); when clicking on the timed segment, the video should jump to this exact time (P33). Other possible improvements could be related to: adding synthetic voice to voice the AD (P2), an option to visualize the video full screen (P34), improving the functions forward and go backward (P6, P37), an option to print the script (P10), a sounded count down (P12).

Regarding missing functionalities, participants would like to navigate using the segment list (P10) and the possibility to choose narration speed in 'characters per second' (not 'words per minute) (P28).

17 participants (70.83%) found the 'Setting angle' function easy to use.

Regarding the preview modes, 11 participants (45.83%) assessed the preview modes positively in their comments: (P5, P8, P10, P11, P13, P24, P31, P32, P33, P53, P36). Some participants did not use the preview options due to unordered TC values (P6, P34) or problems during recording: (P1, P4, P28, P37).

13 participants (54.17%) (P28, P53, P32, P49, P8, P25, P24, P4, P34, P12, P15, P1, P11) consider that it will take more time to describe 360º videos than standard videos because a describer needs to watch the whole scene when selecting content. 1 participant (P52) considers that it will depend of the video, but it will probably take more time to decide which visual elements are the most relevant. 3 participants (12.5%) (P5, P10, P33) consider that it will take slightly longer to write AD than in standard 2D because the angles need to be assigned after writing the scripts, but they also consider than it can be completed fast with practice. 4 participants (16.67%) (P13, P31, P6, P46) consider that it will not take necessarily longer to describe such videos once a describer is accustomed to the new video format and the software. 3 participants (12.5%) (P2, P36, P37) consider that the process of audio describing 360º videos will not take longer time than when providing AD to standard content.

8 participants (33.33%) (P32, P4, P24, P12, P10, P5, P2, P1) consider that 360º videos will impact their work as an audio describer, 4 participants (16.67%) (P15, P13, P36, P49) do not have an opinion on that question. 5 participants (20.83%) (P11, P37, P46, P8, P53)

consider that such content can have impact on their professional life once such videos become more common. 6 participants (25%) (P33, P34, P6, P25, P52, P31, P28) does not consider that this new format will have an impact on their professional life.

The problems related to recording functions (P34, P37, P46, P8) or displaying the video (P32, P34, P37, P28, P32, P52) may be due to the PCs used by participants taking part in the test, or to the lack of experience in recording AD (only 13 participants –54.17% reported recording AD in their practice).

For some participants (P1, P10, P49) several buttons (such as long test or short test) were not displayed correctly which may be due to performing the test with lower screen resolution than necessary. Although one participant (P13) deemed the editor "too complex", this response may be related to the profile of this audio describer, who is 61 years old, does not use any software for producing AD and never audio described a 360º video before.

4 participants (P12, P24, P49, P52) reported that it took them a while to get used to the program as I they had no previous experience of using it. Such factors as lack of familiarity with the new format of 360º videos and the new software, age (question 3 in the pre-questionnaire), usage of AD editors (question 10), not being accustomed to record AD (question 11), and the fact that most participants described the 360º video for the first time (question 6) may explain lower score in SUS questionnaire. This was expressed by several participants in the following comments:

'I would definitely need some technical training and previous practice in order to be able to use this editor for my work.' (P13);

I really like the editor. I wasn't comfortable recording the AD myself. I usually write the script and professional voice actors record it' (P5);

'As any new software you don't feel confident until you've used it a lot. I'd try to use it to be more familiar with it' (P28);

'A tutorial video would be great or a presential class' (P32).

Annex 5. Documents related to the reception studies

Annex 5.1. Demographic questionnaire

[English version]

Some questions about yourself

Please reply to these general questions about yourself.

Participant code: _____

1. **Sex: _____**
   a) Female
   b) Male
   c) Other
   d) I prefer not to reply

2. **Age: _____**

3. **Main language: _____**

4. **Please indicate your level of studies:**
   a) No studies
   b) Primary education
   c) Secundary education
   d) Further education
   e) University

5. **I define myself as…:**
   a) Blind person
   b) Partially-sighted person
   c) Other:

6. **Age at which you started losing your sight:**
   a) From birth
   b) 0–4
   c) 5–12
   d) 13–20
   e) 21–40
   f) 41–60
   g) More than 60

7. **What devices do you use on a daily basis? Multiple replies are possible:**
   a) TV
   b) PC
   c) Laptop
   d) Mobile phone
   e) Tablet
   f) Head-mounted Display (HMD)
   g) Other:

8. **How often do you watch virtual reality content (for instance, 360º videos)?**

	Never	**Occassionally**	**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 dedicated HMD (i.e. Vive, Oculus)					

9. If you have never used virtual reality content such as 360º videos or only occasionally, please indicate why. Multiple answers are possible:

a) Because I am not interested
b) Because it is not accessible
c) Because I have not had the chance to use it
d) Other:

10. Please state your level of agreement with the following statement: "I am interested in virtual reality content (such as 360º videos)."

a) I strongly agree
b) I agree
c) Neither agree nor disagree
d) Disagree
e) Strongly disagree

11. Do you own any device to access virtual reality content?

a) Yes
b) No
c) I don't know or I don't want to reply

12. If you replied "yes" to the previous question, please specify which device(s):

_____

13. Do you like watching the following types of content on television or online?

	I like it very much	**I like it**	**Neither like it nor dislike it**	**I don't like it**	**I don't like it at all**
News					
Fiction (series, films)					
Talk shows					
Documentaries					
Sports					
Cartoons					

14. How many hours a day do you watch audio described content?

a) None
b) Less than 1 hour
c) 1–2 hours
d) 2–3 hours
e) 3–4 hours
f) 4 hours or more

15. What do you use to access online content?

a) Magnifier (for example, ZoomText)
b) Screen reader (for example, JAWA, VoiceOver, TalkBack)
c) Both
d) None

[Catalan version]

Preguntes generals

Si us plau, respon a les següents preguntes.

Codi de participant: _____

1. **Sexe: _____**
   a) Dona
   b) Home
   c) Altre
   d) Prefereixo no responder

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) Altre:

6. **Edat en què vas començar a perdre la visió:**
   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 utilitzes diàriament? Pots seleccionar més d'una opció:**
   a) TV
   b) Ordinador de taula
   c) Ordinador portàtil
   d) Telèfon mòbil
   e) Tauleta
   f) Ulleres de realitat virtual
   g) Altre:

8. **Cada quan veus continguts de realitat virtual (per exemple, videos de 360°)?**

	Mai	**Ocasionalment**	**Almenys un cop al mes**	**Mínim una vegada a la setmana**	**Tots els dies**
Al telèfon mòbil					
En la tablet					
A l'ordinador					
Al telèfon mòbil connectat a unes ulleres de realitat virtual (tipus Google Cardboard o Samsung VR)					
En unes ulleres de realitat virtual (tipus Viu, Oculus)					

9. Si mai has vist contingut en realitat virtual com vídeos de 360° o només ho has fet en alguna ocasió, indica el motiu. Pots triar més d'una resposta:

a) Perquè no m'interessa
b) Perquè no és accessible
c) Perquè no he tingut l'oportunitat de fer-ho
d) Altre:

10. Si us plau, indica el teu nivell d'acord amb la següent afirmació: 'M'interessa el contingut de realitat virtual (com vídeos de 360°):

a) Molt d'acord
b) D'acord
c) Ni d'acord ni en desacord
d) En desacord
e) Molt en desacord

11. Tens algun dispositiu per accedir contingut de realitat virtual (per exemple, Google Cardboard, Samsung VR, Oculus Rift, PlayStation VR, etc.)?

a) Sí
b) No
c) No ho sé o no vull contestar

12. Si has respost "sí" a la pregunta anterior, especifica quin dispositiu(s): _____

13. Quin tipus de continguts t'agrada veure a la televisió o en línia?

	M'agrada molt	**M'agrada**	**Ni m'agrada ni em desagrada**	**No m'agrada**	**No m'agrada gens**
Notícies					
Ficció (sèries, pellícules)					
Programes de debat					
Documentals					
Esports					
Dibuixos animats					

14. Quantes hores al dia passes veient contingut audiodescrit?

a) Cap
b) Menys d'una hora
c) Entre 1 i 2 hores
d) Entre 2 i 3 hores
e) Entre 3 i 4 hores
f) 4 hores o més

15. Què fa server per accedir als continguts en línia?

a) Magnificador (per exemple, ZoomText)
b) Lector de pantalla (per exemple, JAWA, VoiceOver, TalkBack)
c) Tots dos
d) Cap

Annex 5.2. IPQ questionnaire

[English version]

1. In the computer generated world I had a sense of “being there”.
2. Somehow I felt that the virtual world surrounded me.
3. I felt like I was just perceiving pictures.
4. I did not feel present in the virtual space.
5. I had a sense of acting in the virtual space, rather than operating something from outside.
6. I felt present in the virtual space.
7. How aware were you of the real world surrounding while navigating in the virtual world (i.e. sounds, room temperature, other people, etc.)?
8. I was not aware of my real environment.
9. I still paid attention to the real environment.
10. I was completely captivated by the virtual world.
11. How real did the virtual world seem to you?
12. How much did your experience in the virtual environment seem consistent with your real world experience ?
13. How real did the virtual world seem to you?
14. The virtual world seemed more realistic than the real world.

[Catalan version]

1. En el món generat per ordinador, he tingut la sensació de “trobar-m’hi a dins”.
2. He sentit que en certa manera el món virtual m’envoltava.
3. He sentit com si només percebés fotografies.
4. No m’he sentit present en l’espai virtual.
5. He tingut la sensació d’estar dins l’espai virtual, en lloc de mirar-m’ho des de fora.
6. M’he sentit present a l’espai virtual
7. Fins a quin punt eres conscient del món real que t’envoltava quan navegaves pel món virtual (per exemple, sorolls, temperatura de la sala, altres persones, etc.)?
8. No era conscient de l’entorn real que m’envoltava.
9. He continuat parant atenció al món real que m’envoltava.
10. Estava totalment captivat pel món virtual.
11. Fins a quin punt t’ha semblat real, el món virtual?
12. Fins a quin punt l’experiència en el món virtual t’ha semblat comparable a l’experiència en el món real?
13. Fins a quin punt t’ha semblat real, el món virtual?
14. El món virtual m’ha semblat més realista que el món real.

Annex 5.3. Preference questionnaire in the pilot study

[English version]

Please, provide us some feedback on your experience with the videos and audio description.

Please, reply to the below questions with your own words.

Participant code: _____

1. **Which AD type do you prefer in 360º? Rank the AD modes according to your preferences: 1- preferred mode, 2- second prefered mode, 3 - the least prefered mode:**

	1	2	3
AD-C			
AD-S			
AD-D			

2. **Please explain why you ranked the AD modes in this way.**

3. **How could AD be improved in this medium?**

4. **Other comments: _____**

[Catalan version]

Si us plau, comenti la seva experiència amb els vídeos i l'audiodescripció.

Si us plau, respongui a les preguntes amb les seves paraules.

Codi de participant: _____

1. **Quin tipus de l'audiodescripció prefereix als vídeos de 360°? Ordeni'ls segons les seves preferències: 1 – el que més li agradi, 2 – el segon que més li agradi, 3 – el que menys li agradi:**

	1	2	3
AD-C			
AD-S			
AD-D			

2. **Expliqui per què ha ordenat les seves preferències d'aquesta manera.**

3. **Com es podria millorar l'àudiodescripció en aquest mitjà?**

4. **Altres comentaris: _____**

Annex 5.4. Preference questionnaire in the main study

[English version]

Please, provide us some feedback on your experience with the videos and audio description.

Please, reply to the below questions with your own words.

Participant code: _____

1. **Which AD type do you prefer in 360º? Rank the AD modes according to your preferences: 1- preferred mode, 2- second prefered mode, 3 - the least prefered mode:**

	1	2	3
AD-C			
AD-R			
AD-E			

2. **Please explain why you ranked the AD modes in this way: _____**
3. **How could AD be improved in this medium?**
4. **Other comments: _____**

[Catalan version]

Si us plau, comenti la seva experiència amb els vídeos i l'audiodescripció.

Si us plau, respongui a les preguntes amb les seves paraules.

Codi de participant: _____

5. **Quin tipus de l'audiodescripció prefereix als vídeos de 360°? Ordeni'ls segons les seves preferències: 1 – el que més li agradi, 2 – el segon que més li agradi, 3 – el que menys li agradi:**

	1	2	3
AD-C			
AD-R			
AD-E			

6. **Expliqui per què ha ordenat les seves preferències d'aquesta manera: _____**

7. **Com es podria millorar l'àudiodescripció en aquest mitjà?**

8. **Altres comentaris: _____**

Annex 5.5. Audio description scripts used in the main study

Audio introduction

Hola, soc l'Andrew Gold i això és Terra Santa.

Aquesta sèrie documental produïda per Jaunt Ryot consta de cinc parts. David Darg s'apropa als llocs sagrats de Jerusalem. Visitem algunes de les gestes més impressionants de la creació humana. Des del Mur Occidental fins a la Basílica de la Nativitat.

Al llarg de les dècades, aquí s'han rodat més documentals i reportatges que en qualsevol altre lloc, però el que fa especial aquest documental és que s'ha enregistrat amb una càmera de vídeo de 360 graus. Si és el primer cop que veus un vídeo de 360 graus, imagina't que ets al centre d'una esfera amb una pantalla al teu voltant. No et pots desplaçar dins d'aquesta esfera, però sí que pots girar-te, mirar cap amunt, cap avall o fer qualsevol altre moviment que vulguis, mentre et mantens fixat al centre de l'esfera.

Endavant, submergeix-te en el que passa al teu voltant.

Faràs servir unes ulleres de realitat virtual i uns auriculars per veure la sèrie. Un episodi té informació addicional sobre què passa al vídeo. L'investigador et dirà com activar aquesta informació.

Avui només veuràs els tres primers episodis. Cada un dura uns cinc minuts. Si vols veure els altres, estan disponibles de franc a l'aplicació Jaunt VR a Google Play, iTunes i GearVR.

Ara comença l'episodi 1, després d'una petita audiointroducció. Després, veuràs els episodis 2 i 3.

(AMBIENT SOUND: AFRICA_BOTSWANA_CICADES) Si busques les vistes més espectaculars de la ciutat vella, et recomano el mont de les Oliveres. Per descomptat, els bosquets que abans cobrien el vessant han desaparegut, i avui hi trobem principalment sorra seca i algunes àrees amb fullatge desèrtic. Arbusts, arbres petits. A l'estiu sol fer massa calor per venir aquí, però les tardes d'hivern, quan el sol encara és càlid i daurat, val molt la pena.

Quina imatge! La cúpula daurada de la roca il·lumina la pedra rosada característica de Jerusalem. Més lluny, els terrats groguencs i els carrerons estrets i sinuosos de la ciutat vella.

Audio description scripts

Classic presentation mode – Episode 1

— Títol, lletres blanques sobre fons negre: Jaunt Ryot.
— Una dona en una terrassa de pedra mira la Ciutat de Jerusalem.
— Títol: Terra Santa.
— La dona rossa du un mocador blanc al cap i encén una espelma en un santuari, dins d'una església imponent.
— Una capella coberta amb llum daurada.
— Més pelegrins esperen en una llarga cua a l'entrada d'un santuari.
— El santuari és al centre.
— L'envolten grans columnes de pedra que fan semblar nans els pelegrins. Fora hi ha un edifici de pedra rugosa.
— Una botiga ven articles de pell. En una altra, una dona mira collarets de granadura.
— La dona, amb mocador i faldilla llarga, és al costat d'un edifici decorat amb mosaic blau i blanc que té una cúpula daurada.
— Un tram d'esglaons condueix a l'entrada.
— Les teulades de Jerusalem.
— La dona camina per la teulada de la torre.
— Lluny, hi ha les ruïnes d'edificis construïts per civilitzacions anteriors.
— Els homes preguen a la muralla.
— La dona, dalt d'un camell guiat per un home, avança per un paisatge desèrtic.
— El camell s'allunya.
— Crèdits: Presentadora, Naomi Darg: Director, David Darg: Post producció Dirk Wallace, Darryl E. Chong, Travis Cook, Dan Dasho. Jaunt Ryot.

Classic presentation mode – Episode 2

— Títol, lletres blanques sobre fons negre: Jaunt Ryot. Una dona, a camell, es dirigeix cap a una ciutat desèrtica.
— Títol: Terra Santa.
— El camell avança pel terra sec i rocós.
— Una vista dels edificis de pedra clara de Jerusalem.
— La gent passeja vora les muralles cap a l'ombra d'una porta amb volta.
— L'ombra de la gent és allargada.
— Davant la muralla hi ha una plaça amb arbres ben plens de fulles.
— Els merlets de pedra rugosa envolten edificis moderns.
— Les carreteres, transitades, tenen palmeres a banda i banda.
— Una dona asseguda en un buit fet a la muralla toca l'arpa.
— Sostre de pedra, amb volta.
— La Porta té una gran entrada.
— La dona surt de la foscor cap a fora, on brilla el sol. En un altre mur, els homes preguen, amb la cara gairebé tocant els grans blocs de pedra que formen el mur.
— Crèdits: Presentadora, Naomi Darg: Director, David Darg. Post producció: Dirk Wallace, Darryl E. Chong, Travis Cook, Dan Dasho. Jaunt Ryot.

Classic presentation mode – Episode 3

— Títol, lletres blanques sobre fons negre: Jaunt Ryot. Un sol baix il·lumina els turons del desert.
— Títol: Terra Santa. Una dona, dalt d'un burro, es dirigeix a uns arcs de pedra clara al mig del desert rocós. Hi ha una creu negra a la part superior de l'arc central. Una porta hi bloqueja l'accés. El burro s'atura als arcs.
— Una carretera dalt d'un turn on hi ha molts edificis blancs d'altures diverses. La gespa del voltant és resseca.
— La dona és en un pati de lloses de color marró clar envoltat de cotxes aparcats.
— La dona se'n va. Darrere seu hi ha un edifici fet amb maons clars, amb petites finestres quadrades al pis superior.
— A l'interior, la llum d'una gruta il·lumina el rostre d'una multitud de pelegrins. La dona encén una espelma.
— Darrere seu, una bastida omple el centre de l'església.
— Un home s'arrossega sota l'altar per besar l'estrella.
— Una dona de mitjana edat amb un mocador blanc al cap i una faldilla llarga i grisa contempla les pintures religioses de les parets del passadís del costat.
— La gent va amunt i avall. Alguns estan asseguts, esperant el seu torn.
— Una dona puja pels esglaons de la gruta, que són molt alts, mentre un home s'agenolla i pressiona el front contra l'esglaó superior.
— Un altre home es descalça en senyal de respecte.
— Fora, un home serveix menjar d'una paella.
— Murs de formigó, coberts de pintades.
— Un mural de Banksy.
— Un home emmascarat llança unes flors.
— Naomi Darg. David Darg, post producción. Dirk Wallace, Darryl E. Chong, Travis Cook, Dan Dasho. Jaunt Ryot.

Radio presentation mode – Episode 1

— Aquí comencem el nostre viatge.
— Jerusalem, centre espiritual. Marcat pel conflicte. Un lloc on la història mai és un capítol tancat.
— (XIUXIUEJANT/ AMBIENT SOUND: MOSQUE_WITH_TOURISTS) Veniu d'hora si voleu calma i no haver de fer cua.
— (XIUXIUEJANT/AMBIENT SOUND: MOSQUE_WITH_TOURISTS) L'any passat van venir aquí 4 milions de turistes, i cada cop en són més.
— Aedicula vol dir "una caseta". Aquest edicle està envoltat per una columnata arquejada i columnes de pedra imponents.
— (AMBIENT SOUND: MARKET (FRUIT) Practiqueu la cara de pòquer i el regateig abans d'entrar en aquest mercat. Prepareu-vos!
— Hi ha files i files de botigues en aquests carrers empedrats. Obren cada dia i venen coses semblants: joies, roba, quincalla, antiguitats...
— Jo també me'n vaig. [AMBIENT SOUND: CAMEL 1 OR CAMEL 4] Sí, aquest és el meu camell. Anem a veure la posta de sol. [AMBIENT SOUND: CAMEL 1 OR CAMEL 4] Fins al pròxim episodi!

Radio presentation mode – Episode 2

— Hola, soc l'Andrew Gold, i això és el segon episodi dels documentals immersius Terra Santa.

— A l'Antic Testament, aquesta vall de la ciutat vella es coneix, almenys en part, com a Jardí del Rei. Per veure una posta de sol, veniu aquí. Si no voleu caminar, veniu amb camell!

— (AMBIENT SOUND: MARKET (FRUIT) Fa molt de temps, aquests murs eren de defensa. Avui, són una atracció turística més. Quan el sol està alt, com ara, els turistes hi passegen amb les guies de viatge a la mà: formen ombres llargues sobre el terra asfaltat. És increïble!

— (AMBIENT SOUND: MARKET (FRUIT) Com he dit: increïble! Jerusalem és nova i antiga: edificis moderns, carreteres transitades, palmeres tropicals. Tot això més enllà dels merlets de la ciutat vella i la dona que toca l'arpa a la porta de Jaffa. Una estranya harmonia.

— (AMBIENT SOUND: MARKET (FRUIT) No totes les portes són iguals: n'hi que hi passen carreteres. D'altres, són petites i mig amagades. Quan en surts i veus el sol, és com si deixessis el temps enrere.

— Jo també me'n vaig. Ens veiem a l'episodi següent.

Radio presentation mode – Episode 3

— Hola, soc l'Andrew Gold i estàs veient el tercer episodi de Terra Santa.
— (AMBIENT SOUND: TOURIST CROWD) El monestir de Sant Jordi és una meravella! Penja entre penya-segats al desert de Judea. Us semblarà un lloc perfecte per trobar una mica de pau. Més enllà es veuen desenes de blocs de pisos blancs de diferents altures.
— Betlem ja no és el "poblet" de les nadales. Els carrers plens de trànsit, els turistes amb les guies. Hi ha molt de moviment!
— (AMBIENT SOUND: TOURIST CROWD) És antiga, l'estructura de maó de color crema amb finestretes quadrades al pis superior. És història, ho notes.
— (AMBIENT SOUND: MOSQUE) Per entrar a la gruta, us heu d'ajupir i passar per una petita porta amb un nom ben adequat: Porta de la Humilitat. A dins, està molt decorat: llums brillants, frescos i cortines de vellut. De seguida s'omple; haureu de fer cua. Fora, Betlem és una ciutat animada, però no heu d'anar gaire lluny per recordar la història de Maria i Josep. Si ets el típic turista, hi ha molt per fer. Hi ha un antic basar ple de vida.
— (AMBIENT SOUND: BACKSTREET EGYPT) No us perdeu el llançador de flors de Bansky!
— (AMBIENT SOUND: DESERT WIND) Jo també me'n vaig. Fins aviat!

Extended presentation mode – Episode 1

Extended track 1

Hola, gràcies per treure el cap darrere l'escena! Aquí és on us donem tota la informació addicional. Comencem per la protagonista. Naomi Darg, en pantalla, presenta Holy Land. Diria que té entre 30 i 40 anys, és de complexió normal i té els cabells arrissats. Li arriben fins a les espatlles. Com que som a Jerusalem i visitem llocs sagrats, duu roba còmoda i més aviat discreta, de colors neutres, adequada per al clima càlid: vestits llargs, faldilles, samarretes de màniga llarga i mocadors de cotó. A la Cúpula de la Roca, porta un mocador fosc que li cobreix el cap. (PAUSA) a l'escena inicial, quan la veiem en una plaça elevada amb vistes a la façana de pedra calcària de les estructures antigues, duu una túnica de color verd oliva i pantalons negres.

Extended track 2

Hola, gràcies per treure el cap darrere l'escena! Ara que hi ha tant de moviment, per què no parem un moment? Orientem-nos abans de continuar. Us hem preparat algunes indicacions sobre la ciutat vella, unes orientacions bàsiques sobre la ruta que seguirem en aquest episodi.

A l'entrada de la basílica del Sant Sepulcre hi ha quatre arcs magnífics, amb unes llindes decorades amb creus de Jerusalem. Molts cristians creuen que la basílica es va construir sobre el calvari bíblic, o Gòlgota, on Jesús va ser clavat a la creu, va morir i va ressuscitar.

La basílica es troba encaixonada en un pati a la frontera dels barris cristià i musulmà, i sembla sorgida del no-res. Sempre està plena de turistes i pelegrins, sobretot perquè té les cinc últimes Estacions de la Creu: de la deu a la catorze. Us heu de vestir discretament: els vigilants poden ser estrictes i potser no us deixaran entrar amb les cames, espatlles o esquena descobertes.

Extended track 3

Hola, gràcies per treure el cap darrere l'escena! Estem parlant del mont del Temple, un turó situat a la Ciutat Vella de Jerusalem. El lloc actual és una plaça plana envoltada de murs de contenció (inclòs el Mur Occidental). Està dominada per tres estructures monumentals, una de les quals és l'edifici més fotografiat del món, la Cúpula de la Roca. La Cúpula de la Roca està situada aproximadament al centre de la plaça, que està envoltada de xiprers. La superfície plana i pavimentada, rodejada d'atractius edificis del període mameluc, ocupa 140 hectàrees. Passejar per aquest lloc històric és un veritable contrast amb el soroll i la congestió dels carrerons i carrers empedrats. Avui, el recinte és l'espai públic més gran de Jerusalem Oriental. En contrast amb l'austeritat de l'entorn, el temple és una barreja exòtica de blau marí, lapislàtzuli, verd maragda, ocre i malva. Es diu que la cúpula d'or original va desaparèixer fa molt temps, i que les 5.000 làmines d'or que cobreixen la cúpula avui han costat al rei de Jordània 8,2 milions de dòlars.

Extended track 4

Hola, gràcies per treure el cap darrere l'escena! La torre de David es troba a la porta de Jaffa. El museu recorre 4.000 anys d'història antiga de la ciutat vella de Jerusalem i hi ha una magnífica vista de la ciutat i les ruïnes que hi ha a sota de la torre. Els fonaments d'aquesta estructura tenen almenys 2.000 anys, i els van construir els asmoneus. La Ciutadella de David ha tingut moltes funcions, a més de la seva funció principal com a avançada militar. Al segle IV, era un monestir. Els croats van construir el fossat al seu voltant per protegir els pelegrins. Durant 400 anys va ser una guarnició turca. Els britànics van ser els primers d'utilitzar la torre com a centre cultural.

Extended presentation mode – Episode 2

Extended track 1

Hola, gràcies per treure el cap darrere l'escena! La vall del Cedró es troba a la part oriental de la ciutat vella de Jerusalem. Separa el mont del Temple del mont de les Oliveres. Travessa el desert de Judea a Cisjordània, cap a la mar Morta, i descendeix més de 1.200 metres en el seu recorregut de 32 kilòmetres. L'antic monestir de Mar Saba es troba a la part baixa de la vall.

Extended track 2

Hola, gràcies per treure el cap darrere l'escena! La ciutat de Jerusalem està envoltada de muralles de temps antics. Les muralles daurades de pedra calcària que hi ha avui aquí es van construir al segle XVI, durant l'imperi otomà. Selim el Gran va portar Jerusalem sota domini turc el 1517, però va ser el seu fill, Solimà el Magnífic, qui va ordenar construir aquestes muralles tan conegudes. Tant les muralles com la ciutat vella de Jerusalem són patrimoni mundial de la UNESCO. S'hi van incloure el 1981.

Extended track 3

Hola, gràcies per treure el cap darrere l'escena! Actualment, el Mur Occidental, construït amb pedra calcària, fa uns 50 metres de llarg i uns 18 metres d'alçada, però s'estén molt més sota terra. Això només és una part relativament petita d'un antic mur de contenció molt més llarg, conegut en el seu conjunt com a "Mur occidental". Si veniu aquí, potser veureu visitants que entaforen paperets a les esquerdes del mur. Són pregàries i peticions, però no fan cap favor al mur. Un estudi del 2014 va descobrir que algunes parts s'estaven erosionant i un arqueòleg encarregat del patrimoni d'Israel va proposar de fer una revisió a fons de tot el mur.

Extended presentation mode – Episode 3

Extended track 1

Hola, gràcies per treure el cap darrere l'escena! Actualment la principal via de Betlem és el carrer del Pessebre, que s'estén des de la tomba de Raquel fins a la plaça del Pessebre, el punt central de la ciutat. Aquesta plaça és testimoni de la barreja de cultures de Betlem, amb la Basílica de la Nativitat a un costat i la Mesquita d'Omar a l'altre. Més amunt de la plaça del Pessebre hi ha la ciutat vella i el basar. La millor manera de conèixer-ho és a peu.

Extended track 2

Hola, gràcies per treure el cap darrere l'escena! El llançador de flors, de l'artista Banksy, és un grafit d'un home amb mocador i gorra de beisbol que llança un ram de flors. L'home i l'embolcall del ram són en blanc i negre; les flors i les tiges que sobresurten de l'embolcall són de color. Els grafits de Bansky són un símbol, sovint polític. Costa d'imaginar un lloc més adient per aquesta obra. El grafit d'aquest artista sigil·lós és només un dels molts que han convertit el mur de separació israelià en un ampli llenç. Per molts turistes, són l'atractiu principal de Betlem.

Annex 5.6. Presentation modes in the main study (electronic format)

Link to the website: https://media.i2cat.net/imac/adtest_cat/