

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 (Reviere 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 360° 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 360° 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; Kinatader 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.

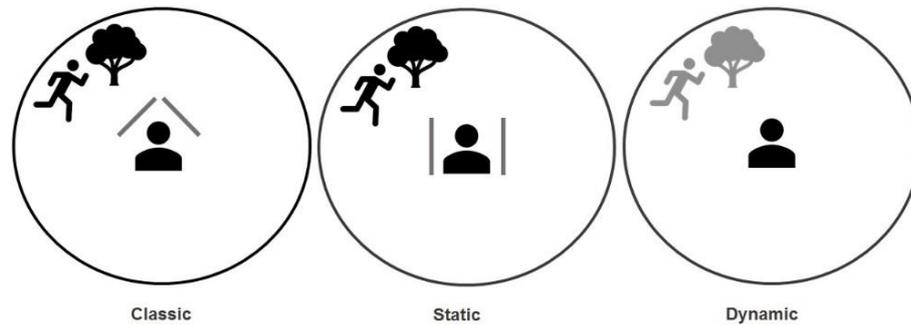


Figure 1 *AD presentation modes used in the preliminary study*

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, Reviere 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)

<p>Més pelegrins esperen en una llarga cua a l'entrada d'un santuari.</p> <p>[back translation] 'More pilgrims wait in a long line outside the entrance to a shrine.' (ep. 1)</p>	<p>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.</p> <p>[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)</p>
---	---

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):

<p>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.</p> <p>[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)</p>	<p>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.</p> <p>[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)</p>
--	---

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 adicional. 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 Banksy 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 graffiti 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 tidbits 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/nubxhjiwc091/5WypbBxFEQqS8I6WaKUUQgg/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 Millennium*, 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 approach 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](https://doi.org/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. Southampton: 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_reviuers.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, FELLS, 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, FELLS, 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.