Audio description in 360º videos: Results from focus groups in Barcelona and Kraków

Anita Fidyka
Universitat Autonoma de Barcelona

Anna Matamala
Universitat Autonoma de Barcelona

Abstract

This article discusses how audio description could be integrated into 360º 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, 360º videos, virtual reality

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), 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 gives an overview of virtual environments. Section 3 reviews previous work in the field of AD in high-immersive environments. Section 4 outlines the methodological aspects of the focus groups, and Section 5 offers a discussion of the results. Finally, conclusions and future research possibilities are presented in Section 6.

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."

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 2018), 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).

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.

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.

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.

5. 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.

5.1. 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.

5.2. 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.

6. 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 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

Author’s address

Anita Fidyka
https://orcid.org/0000-0003-4135-2654
Universitat Autònoma de Barcelona
Department of Translation, Interpreting and East Asian Studies
K-1002, Campus de la UAB
Bellaterra (Cerdanyola del Vallès), 08193 Barcelona
Spain
anita.fidyka@uab.cat

Anna Matamala
https://orcid.org/0000-0002-1607-9011
Universitat Autònoma de Barcelona
Department of Translation, Interpreting and East Asian Studies
K-1002, Campus de la UAB
Bellaterra (Cerdanyola del Vallès), 08193 Barcelona
Spain
anna.matamala@uab.cat