Título: (Sub)títulos en realidad virtual cinematográfica: un estudio descriptivo

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

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

1. Introduction[1]

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

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

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

"disrupt [the immersive experience] as little as possible" (Brown *et al.*, 2018: 1), guaranteeing both accessibility and usability.

A necessary first step before defining guidelines for the creation of subtitles in immersive environments is to describe the limited existing practices and gather viewers' feedback. This article will focus on the descriptive aspect and will analyse the (sub)titles found in a multimedia corpus of cinematic virtual reality content generated by The New York Times and the BBC[2].

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

2. Immersive content

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

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

---

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

used, it refers to the concept of cinematic virtual reality (CVR) defined by Mateer (2017: 15):

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

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

2.1. Features of cinematic virtual reality

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

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

A well-written story is crucial to achieve an immersive experience in CVR content. As stated before, one of the main challenges for CVR content directors is the lack of control over viewers' gaze directional behaviour, because they are free to look at any point in the

viewing sphere. Other challenges that have been reported are the difficulties finding the right shots, hiding the crew kits and engaging the viewers (EBU, 2017). Due to these and other hindrances, a new grammar of filmmaking needs to be developed. In Dooley's words (2017: 165): "Just as the filmmakers of the late nineteenth century took some time to experiment with screen grammar and establish the rules of narrative storytelling on the two-dimensional screen, so too are VR developers now exploring a new screen grammar for the 360-degree, interactive space."

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

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

2.2. Genres

According to the EBU report on VR (2017: 9), "a majority of the stories are either history or news and current affairs products." They also point out the potential of VR content for music (with binaural audio) and sport events, and to a lesser extent for fiction products and promotional material. Mateer (2017) also agrees that most CVR content features non-

fiction stories, and he refers to promotional material, travel and sport events. He also provides some examples of action-based contents, such as "Lewis' Escape The Living Dead (2016) and Lin's HELP (2015)" (Mateer, 2017: 15). According to Grambart (2015), the current state of CVR can be compared to the early filmmaking, because both have started recording documentary or journalistic work. When cuts were introduced, and cinematography was born, filmmakers had the filmic language and strategies to tell more complex stories. CVR is now in that early period focusing on non-fictional content and a new cinematic grammar needs to be created in order to turn CVR into real storytelling (Grambart, 2015).

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

3. **Subtitles in 360º content**

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

The BBC Research & Development team has recently published a White Paper (Brown *et al.*, 2018) on subtitles in 360º videos. According to them, the main challenges presented by this new medium are caused by the freedom of movement that viewers have in the 360º sphere. This causes that "nowhere in the scene is guaranteed to be visible to the viewer" and "there will always be something behind the subtitle, and we may not know what this

will be" (Brown *et al.*, 2018: 1). Other challenges are related to immersion, because preserving the immersive experience is vital for 360º content, and "it will be important that the subtitles disrupt this as little as possible" (*ibid*., 1). Also, as there is the possibility that subtitles appear outside the Field of View (FoV) of the viewers, the effort to find the subtitles should be minimum. Finally, they also highlight the challenge of minimising VR sickness, because some viewers may experience dizziness when using VR and the implementation of the subtitles should not contribute to that (Brown *et al.*, 2018). For Rothe *et al.* (2018) the challenges are similar: position (where to place the subtitles), speaker location (and they insist that this is difficult for CVR content, because speakers can appear outside the FoV) and VR experience (related to the feeling of being immersed).

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

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

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

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

4. Methodology: multimedia corpus

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

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

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

4.1. Selection criteria

Audiovisual content developed in 360 degrees is not as available as 2D content, especially when it comes to professional quality. Also, accessing 360º videos and being able to

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

Following these criteria, it was decided to extract all the videos that were found in the NYT VR app and the BBC YouTube channels in April 2018. From the NYT VR App, 472 videos were selected on the 9th April 2018. From the 472 videos analysed in the NYT VR platform, two were discarded. One for not fulfilling the selection criteria (it was CGI created) and another for having been removed from the platform. The entire duration of the sample was 18 hours, 42 minutes and 42 seconds and the average duration per video was 2 minutes and 23 seconds. Those videos could be accessed via app in a smartphone or via browser (in YouTube or NYT VR Player) in any computer. From the BBC YouTube channels, 99 videos were selected on the 10th April 2018. From the 99 videos analysed in the BBC YouTube channels, six were discarded for not fulfilling the selection criteria (were CGI created). The entire duration of the sample was 6 hours, 23 minutes and 22 seconds, and the average duration per video was 3 minutes and 52 seconds (see Table 1). The videos could be accessed via YouTube in a smartphone or in a computer.

Broadcaster	No. of videos	Total duration of the sample	Average duration per video
The New York Times	472	18:42:42	00:02:23

BBC	99	06:23:22	00:03:52

Table 1. Summary of the analysed sample.

4.2. Data extraction process

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

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

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

b. Translation: titles that provided a written translation for the voice-over.

10

c. Speaker identification: titles that contain the name of the speaker.

d. Location identification: titles that contain the name of the location.

e. Directions: titles that contain directions to indicate the viewers to direct their gaze to a specific location (for example, 'look up').

f. Additional explanatory information: long titles including several sentences and/or paragraphs that add extra information about the main story of the video to inform the viewers.

g. Explanatory titles: short titles that include information other than speaker, location or directions.

6. A column labelled "Speaker identification", to indicate the different strategies for character identification. A picklist was provided (Colours, Tags, Placement, Other (to be defined)).

7. A column labelled "Non-speech information", to indicate the different types of non-verbal information. A picklist was provided (Sound effects, Music, Paralinguistic information).

8. A column labelled "Placement". A picklist was provided (Evenly Spaced, Follow Head Immediately, Fixed position, Other (to be defined)). For the position of the titles, adapted version of the classification by Brown *et al.* (2018) is used:

a. Evenly Spaced: subtitles are placed into the scene in different fixed positions, equally spaced around the video.

b. Follow Head Immediately: subtitles are always visible in front of the viewer. If the viewers move their head, the subtitles move with them, always visible at the same location.

c. Fixed position: subtitles are place into the scene in a single fixed position.

d. Other (to be defined): this last category was added in case an unclassified type of implementation appeared.

9. A column labelled "Position". A picklist was provided (Middle of the image, At the bottom, At the top, Next to the speaker, Next to an object or a person, On top of character(s), Integrated in an area of the image).

10. A column labelled "Font colour", in an open field.

11. A column labelled "Background box". A picklist was provided (Yes, No).

12. A column labelled "Graphic strategies", in an open field.

13. A column labelled "Font type". A picklist was provided (Sans-serif, Serif, Mixed).

14. A column labelled "Series", which identify if the video belonged to a specific series developed by the broadcasters, in an open field.

15. A column labelled "Category". This metadata would serve to understand which type of genres are the most popular among 360-degree content. The following picklist was created: News, Life Stories, Learning, Music, Sports, Travel, Science & Nature, History, Art, Culture & the Media, Comedy, Politics, Horror & Supernatural, Action & Adventure, Thriller, Crime, Drama, Promotional. This picklist is based on BBC genre classification[3]. The list was shortened for the sake of simplification, according to the most popular genres for 360-degree videos.

16. A column labelled "Date", which indicated the date of selection, in date format.

17. A column labelled "Timecode", which indicated the exact time in the video were the title appears, in time format.

18. A column labelled "Screenshot ID", in an open field. When analysing the videos, screenshots with their corresponding timecodes of all examples included in the corpus were created and stored.

19. A column labelled "Comments", in an open field.

20. A column labelled "URL", in an open field.

5. Results from a qualitative analysis

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

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

5.1. Results from NYT VR

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

---

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

1. Function

The percentage of videos that included each function has been calculated (see Table 2). Please notice that the number of instances is not analysed but the fact the video included a function or not. Also, one video can include titles with more than one function.

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

Table 2. Information about function of titles in the videos from NYT VR app.

2. Placement

The percentage of videos that included each type of placement implementation has been calculated (see Table 3). One video can include more than one type of placement.

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

Follow Head Immediately	0	0%
Other (to be defined)	0	0%

Table 3. Information about placement of titles in the videos from NYT VR app.

3. Position

The percentage of videos that included each type of positioning has been calculated (see Table 4). One video can include more than one type of placement.

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

Table 4. Information about position of titles in the videos from NYT VR app.

4. Speaker identification

In Table 5 below, the different speaker identification strategies found in the analysis are reported.

Speaker identification strategies	No. of videos	Description

Tags	2	The tag is displayed below the subtitle, indicating who is speaking, to differentiate the subtitle from others appearing before and after.
Placement	1	The subtitle is displayed next to the speaker.
Colours	0	
Other (to be defined)	0	

Table 5. Information about speaker identification strategies in the videos from NYT VR app.

5. Non-speech information

Only one occurrence from 1185 titles reads: "Crowd chanting".

6. Font colour

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

7. Background box

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

8. Graphic strategies

From 1185 titles, 53 (4.5%) present some graphic strategies. The variety of graphic effects implemented in the videos is considerable. There are some effects applied to the font type, such as the usage of customised font to highlight the title of the video, or fonts with metallic texture and sparkling effects. Also, different fading in and out effects are implemented. Some videos present creative or integrated titles in different forms, for example: titles integrated in the form of a newspapers layout; titles integrated in the image

as if they were a poem; titles integrated inside a drawn map; titles integrated in the roof of the house that appears in the image; title integrated as if it was a monitor for the biker; titles integrated as if they were a recipe; titles integrated in rear and front parts of the plane and titles integrated in the form of a map. A video where three arrows are used to indicate who the speaker is was also found.

An interesting video in the corpus is titled "Sensations of Sound"[4]. The video features the story of Rachel, who gained partial hearing at age 20. She explains how she felt music, before and after receiving a cochlear implant. This video is about deafness, and they carefully created subtitles for it and all kind of integration strategies for titles can be found.

9. Font type

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

10. Category

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

5.2. Results from BBC

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

1. Function

The percentage of videos that included each function has been calculated (see Table 6).

Function	No. of videos	% (from total 37 videos with text on screen)
Explanatory titles	21	56.8%

---

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

Additional explanatory information	14	37.8%
Location identification	10	27%
Directions	7	18.9%
Speaker identification	5	13.5%
Transcription	3	8.1%
Translation	0	0%

Table 6. Information about function of titles in the videos from BBC VR videos.

2. Placement

The percentage of videos that included each type of placement implementation has been calculated (see Table 7).

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

Table 7. Information about placement of titles in the videos from BBC VR videos.

3. Position

Th percentage of videos that included each type of positioning has been calculated (see Table 8).

Position	No. of videos	% (from total 37 videos with text on screen)
Middle of the image	31	83.8%
Next to an object or person	12	32.4%
Next to the speaker	5	13.5%
At the bottom	2	5.4%
Integrated in an area of the image	2	5.4%
On top of character(s)	1	2.7%
At the top	0	0%

Table 8. Information about position of titles in the videos from BBC VR videos.

4. Speaker identification

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

5. Non-speech information

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

6. Font colour

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

7. Background box

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

8. Graphic strategies

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

9. Font type

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

10. Category

From the analysed 93 videos: 26 (28%) correspond to Science and Nature; 17 (18.3%) correspond to News; 15 (16.1%) correspond to Life Stories; 14 (15.1%) correspond to Art, Culture and the Media; 6 (6.4%) correspond to Politics, 4 (4.3%) correspond to Sports; 4 (4.3%) correspond to Travel; 3 (3.2%) correspond to Music; 2 (2.1%) corresponds to Promotional; 1 (1.1%) correspond to Drama; and 1 (1.1%) correspond to History.

6. Current (sub)titling practices in immersive media

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

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

suitable for immersive journalism (De la Peña *et al.*, 2010), documentaries and potentially music and sport events.

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

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

Regarding placement, titles were mostly placed evenly spaced (97.5% for NYT and 35.1% for BBC) or in a fixed position (50.9% for NYT and 64.9% for BBC). Titles that followed

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

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

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

As for graphic strategies, NYT used them mainly for aesthetical purposes in 4.5% of its videos, to improve the video content or the story. BBC used it more (45.9% of the videos), but apparently without a specific intention which might be considered distracting in some cases. In general, the integration of titles in the CVR content was more creative and aesthetical than in 2D content. An example is the graphic strategies found in the video

"Sensations of Sound" by NYT, as an example of what could be done. Some strategies that were found in the video: (1) words appear one by one, and they are vertically positioned next to the speaker; (2) sets of words appear one by one as the speakers pronounce them, and they are vertically positioned next to the speakers; (3) some words (for example, 'blurred') are emphasised by adding extra spacing between characters; (4) some titles appear integrated with illustrations (a figure playing the piano or the guitar) that somehow represent what is being said; (5) some titles appear following the rhythm of a metronome; (6) the word 'vibration' is emphasised by being represented with a vibration effect; (7) some titles appear inside a comic bubble; (8) the word 'jump' is emphasised by using a zoom in/zoom out effect, like if the word was actually jumping; (9) some titles are integrated inside a stave for musical notes; (10) some titles appear integrated inside a stave, but vertically, as if each word were a musical note; (11) the word 'world' is emphasised with a flickering effect; (12) the word 'signing' is emphasised by appearing the letters one by one, probably for the viewer not to mix it with 'singing', because the context could be misleading.

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

7. **Conclusions**

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

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

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

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

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

8. References

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

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

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

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

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

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

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

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

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

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

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

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

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

FRAILE, Isaac *et al.*, 2018: "Personalized and Immersive Presentation of Video, Audio and Subtitles in 360º Environments: An Opera Use Case", paper presented in ACM TVX 2018 [available at https://figshare.com/articles/Demo_Personalized_and_Immersive_Presentation_of_Video_Audio_and_Subtitles_in_360o_Environments_An_Opera_Use_Case/6527531, date of retrieval: March 23rd, 2019].

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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