doi:10.20944/preprints202107.0621.v1

Article

Subtitling 3D VR Content with limited 6DoF: Presentation Modes and Guiding Methods

Mario Montagud [1, 2,*], Cristian Hurtado [1], Juan Antonio De Rus [1] and Sergi Fernández [2]

[1] Universitat de València, Departament d'Informàtica. Spain ({hursan, juanrusa}@alumni.uv.es)
[2] Fundació i2CAT, Barcelona. Spain ({mario.montagud, sergi.fernandez}@i2cat.net)
* Correspondence: mario.montagud@i2cat.net

Featured Application: Requirements, guidelines and insights on how to subtitle 3D Virtual Reality (VR) content.

Abstract: Every (multimedia) service needs to be accessible. Accessibility for multimedia content is typically provided by means of access services, of which subtitling is likely the most widespread one. Up to date, many recommendations and solutions for subtitling classical 2D audiovisual services are available. Likewise, recent efforts have been devoted to devising adequate subtitling solutions for VR360 video content. This paper, for the first time, goes a step beyond, by exploring two key requirements to fulfill remaining challenges towards efficiently subtitling 3D Virtual Reality (VR) content: presentation modes, and guiding methods. By leveraging insights from earlier work on VR360 content, the paper proposes novel presentation modes and guiding methods to not only deal with the freedom to explore the omnidirectional scenes, but also with additional specificities of 3D VR compared to VR360 content: depth, 6 Degrees of Freedom (6DoF), and viewing perspectives. The obtained results prove that always-visible and a novel proposed comic-style presentation mode are far more appropriate than state-of-the-art fixed-positioned subtitles, mainly in terms of immersion, ease and comfort of reading, and identification of speakers, when applied to professional pieces of content with limited displacement of speakers and with limited 6DoF (i.e. users are not expected to largely navigate around the virtual environment). Likewise, even in such limited movement scenarios, the results show that the use of indicators (arrows), as guiding methods, is well received. Overall, the paper provides relevant insights and paves the way toward efficiently subtitling 3D VR content.

Keywords: Accessibility, Guiding Methods, Immersive Media, Immersion, Subtitling, Virtual Reality

1. Introduction

Every (multimedia) service needs to be accessible. This is not only to adhere to worldwide regulation frameworks, but to secure global e-inclusion and equal access to information to everyone [1]. In multimedia services, accessibility is typically provided by means of access services (e.g. subtitling, audio description, and sign language interpreting), appropriate interaction modalities (e.g. User Interfaces, voice control), and assistive technologies (e.g. voice readers, content processing to address specific audiovisual impairments…). The study in [1] provides statistics about the percentages of the worldwide population with some form of audiovisual impairments, and about the increasing ageing process, which is closely related to disability rates. These statistics reveal the significant percentage of the population with accessibility needs.

Multimedia content includes audio and visual elements which, at the same time, may contain linguistic and non-linguistic components. Effective content comprehension can only be achieved by appropriately processing these audio and visual elements, but this is unfortunately not possible to all spectrum of consumers without the provision of access

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2021 doi:10.20944/preprints202107.0621.v1

services. One of the most deeply explored and adopted access service in traditional audiovisual services is subtitling. Subtitling is not just an essential service for the interpretation of the audio information for the Deaf and Hard-of-hearing (SDH) [2, 3], but has also become a powerful tool for the hearing audience whose barrier is linguistic and not sensorial (*interlingual* subtitles) [4]. Furthermore, subtitles can provide benefits in a variety of scenarios and contexts, like in educational and training environments, culture, as an assistive tool for users with cognitive impairments, and in noisy/public environments where the audio cannot/should be heard [5]. Therefore, subtitles not just have become a key access service, but an intrinsic component of media services. Due to this, this study considers subtitling from a universal point of view, encompassing the two classical categories of subtitles: SDH (mainly *intralingual*, including paralinguistic information, and aimed at persons with hearing loss); and subtitles for hearing users (mainly *interlingual*, but with extra applicability).

Up to date, many subtitling solutions for traditional 2D video content have been proposed, even including support for personalization options (e.g., [6], [7], [2]) and advanced presentation methods for better speaker identification (e.g., [8]). In addition, recent studies have started exploring how accessibility, and subtitling solutions in particular, can be efficiently provided for immersive media services, like VR360 video. While the audiovisual content is framed onto a flat 2D window in traditional video formats, the users have freedom to explore around omnidirectional 360º scenes in VR360 video. This bring new challenges and requirements, which are deeply analyzed in [9].

Similar to VR360 video, 3D Virtual Reality (VR) content and services are becoming mature and very popular, not only for entertainment, but also for many other relevant sectors, like culture [1], communication and collaboration [10], education and training [11], consultation and therapy [12], etc. However, accessibility for 3D VR content is still in its infancy. This paper, for the first time, explores two key requirements to fulfill remaining challenges towards efficiently subtitling 3D VR content: presentation modes, and guiding methods. By leveraging insights from earlier work on VR360 content (e.g. [9] [13]), the paper proposes novel presentation modes and guiding methods to not only deal with the freedom to explore the omnidirectional 360º scenes, but also with additional specificities of 3D VR compared to VR360 content: depth, 6 Degrees of Freedom (6DoF), content blocking and viewing perspectives. In particular, the paper aims at shedding some light with regard to two key Research Questions (**RQ**) when it comes to the implementation of subtitling, as the most widespread access service, in 3D VR environments:

RQ1) What are the most appropriate subtitling presentation modes in 3D VR environments?

RQ2) Do visual indicators, like arrows, provide benefits for guiding the users toward the target speakers in 3D VR environments?

By implementing state-of-the-art and novel alternatives, applying them to professionally created 3D VR test sequences, and running user tests, the obtained results prove that always-visible and a novel proposed comic-style presentation mode are far more appropriate than state-of-the-art fixed-positioned subtitles, mainly in terms of immersion, ease and comfort of reading, and identification of speakers. Likewise, the results show that the use of indicators (arrows), as guiding methods, is well received and provides benefits. Although in the adopted test sequences the key actions happen with delimited spaces with limited displacement of speakers, and without requiring large 6DoF by the users, these test sequences are still worth to evaluate, as they are applicable to, and representative of, a wide variety of use cases, like: e-learning, content watching, sports / cultural events, etc.

In conjunction, the contributions of the paper are expected to become a valuable resource for the interested audience in this field, including: 1) users with accessibility needs and frequent subtitles users, to know about potential options and the ones with higher acceptance; 2) the development community, content producers and service providers, in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2021 doi:10.20944/preprints202107.0621.v1

order to improve their solutions; and 3) the standardization bodies and research community, in order to have an overall view of solved and critical missing aspects, and/or assess to what extent the existing requirements and guidelines are met.

The remainder of this paper is structured as follows. Section 2 briefly reviews the state-of-the-art. Next, Section 3 describes the proposed and implemented solutions, and the adopted 3D VR content, as stimuli for the tests. Section 4 presents the methodology for the user tests, and the obtained results. Section 5 provides a discussion about the obtained results and insights. Finally, Section 6 provides the conclusions and ideas for future work.

2. Related Work

This section reviews the state-of-the-art with regard to subtitling solutions for traditional 2D video, VR360 video and 3D VR content.

2.1. Traditional 2D Video Subtitling

In the last years, research efforts on technological and user experience aspects for subtitles have mostly focused on traditional 2D video, addressing a variety of key aspects, as e.g. reviewed in [2]. Relevant examples are: positioning (e.g. [14]); representation of non-speech information (e.g. [15]); reading speed (e.g. [16]); presentation methods for better speaker identification (e.g., [8]); and personalization (e.g. [2]). Although most of these studies focused on TV-related services, a few of them also focused on live events, like opera performances (e.g. [17]).

2.2. Traditional 2D Video Subtitling

Research studies on VR360 video subtitling have also recently emerged, due to the increasing popularity of this immersive *medium* [1].

First, a study carried out by BBC [18] compared four alternatives for subtitles presentation modes: i) *Fixed-positioned* (*Evenly spaced*): subtitles are equally spaced with a separation of 120º in a fixed position below the eye line; ii) *Follow head immediately* or *Always-visible*: subtitles are displayed always in front of the viewer, and follow him/her when looking around; iii) *Follow with lag*: subtitles appear directly in front of the viewer, and they remain there until the viewer looks somewhere else; then, the subtitles rotate smoothly to the new position in front of the viewer; and iv) *Appear in front, then fixed*: subtitles appear in front of the viewer, and then are kept fixed until they disappear. The obtained results from the conducted user tests (*n=24* participants), using short clips, reflected that the *Always-visible* mode was preferred by users, mainly because: i) subtitles were easy to locate; and ii) viewers did not miss the subtitles when exploring the 360º environment. However, the particular implementation in [18] resulted in blocking effects (i.e. subtitles blocked important parts of the images and were considered obstructive). Second, the study in [19] also compared the first two previous presentation modes (*n=34* participants). Although no clear differences between their appropriateness and on users' preferences were identified, the questions about presence, VR sickness and workload slightly favored the *Always-visible* mode. Both studies remarked the need for further research on the topic, not only considering longer pieces of content and different content genres, but also exploring additional presentation modes.

An additional recent study [20] has contributed with a VR360 video corpus to characterize commercially available subtitling solutions and practices, by analyzing the YouTube channels of two major providers: BBC and NYT. On the one hand, this analysis reflects that the *Always-visible* presentation mode, which seems to be the most preferred by users, is not yet adopted by these providers, although a few VR360 players have recently started considering it [13]. On the other hand, this analysis confirms that many key requirements for efficiently subtitling VR360 content, as analyzed in [9], are not yet met. The offered subtitling solutions up to date mostly rely on burned-in subtitles added at

 doi:10.20944/preprints202107.0621.v1

post-processing stages, having a much stronger focus on enhancing the narrative of the videos than on making the content accessible. In the same line, the study in [13] provides an analysis and categorization of the subtitling features provided by the key existing VR360 video players. It is shown that further work on this topic is needed to fulfill key requirements in this context.

Finally, the study in [9] identifies key challenges and requirements for efficiently subtitling VR360 content, and reports on a set of user tests to shed some light on the most appropriate solutions regarding key aspects: presentation modes, guiding methods, representation of non-speech information, comfortable Field-of-Views (FoVs), and use of easy-to-read language. First, it is shown that *Always-visible* subtitles provide the best user experience, but that subtitles *attached-to-the-speaker* are also well received and have potential, if appropriately implemented. Second, it was shown that arrows are the simplest and most intuitive method for visually guiding the user toward the target speaker.

2.3. 3D VR Content Subtitling

Recent works have reported on relevant issues when it comes to the integration and presentation of subtitles in 3D environments. The study in [21] points out the necessity to further research on how to appropriately subtitle 3D VR content. That study claims that appropriate subtitling solutions are necessary, not just to contribute to the narrative and maximize the user's engagement and immersion, but critically to avoid the occurrence of ghosting effects that may hinder readability, and cause eyestrain and simulation sickness. These conclusions were reached out after reviewing and analyzing, from a descriptive approach, a variety of 3D subtitled movies, using both stereoscopic video and computer-generated 3D content as the production formats. For instance, it is remarked that a non-curated super-imposition of traditional 2D subtitles on top of the 3D content can destroy the viewing experience, having an impact on the depth estimation, and causing ghosting, shadowing and dizziness effects. An appropriate implementation can however minimize the impact of these issues, e.g. by positioning the subtitles close to the rendering plane, and applying appropriate lighting, shades and colors [22].

From the analyzed corpus in [21], interesting insights were found out. Most of the 3D movies place the subtitles in the perspective closest to the viewer, but are not really integrated with the movie dynamics and effects. Some of them adjust the depth of subtitles according to the scene and speaker's location, but neither guidelines are provided nor concrete patterns were identified for these adjustments. Interestingly, one of the analyzed movies (*Avatar, Director: James Cameron, 2010*) adopts a dynamic subtitling strategy based on changing the subtitles' position depending on the location of the speaker on the screen and the type of scene to facilitate reading and improve the tracking of the action. In essence, the study in [21] claims for the need for better subtitling solutions for 3D content, exploring requirements and appropriate alternatives. The goal must be to ensure accessibility and content comprehension, but without having a negative impact on the user experience.

Similarly, an exploratory study concerning the usage of subtitles in video games was conducted in [23]. It was found out that subtitles are not yet efficiently provided in the game industry. That study also highlights the need for appropriate solutions, standards and recommendations for subtitling this relevant medium, given its interactive and immersive nature, and high dynamism. This would not only enhance game accessibility, but the gaming experience for all players, in a variety of application contexts (e.g. entertainment, training, therapy…). In the same line, the study in [24] claimed that the appropriate subtitling solutions for VR "*have the potential to make the overall viewing experience less disjointed and more immersive*".

These two last studies [21, 23] have served as motivation for conducting the research presented in this work, which goes even one-step beyond considering 3D movies, but interactive 3D VR content with limited 6DoF, thus moving closer to the gaming sector. In particular, our work departs from the lessons learned in [9] regarding the use and benefits

 doi:10.20944/preprints202107.0621.v1

of presentation modes and guiding methods for VR360 video, in order to examine their adoption and improvement for their integration with this relevant, but still unexplored, *medium* when it comes to the adoption of appropriate accessibility – and subtitling, in particular – solutions.

3. Materials and Methods: Subtitling Solutions and Stimuli

This section presents the adopted and novel proposed subtitling solutions, in terms of: presentation modes (Section 3.1), and guiding methods (Section 3.2). Then, it describes the VR content stimuli adopted for conducting the tests and assessing the appropriateness, benefits and limitations of the proposed solutions (Section 3.3).

3.1. Presentation Modes

By presentation modes, we refer to how the subtitles are integrated and presented with the audiovisual content, even when exploring the immersive environment.

On the one hand, the two most widely implemented presentation modes for VR360 video (e.g. [9, 13, 18, 19]) have been adapted and adopted for their adequate integration with 3D VR content:

- *Mode A - Always-visible subtitles*: the approach for 3D VR is identical than for VR360 content. Subtitles are attached to the virtual camera, thus being always visible at the bottom center of the FoV (although the positioning could be dynamically personalized, as in e.g. [5, 9]), regardless of where the user is looking at (see Figure 1, top left image).
- *Mode B - Fixed-positioned subtitles*: Unlike in VR360 video where the users can just explore the 360º environment (i.e. 3DoF), 3D VR content brings a depth dimension allowing for free navigation within the environment (i.e. 6DoF). Therefore, apart from the latitude and longitude coordinates typically used in VR360 subtitling [9] [26], the depth dimension acquires increased relevance when implementing *Fixed-positioned subtitles* in 3D VR content, and e.g. a 3D Cartesian coordinate system (x, y, z) needs to be adopted as the reference. However, when the speakers' locations do not largely vary, the implementation of this presentation mode for 3D VR content is similar to the one for VR360 content (see Figure 1, top right image). This is the case of the scenarios considered in this paper, as detailed later.

In addition, previous works, like [9], [25] and the *Avatar movie* (Section 2), have revealed the potential of adopting more dynamic presentation modes for subtitles, by rendering them in a close position to the associated speaker. By leveraging the insights and lessons learned from these state-of-the-art implementations and studies, a third novel presentation mode is proposed in this paper:

- *Mode C - Comic-style subtitles*: It consists of adding bubbles associated to the speakers' faces, like in comics, and presenting the subtitles in the planes making up these bubbles (see Figure 1, bottom left image). In this mode, the size of the bubble is slightly adjusted for an adequate fit of the active subtitles text. The size is not meant to largely vary, as the amount of characters per subtitle frame are typically limited to 2 lines and 40 characters per line [13] [27], for any presentation mode. Likewise, being the subtitles rendered in 2D planes, a dynamic algorithm has been adopted to render these planes always orthogonally to the users' viewing perspective to maximize readability. The bubbles are only made visible while associated subtitles need to be presented.

Figure 1. Adopted and Proposed 3D VR Subtitling Solutions; top left) *Always-visible* subtitles; top right) *Fixed-positioned* subtitles; bottom left) *Comic-style* subtitles; bottom right) 3D arrow (indicator) as guiding method.

3.2. Guiding Methods

Regardless of the presentation mode being provided / enabled, users will have freedom to explore the 3D VR environment. During exploration and navigation, it could happen that the active speaker is outside of the current user's FoV. Although spatial audio could support users in perceiving where the speaker is, deaf and hard of hearing users could not hear it, or the audio could not be listened in noisy/public environments.

Therefore, appropriate visual guiding methods are additionally needed to help the users in intuitively finding the speaker or main action in the 3D VR environment when audio cues are lacking or cannot be accessed.

The study in [9] reported on a series of iterative user tests comparing visual guiding methods for VR360 video: sided text, arrows, and a radar. It was shown that the use of arrows was the preferred option by users, being perceived as the most simple, yet intuitive and effective, method to support in finding the target speaker in the omnidirectional environment.

The implementation for VR360 video in [9] relied on just considering the position of the speaker in the horizontal direction (latitude angle, from -180º to +180º), as it was found out that considering the vertical direction (longitude angle, from -90º to +90º) was not really necessary, as identifying the speaker in the vertical axis was very intuitive, and it is not common to find speakers near the poles of the sphere. Therefore, the arrows simply pointed to the right or left, depending on the relative position of the speaker compared to the current user's FoV. This was proved to be effective enough, and the simplest less invasive and most natural solution. However, this approach cannot be adopted in 3D VR content, as the users can freely explore and navigate around the environment (i.e. 6DoF). Accordingly, a 3D arrow able to point at all directions has been implemented in this work as the visual guiding method (see Figure 1, bottom right image). The arrow just points to the target active speaker, if any, by comparing the current user's position (and associated FoV) to the speaker's one. The arrow is made visible every time a speaker is active, and thus subtitle frames need to be shown, and it could be optionally just presented when the speaker is out of the FoV and/or at a distance higher than a threshold. This optional feature has not been enabled in the conducted tests though. The position of the arrow can also be

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2021 doi:10.20944/preprints202107.0621.v1

customized, but according to the results in [9], users preferred it to be positioned close to the subtitles for an easier identification and better integration.

3.3. Stimuli

In order to assess the appropriateness and benefits of the proposed subtitling presentation modes and guiding methods, a professionally created VR content episode has been adopted [28]. Although the content was initially targeted at assessing the potential of Social VR [29] for shared video watching, it has been adopted for this study due to its professional quality, the availability of a full 3D version of the episode (including a virtual space and more than one character), and of sufficient test sequences with the adequate length.

Next, its theme and production process are briefly described to provide a general overview of the used stimuli and sequences.

3.3.1. VR Content Theme

The VR episode consists of a thriller-like plot revolving around a crime investigation story. Being inspired by movies like *The Usual Suspects*, the VR episode departs from the murder of a celebrity, and revolves around the interrogation to two suspects. Its initial and main objective was to use it to assess the potential of Social VR for shared video watching in sessions of two remote users, who observe the interrogations and play the role of inspectors. As the focus of Social VR is the interaction between users within the context of shared media consumption, and not that much the interaction with the environment and other characters, the episode recreates interrogation scenes behind one-way mirrors, like in classical police stations. Likewise, unlike traditional watching apart together scenarios, in which the users watch exactly the same content (e.g. [29]), the adopted VR episode includes a shared observation space giving connection to two separate interrogation rooms, where a different suspect is being interrogated by a police inspector. Therefore, although the integrated users in the 3D environment share a common space and can directly see and talk to each other, they can only see and hear only one of the two interrogation scenes belonging to the same story. The goal of this innovative approach was to boost interaction and the exchange of impressions and findings between the two users to reach a conclusion about the authority of the crime.

Figure 2 illustrates the overall view of the ideated and recreated 3D scenario, as well as how the 3D VR scenario looks like and the viewpoints for each user.

Although no users are captured and integrated in this work, the availability of two related and similar VR content pieces for each of the two interrogation scenes, each one with a duration of 8 min, allows having enough sequences for each test condition to overcome memory and order effects.

3.3.2. Production Process

After the selection of the theme and scenario, the next steps consisted of writing the script and casting the actors. The story was further developed, revolving around the murder of Ms. Yelena Armova, a wealthy British celebrity at the peak of her career, in still unknown circumstances. Two persons are the main suspects: *suspect 1)* Mr. Ryan Zeller, the lover of the victim; and *suspect 2)* Ms. Christine Gérard, her assistant. The two suspects are being interrogated by a police inspector in each of the interrogation rooms (Figure 2), and the story shows that they have a very distinct version about what happened, and that both of them have things to hide.

A casting process was conducted to select the actors representing the roles of the police inspector and the two suspects. Their participation was necessary, mainly because the objective was to recreate a hyper-realistic VR story resembling real characters, and not cartoon-like avatars.

With respect to content production and consumption, different versions of the VR episode were produced by adopting a variety of content formats, ranging from stereoscopic 360º video to full 3D, and also including a hybrid version with the 3D environment

 doi:10.20944/preprints202107.0621.v1

and inserted video billboards. The goal was to gain insights about direct implications on the required infrastructure, complexity and costs, but especially to assess the impact of such content formats and combinations on the user experience [28]. Given that the goal of this experiment is to explore appropriate subtitling solutions for 3D environments, the 3D version of the episode has been adopted.

Regarding the VR environment, the two separate interrogation rooms and the shared observation space, together will all associated elements (e.g. chairs, desks, screens...), were modelled in photorealistic 3D, resembling a 70's look police station room.

In order to create the 3D interrogation scenes, the actors were 3D scanned with a photogrammetric scanner consisting of 96 cameras to obtain the 3D surface of their bodies (see Figure 3.a). Then, a Motiop Capture (MoCap) session was recorded by using a Vicon MXT40S system with 30 cameras (see Figure 3.b), in which each actor wore 59 retro-reflective markers to track their movements. For facial capture, a tool was developed to record the faces by using an iPhone X and the ARKit framework[1], with the help of an ad-hoc helmet (see Figure 3.c). Then, the captured facial content data were synchronized with the MoCap data, and converted into an appropriate format for further 3D editing and adjustments.

Realistic lighting conditions were recreated to provide a natural integration of the users and characters into the 3D virtual environment and to provide a thriller-like atmosphere (e.g. direct light in the interrogation rooms and dimmed lights in the dark shared room where the users are placed). Spatial ambient sound was also prepared, coming from the direction of the actions. This is the case for the doors opened, actions by the actors, sound from the other user, etc.

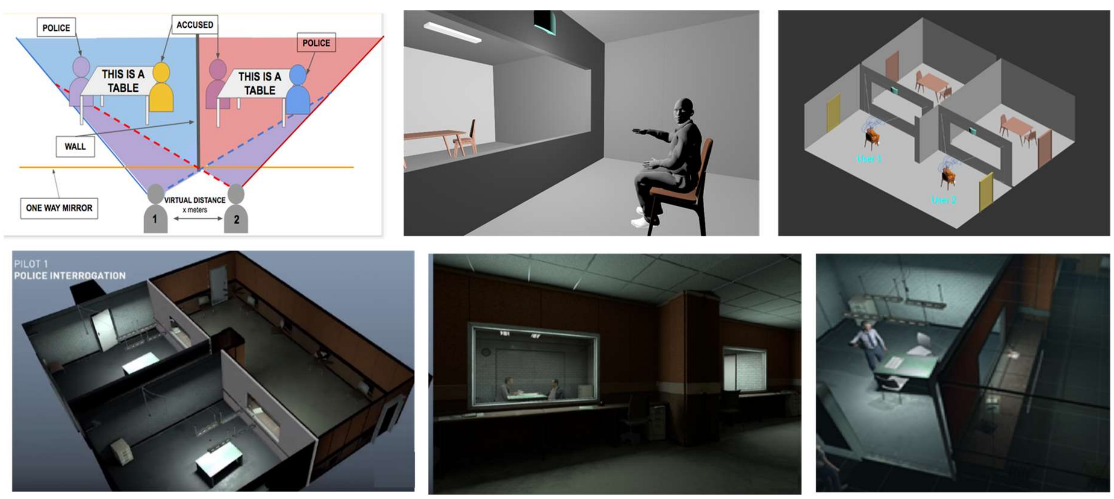

Figure 2. Overview of the 3D modelled and recreated VR scenario, including the layout and view of the 3D environment, the interrogation rooms and scenes, and the users' viewpoints toward them.

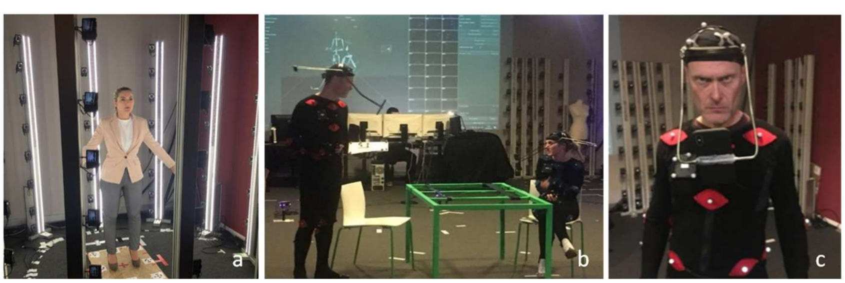

Figure 3. (a) 3D body model creation, (b) MoCap recording, and (c) facial gestures recording.

[1] Arkit Framework: https://developer.apple.com/documentation/arkit

 doi:10.20944/preprints202107.0621.v1

After the recording and modelling of all assets, post-production processes were conducted for all the raw material, including the required adjustment tasks for an appropriate compositing and seamless blending. Finally, all assets making up the experience were integrated in a Unity project.

More details about the developed story, the production analysis and process are provided in [28]. All assets are available as *open-access* on Zenodo: https://zenodo.org/communities/vrtogether-h2020 A video describing the created VR content, and summarizing the production process, is available at: https://www.youtube.com/watch?v=aHO5M1qNmjY Finally, a demo video showcasing how the proposed and adopted subtitling solutions look like when applied to this VR content episode, as well as some implications of their usage, can be watched at: https://www.youtube.com/watch?v=SzastPjzzeM

4. Results: User Tests

This section reports on the user tests conducted to determine the most appropriate subtitles presentation modes, with their pros and cons, and the potential benefits of using guiding methods, in particular arrows, pointing to the target speaker(s).

The followed methodology is explained in Section 4.1, while the results are presented in subsequent subsections.

4.1. Methodology

The user tests were divided into two parts: a first part focused on evaluating the appropriateness of the presentation modes, and a second one on determining the benefits of using an arrow as visual guiding method.

4.1.1. Test Condition: Presentation Modes

The first part of the tests explored the appropriateness of the considered 3D VR subtitling presentation modes (Modes A, B, C). Given that three presentation modes are considered, the two interrogation scenes were divided into two parts of the same length (all with a duration of ~240s):

- Clip 1A: Initial Part of the Interrogation to suspect 1
- Clip 1B: Final Part of the Interrogation to suspect 1.
- Clip 2A: Initial Part of the Interrogation to suspect 2.
- Clip 2B: Final Part of the Interrogation to suspect 2

In order to avoid order effects, the presentation of test conditions was counterbalanced, always starting by the first part of the interrogation scenes to be able to adequately follow the story (Table 1). The clips were presented without audio, so the subtitles become key to understand the story, even if participants do not have hearing impairments.

Table 1. Counterbalancing for Test Condition 1 (Subtitling Presentation Modes) to avoid Order Effects

Participant ID	Clip 1A	Clip 1B	Clip 2A
1-7-13-19	Mode A	Mode B	Mode B
2-8-14-20	Mode C	Mode A	Mode B
3-9-15-21	Mode B	Mode C	Mode A
4-10-16-22	Mode A	Mode C	Mode B
5-11-17-23	Mode B	Mode A	Mode C
6-12-18-24	Mode C	Mode B	Mode A

 doi:10.20944/preprints202107.0621.v1

4.1.2. Test Condition: Use of Guiding Methods

The second part of the tests consisted of adding a visual guiding method (arrow) that dynamically points at the target speaker, regardless of the user's position and viewpoint. The arrows were added to the Modes B and C, in a counterbalanced manner (Table 2). That test condition was not considered for Mode A, as in such a mode it may be possible to have the subtitles and the target speakers in different FoV, and therefore the use of the indicator could cause confusion in such situations.

Table 2. Counterbalancing for Test Condition 2 (Use of Guiding Methods) to avoid Order Effects

Participant ID	Clip 2A	Clip 2B
Odd ID (1, 3, 5…)	Mode B	Mode C
Even ID (2, 4, 6…)	Mode C	Mode B

4.1.3. Equipment and Setup

- Gaming laptop (MSI, i7-10750H, 16GB DDR4-2666MHz, GeForce® GTX 1660 Ti, GDDR6 6GB).
- VR headset: Oculus Quest connected to the laptop via the Oculus link cable.
- Noise cancelling headphones

The participants sat in a comfortable swivel chair in a spacious room, with an appropriate lighting and temperature, and with just the presence of the experiment facilitator.

4.1.4. Procedure

- Step 1 (~5min). Participants are welcome, and introduced to the test.
- Step 2 (~3min). Participants fill in a consent form.
- Step 3 (~3min). Participants fill in a demographic and background information questionnaire.
- Step 4 (~3min). Participants fill in the Simulation Sickness Questionnaire (SSQ) [30].
- Step 5 (~5min). Part 1 - First Test Condition.
- Step 6 (~5min). Participants fill in the Igroup Presence Questionnaire (IPQ) [31] and SSQ questionnaires.
- Step 7 (~5min). Part 1 - Second Test Condition.
- Step 8 (~5min). Participants fill in the IPQ and SSQ questionnaires
- Step 9 (~5min). Part 1 – Third Test Condition
- Step 10 (~5min). Participants fill in the IPQ and SSQ questionnaires
- Step 12 (~5min). Participants fill in the ad-hoc questionnaire on subtitling presentation modes.
- Step 13 (~5min). Part 2 – First Test Condition.
- Step 14 (~5min). Part 2 – Second Test Condition.
- Step 15 (~5min). Participants fill in the ad-hoc questionnaire on indicators.
- Step 16 (~5min). Participants are thanked and said goodbye

Before each test condition, the experiment facilitator helped the participants with the setting and collocation of the VR equipment, and launched the experience.

Overall, the test session for each user had a duration between 60-80 minutes.

4.1.5. Forced Camera Movements

In order to test the effect of the distance and viewing perspectives, for each one of the subtitling presentation modes, smooth camera movements and transitions were programmed and triggered in each test condition, for the two clips. These transitions between positions and implicit viewpoint are sketched and listed in Figure 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2021 doi:10.20944/preprints202107.0621.v1

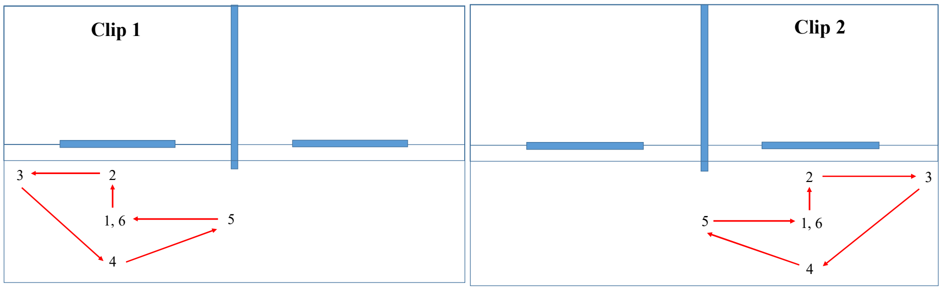

Figure 4. Forced Camera Movements in Clips 1 and 2, with smooth transitions: P1(0-30s): centered position; P2(30-60s): close to one-way mirror; P3(60-90s): corner, low visibility; P4(90-120s): centered position, but far; P5(120-150s): sided, far, crosswise (sided viewing perspective); P6(150-180s): centered position.

4.2. Sample of Participants

In total, 24 users participated in the study, 12 of them were female, and they were aged between 18 and 65 years old (average of 35.12, standard deviation of 14.72), being 13 of them young adults (18-35 years), 9 of them middle-age adults (36-55 years) and 2 old adults (>55 years).

Regarding their study level, 20.8% of them had a secondary school level, 25% were undergraduate university students, 29.1% held a university degree, 20.8% held a PhD degree, and 4.1% preferred not to indicate the study level. All participants were hearing users, but the audio was muted, so the subtitles became a key element to understand the story. 58.3% of the participants had previous experience with VR content consumption using VR headsets (half of them less than once a year, 28.5% of them between 1-5 times per year, 14.2% of them in around a monthly basis, and 7.1% in around a weekly basis).

It must be remarked that no particular filter was applied for the participants' recruitment, beyond having good English level, as the subtitles were presented in that language. It is due to the fact that the considered subtitling modes and use of indicators can provide benefits to the general audience, being potentially applied in many different scenarios, as highlighted in [5].

4.3. Results

This subsection reports on the results for the two parts of the user tests.

4.3.1. Test Condition: Presentation Modes

The impact on presence of each of the considered subtitling presentation modes was assessed by using the IPQ questionnaire [31], which is composed of 14 statements/items to be rated on a seven–point scale (1 to 7). In turn, the 14 items are distributed into four sub-scales:

- *General Presence*: One single item that assesses the general '*sense of being there*'.
- *Spatial Presence*: five items that measure the sense of being physically and bodily present in the virtual environment.
- *Involvement Scale*: four items that measure the attention that the subject pays to the virtual environment and the involvement experienced.
- *Experienced Realism*: four items that measure the subjective experienced sense of realism attributed to the virtual environment

Tables 3-6 provide a summary of the mean and standard deviation values of the answers by the participants for each of the scales, together with the statistical analysis to determine whether significant differences exist between the results obtained for each presentation mode, by using a Wilcoxon Signed Rank test (with 95% Confidence Interval).

Similarly, Figure 5 shows the boxplots of the obtained results for each presentation mode and IPQ scale.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2021 doi:10.20944/preprints202107.0621.v1

The results for all three tested conditions were quite positive, especially for the *Always-visible* (Mode C) and *Comic-style* (Mode B) presentation modes. Indeed, the statistical analysis show that, in the considered scenario and for the considered implementation, the use of *Always-visible* subtitles provided higher presence than *Fixed-positioned* subtitles (Mode A) for each one of the IPQ scales. It is also the case for *Comic-style* (Mode B) subtitles when compared to *Fixed-positioned* ones (Mode A) in terms of the General Presence and Involvement scales. *Always-visible* subtitles (Mode C) also provided higher presence than *Comic-style* ones (Mode B) with regard to the Experienced Realism scale. It may be due to the fact that the bubbles where subtitles are presented resemble comics and animation content, thus having an impact on the experienced realism.

With regard to the results from SSQ, no significant effects / symptoms were noticed to be caused by the VR experience, in any of the test conditions.

Table 3. IPQ – General Presence Scale.

Presentation Mode	Mean	Standard Deviation	Compared to Always-Visible	Compared to Fixed-Positioned	Compared to Comic-style
Always-visible	6.291	0.806	-	**p-value = 0.0032**	p-value = 0.0726
Fixed-positioned	5.583	0.928	**p-value = 0.0032**	-	**p-value = 0.0125**
Comic-style	5.958	0.859	p-value = 0.0726	**p-value = 0.0125**	-

Table 4. IPQ – Spatial Presence Scale.

Presentation Mode	Mean	Standard Deviation	Compared to Always-Visible	Compared to Fixed-Positioned	Compared to Comic-style
Always-visible	5.05	1.873	-	**p-value = 0.04**	p-value = 0.4054
Fixed-positioned	4.91	1.785	**p-value = 0.04**	-	p-value = **0.0609**
Comic-style	5.025	1.907	p-value = 0.4054	p-value = **0.0609**	-

Table 5. IPQ – Involvement Scale.

Presentation Mode	Mean	Standard Deviation	Compared to Always-Visible	Compared to Fixed Positioned	Compared to Comic-style
Always-visible	5.146	1.025	-	**p-value = 0.015**	p-value = 0.3458
Fixed-positioned	5.01	1.041	**p-value = 0.015**	-	**p-value = 0.0039**
Comic-style	5.188	1.019	p-value = 0.3458	**p-value = 0.0039**	-

Table 6. IPQ – Experienced Realism Scale.

Presentation Mode	Mean	Standard Deviation	Compared to Always-Visible	Compared to Fixed-Positioned	Compared to Comic-style
Always-visible	3.635	1.37	-	**p-value = 0.04249**	**p-value = 0.03**
Fixed-positioned	3.541	1.321	**p-value = 0.04249**	-	p-value = 0.1187
Comic-style	3.448	1.329	**p-value = 0.03**	p-value = 0.1187	-

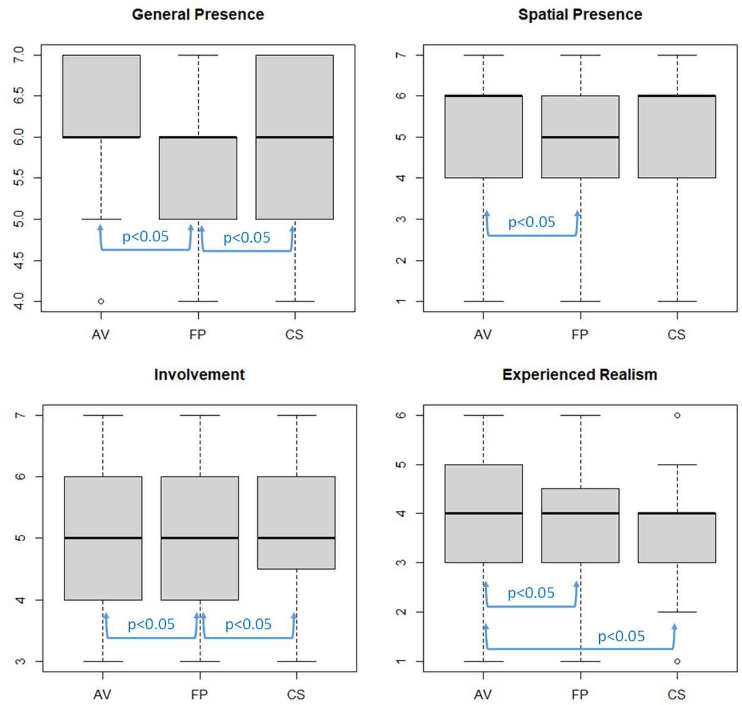

Figure 5. Boxplots of IPQ Scales for each VR Subtitling Presentation Mode (AV = Always-Visible; FP = Fixed-Positioned; CS = Comic-Style).

With regard to the results from the ad-hoc questionnaire on preferences, the answers can be checked in the bar charts in Figure 6. In terms of avoiding content blocking / obstruction, the most preferred mode was *Always-visible* subtitles, closely followed by the other two ones. In terms of ease of reading, *Always-visible* subtitles was also the most preferred mode, followed by *Fixed-Positioned* subtitles. The reason why *Comic-style* was the third preference may be due to the fact that subtitles in such a presentation mode were a bit small, and thus not easy to read, once the distance between the user and the speaker was large. In terms of reading comfort, *Fixed-Positioned* subtitles were the third preference. It may be due to the fact that users had to deviate the viewing patters to read the subtitles and see the speakers when using that mode. In terms of speaker's identification, *Comic-style* was the most preferred option, as subtitles were associated to the target speaker when using such a mode, and *Always-visible* subtitles were the second preferred option. As expected, *Fixed-Positioned* subtitles scored worst with regard to that aspect. In terms of integration with the VR content, *Comic-style* subtitles was again the preferred mode, as subtitles were indeed integrated with the associated speaker(s) on purpose. No significant differences were obtained between the other two modes with regard to that aspect.

In terms of the impact of distance and viewing perspective, *Always-visible* subtitles was clearly the preferred mode, as these two factors heavily have an impact on the other two modes. These other two modes could had been adjusted such that the subtitles rendering planes are always kept normal/orthogonal to the user's viewing perspective, and their size is dynamically adapted according to the distance, but this needs further investigation and fine-tuning. Finally, in overall terms, *Always-visible* subtitles was the most preferred mode, closely followed by *Comic-style* subtitles. This is in line with the results from the IPQ questionnaire, and confirms that the two innovative 3D VR subtitling presentation modes that have been proposed do provide benefits to the user experience.

Figure 6. Preferences on 3D VR Subtitling Presentation Modes.

Finally, the user's position and camera's position (i.e. the viewpoints) were recorded every 0.5s during all test conditions. By doing this, the goal was to compare the users' viewing patters for each one of the considered subtitling presentation modes to assess whether these modes have any impact on the omnidirectional 3D scene exploration. The obtained results for a sub-sample of eight users (those having watched each presentation mode for Clip 1, Table 1) are in Figure 7. These results indeed confirm that when using the *Always-visible* mode the users did explore further the omnidirectional 3D environment, as the subtitles were never missed out. In contrast, a free exploration of the 3D environment when using the *Fixed-positioned* modes, including the *Comic-style* one, can cause the subtitles being out of the FoV and thus a loss of information (remember that the audio was muted and, even if available, users may have hearing impairments or not understand the spoken language). Due to this, a full exploration of the 3D environment was less common for these two modes. The three upper graphs in Figure 7 provide the heat maps from 3D front views, while the three lower ones from 2D aerial projections of the previous ones.

4.3.2. Test Condition: Use of Guiding Methods

In the second part of the tests, the participants watched again the two parts of Clip 2 (2A, 2B), in that order, to appropriately understand the story, with the 3D arrow indicator enabled for the Modes B and C, in a counterbalanced manner (Table 2). The arrow was positioned at the bottom right of the FoV to not block the subtitles in Mode C. In these test conditions, the forced camera movements were especially useful to demonstrate how the arrow can point anytime at the target speaker, regardless of the user's position and viewpoint.

Table 7 provides the answers to the ad-hoc questionnaire focused on assessing the benefits of using indicators. The obtained results shows that many participants (above 60%) believe that the arrow is beneficial for a better positioning within the VR environment and also for a better identification of the active speaker, both when using *Always-visible* and *Comic-style* presentation modes. Participants generally agreed that the use of indicators can contribute to a better content comprehension (almost 70%), and that their inclusion does not have a negative impact on immersion (above 80%), but that even might have a positive impact (especially if relevant for content comprehension). Participants were also quite satisfied with the graphic design of the arrow, although it seems to have room for improvement.

Participants were additionally asked about improvement suggestions with regard to the (use of) visual indicators as guiding methods, and the next ones can be highlighted: use of colors associated with subtitle colors (12.5%); assess the use of 3D gaming radars (12.5%); and improve the design (8.3%), but no concrete suggestions were given to achieve this, beyond enlarging a little bit the size (8.3%) and adding intermittence effects to the arrow (8.3%).

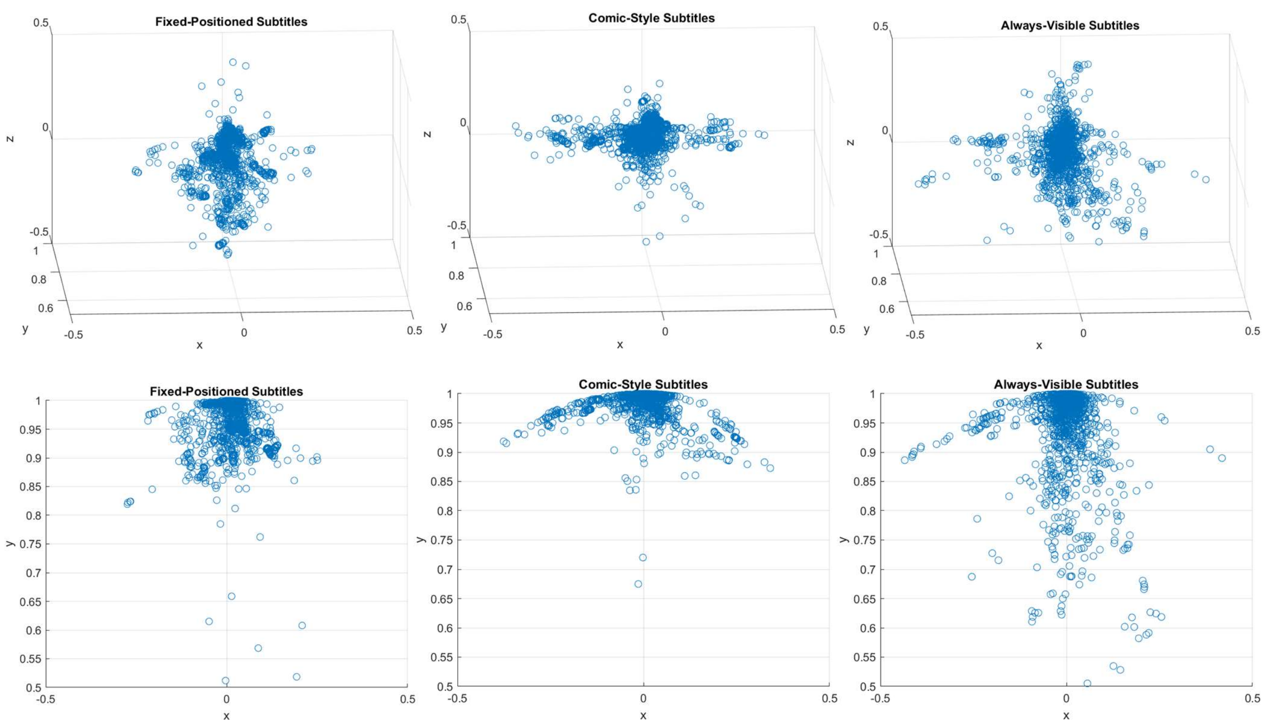

Figure 7. Heat maps of viewing patterns when using each of the considered 3D VR subtitling presentation modes in Clip 1. The upper graphs are 3D front views for each mode, while the lower three ones are 2D aerial projections of the upper ones.

4.3.3. Final Questions

Finally, participants were asked general questions about the relevance of subtitling for hearing users, and of VR subtitling in particular (Table 8).

Over 90% of participants believe both that subtitles are beneficial for hearing users and that an appropriate subtitling of VR content is a relevant feature to be explored and provided.

Participants were also asked about the particular benefits of subtitles according to their opinion, and the main answers are listed next: when consumers do not speak the content language (33.3%); for language learning and improvement (25%); in noisy environments (20.8%) or if/when the audio volume is low (16.7%); to train the reading skills (16.7%); to understand / get the spelling of specific uncommon words and names (12.5%); and when the audio quality is not good (8.3%).

Finally, participants were encouraged to make final free comments, and key ones were: subtitles helped them to understand the story (25%), in fast action scenes it was difficult to read subtitles (16.7%), and that hybrid modes (e.g. combining Modes B and C) could help to overcome the impact of distance and viewing perspectives, but also to increase the immersion once having the speakers within the current FoV (20.8%).

Table 7. Benefits and Appropriateness of Guiding Methods (3D Arrow)

Question Items / Answers	Totally Agree	Partially Agree	Neutral	Partially Disagree	Totally Disagree
The availability of indicators (arrow) is beneficial for a better positioning within the VR environment	7 (29.2%)	8 (33.3%)	5 (20.8%)	4 (16.7%)	-
The availability of indicators (arrow) is beneficial for a better identification of the active speaker when using *Always-visible* subtitles	9 (37.5%)	9 (37.5%)	4 (16.7%)	2 (8.3%)	-
The availability of indicators (arrow) is beneficial for a better identification of the active speaker when using *Comic-style* subtitles	6 (25%)	7 (29.2%)	7 (29.2%)	3 (12.5%)	1 (4.2%)
The availability of indicators (arrow) can contribute to a better content comprehension	6 (25%)	8 (33.3%)	6 (25%)	4 (16.7%)	-
The inclusion of indicators (arrow) can have a negative impact on immersion while watching VR content	-	1 (4.2%)	3 (12.5%)	8 (33.3%)	12 (50%)
The inclusion of indicators (arrow) can have a positive impact on immersion while watching VR content	6 (25%)	7 (29.2%)	5 (20.8%)	4 (16.7%)	2 (8.3%)
The provided indicator (arrow) is visually attractive	6 (25%)	8 (33.3%)	5 (20.8%)	5 (20.8%)	-

Table 8. Relevance of (3D VR) Subtitling

Question Items / Answers	Totally Agree	Partially Agree	Neutral	Partially Disagree	Totally Disagree
Subtitles can contribute to a better content comprehension for hearing users	14 (58.3%)	8 (33.3%)	2 (8.3%)	-	-
In general, I believe that VR subtitling is a relevant feature	15 (62.5%)	7 (29.2%)	2 (8.3%)	-	-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2021 doi:10.20944/preprints202107.0621.v1

5. Discussion

This paper has aimed at shedding some light with regard to the accessibility of immersive media, by focusing on 3D VR content subtitling. As a response to the increased relevance of this content format and medium, jointly with the lack of accessibility solutions for it, two key research questions in this topic have been explored: the appropriateness of different presentation modes, and the potential benefits of using guiding methods.

As with recent studies on VR360 video subtitling [9], the results from the conducted user tests have shown that an *Always-visible* presentation mode is by far more appropriate than a *Fixed-positioned* presentation mode for subtitles. The differences are even larger in 3D VR environments than in VR360 video environments, especially due to two key factors: depth and 6DoF (i.e. freedom to explore). Unlike in VR360 video environments where subtitles can be e.g. replicated every 120º in the sphere (strategy proposed and adopted by BBC, among others [9][13]), this is clearly not an appropriate and sufficient solution in 3D VR environments, because of potential content blocking and the depth dimension, respectively. The conducted user tests have provided relevant insights on these relevant issues. In addition, a novel *Comic-style* presentation mode has been proposed and it has been very well received by users. Indeed, this mode has been shown to provide similar levels of presence than the *Always-visible* mode, for all scales of the IPQ questionnaire, but for the Experienced Realism one (maybe due to the fact that the addition of comic-like bubbles has an impact on the experienced realism, as the story resembles a comic or animation film).

Comic-style has been in general the second most preferred mode by users, but the most preferred one in terms of relevant aspects, like speaker's identification and integration with the VR content. This reflects the potential of this mode, and the convenience of keep researching on assessing its benefits, by ideally overcoming some of its identified drawbacks (e.g. ease of reading, impact of distance and viewing modes, content blocking…).

It should be also remarked that the considered VR scenario brings its own particular conditions that may potentially impact the obtained results. First of all, all actions happen within a delimited area (inside the interrogation rooms), where the speakers do not significantly move. The area is also separated by a mirror from the user's area, so the users e.g. cannot move around the speakers. In addition, despite forcing camera movements, the scenario does not incite the users to navigate around the 3D VR environment, and therefore the scenario has limited 6DoF. Still, these types of test sequences are worth to evaluate, as they are very common in, and applicable to, a wide variety of use cases, like: e-learning, content watching, sports / cultural events, etc.

Despite that the results from this study may not be generalizable for all type of VR content and scenarios, especially those inciting 6DoF and with actions happening in different places and depths, its findings provide valuable answers to RQ1 (what are the most appropriate subtitling presentation modes for 3D VR content?), and have also opened to door to further research on this topic. One example is the exploration of hybrid approaches, where *Always-subtitles* subtitles are enabled when the speakers are out of the FoV, and that mode is switched to *Comic-style* (or to some form of subtitles attached-to-the-speaker) as soon as, and while, the speaker is within the FoV.

In addition, the conducted tests have provided valuable feedback about the benefits of using guiding methods. Even though the use of a 3D arrow as visual guiding method has been well received, the hypothesis is that the results would had even been more positive in scenarios with 6DoF and/or in scenarios in which the speakers and the users can freely and significantly move around the VR environment. Besides, although it was shown in [9] that the use of arrows was preferred to the use of a radar for VR360 video environments, a few users also suggested exploring the use of the radar as guiding method, especially in 3D VR environments with 6DoF. Having determined that guiding methods are beneficial and well received, specific tests on comparing these two methods should be conducted in the future in 6DoF scenarios to gain insights into the most appropriate one(s).

In general, the conducted tests not only have served to validate the appropriateness, benefits and positive reception of the adopted solutions and variants to meet the key identified requirements for efficiently subtitling 3D VR content, but also to receive valuable feedback to fine-tune them, and even to identify further research opportunities.

Finally, the tests have been conducted with a reasonable number of hearing users, but with no particular recruitment filters, given the wide applicability of both subtitles and VR services. A very valuable output for the research community would be the execution of larger-scale tests to gain insights about the impact and benefits of the proposed aspects (presentation modes and guiding methods) in/for different content genres, VR scenarios and audience profiles.

5. Conclusions and Future Work

Immersive media are on the rise, and need to be accompanied with appropriate accessibility solutions. This paper has explored for the first time two key requirements for subtitling 3D VR content: presentation modes and the use of guiding methods. On the one hand, the conducted tests have shown that *Always-visible* subtitles are by far preferred to classic *Fixed-Positioned* subtitles, being in line with the results of recent studies focused on VR360 video [9]. Interestingly, the results have shown that the newly proposed *Comic-style* presentation mode has been in general very well received, even being the most preferred option in terms of relevant aspects (e.g. integration with the VR content, identification of speakers…). This reveals its potential, especially for specific content genres (e.g. gaming, animation…) and audience profiles (e.g. young consumers). On the other hand, the results have proved that the use of guiding methods in 3D VR content, in particular of a 3D arrow as the visual indicator, are beneficial and welcome by users. This is also in line with the results of previous research works on VR360 video [9].

In general, the obtained results not only have served to shed some light on the appropriateness and benefits of the proposed and adopted solutions, but also to get valuable feedback on refining them and identifying opportunities / needs for future research. Given the lessons learned and still open questions, future work will be targeted at:

- Exploring the adoption and benefits of hybrid and advanced presentation modes and strategies, based on the dynamic user's and speaker's positions and viewpoints, as well as on the presented VR scenes.
- Exploring the adoption and benefits of Easy-to-Read subtitles [33] in 3D VR scenes and environments, due to identified difficulties in following the subtitles for specific (fast and high-motion) scenes.
- Exploring the adoption and benefits of all proposed solutions, including 3D radars as guiding methods, in 3D VR scenarios with 6DoF – or with much less limited 6DoF –, likes the ones presented and envisioned in [34].
- Exploring accessibility solutions for social viewing scenarios, like the ones described in [29] and [35].

These future works will also take into account the valuable feedback on refining the proposed implementations. In essence, this work has not only provided innovative solutions and valuable insights and recommendations for the community, but it additionally paves the way for further research in this topic, identifying further relevant opportunities.

Supplementary Materials: Demo video showcasing how the proposed and adopted subtitling solutions look like when applied to adopted 3D VR content pieces, as well as some implications of their usage: https://www.youtube.com/watch?v=SzastPjzzeM

Author Contributions: All authors have contributed substantially to the presented study. Mario Montagud and Sergi Fernández were involved in all phases, from the problem identification and conceptualization to the validation, analysis and documentation. Cristian Hurtado and Juan Antonio De Rus were especially involved in the implementation and evaluation phases.

Funding: This work has been partially funded by the European Union's Horizon 2020 program, under agreement nº 762111 (VR-Together project). Work by Mario Montagud has been additionally funded by the Spanish Ministry of Science, Innovation and Universities with a Juan de la Cierva – Incorporación grant (reference IJCI-2017-34611) and with the ALMA Excellence Network (reference RED2018-102475-T), and by the Fundación BBVA with a Leonardo Grant entitles "Accesibilidad en Medios Inmersivos".

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. M. Montagud, P. Orero, A. Matamala, "Culture 4 all: accessibility-enabled cultural experiences through immersive VR360 content", Personal Ubiquitous Computing, January 2020.
2. A. Matamala, P. Orero, "Listening to Subtitles. Subtitling for the Deaf and Hard-of-Hearing", Berna: Peter Lang, 2010.
3. P. Romero-Fresco, "Subtitling for the deaf and hard of hearing", In Routledge encyclopedia of translation studies, Editors: M. Baker, G. Saldanha, pp. 549-554, Routledge, 2020.
4. D. Kapsaskis, "Subtitling. Interlingual", In Routledge encyclopedia of translation studies, Editors: M. Baker, G. Saldanha, pp. 554-560, Routledge, 2020.
5. M. Montagud, F. Boronat, D. Marfil, F.J. Pastor, "Web-based platform for a customizable and synchronized presentation of subtitles in single- and multi-screen scenarios", Multimedia Tools and Applications, 79, 21889–21923, April 2020.
6. A. Rodriguez, G. Talavera, P. Orero, J. Carrabina., "Subtitle Synchronization across Multiple Screens and Devices", Sensors, 12(7), 8710-8731, June 2012.
7. P. Orero P., C. A. Martín, M. Zorrilla, "HBB4ALL: Deployment of HbbTV services for all", IEEE BMSB'15, Ghent (Belgium), June 2015.
8. Y. Hu, J. Kautz, Y. Yu, W. Wang, "Speaker-Following Video Subtitles", ACM TOMCCAP, 11, 2, Article 32, 17 pages, Jan. 2015.
9. M. Montagud, O. Soler, I. Fraile, S. Fernández, "VR360 Subtitling: Requirements, Technology and User Experience", IEEE ACCESS, Available Online in December 2020.
10. J. Li, et al. "Social VR: A New Medium for Remote Communication and Collaboration", ACM CHI EA '20, 1–8, Honolulu (USA) April 2020.
11. J. Radianti, T. A. Majchrzak, J. Fromm, I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda", Computers & Education, Volume 147, 103778, 2020.
12. T. Xue, J. Li, G. Chen, P. Cesar, "A Social VR Clinic for Knee Arthritis Patients with Haptics", Adjunct Proceedings of ACM TVX/IMX 2020, Barcelona (Spain), June 2020.
13. C. Hughes, M. Montagud, "Accessibility in 360º video players", Multimedia Tools and Applications, October 2020.
14. E. Bartoll E, A. Martínez-Tejerina, "The positioning of subtitles for the deaf and hard of hearing", In Listening to subtitles: Subtitles for the deaf and hard-of-hearing, Frankfurt: Peter Lang, 69-86, 2010.
15. C. Civera, P. Orero, "Introducing icons in subtitles for deaf and hard of hearing: Optimising reception?", In Listening to subtitles: Subtitles for the deaf and hard-of-hearing, Frankfurt: Peter Lang, 149-162, 2010
16. A. Szarkowska, O. Gerber-Morón, "Viewers can keep up with fast subtitles: evidence from eye movements", PLOS ONE, 13(6), 2018.
17. S. Eardley-Weaver, "Opera (sur)titles for the deaf and hard-of-hearing", In J. Díaz-Cintas & J. Neves (Eds.), Audiovisual translation. Taking stock (261-276), Cambridge Scholars Publishing, 2015.
18. A. Brown A., et al., "Exploring Subtitle Behaviour for 360° Video", White Paper WHP 330, 2018.
19. S. Rothe, T. Kim, H. Hussmann, "Dynamic Subtitles in Cinematic Virtual Reality", ACM TVX 2018, Seoul (South Korea), June 2018.
20. B. Agulló, A. Matamala, "(Sub)titles in cinematic virtual reality: a descriptive study", Onomázein, Vol. 53, 2021.
21. B. Agulló, P. Orero, "3D Movie Subtitling: Searching for the best viewing experience", CoMe, 2, pp. 91-101, 2017.
22. D. González-Zúñiga, Diekus, J. Carrabina, P. Orero, "Evaluation of Depth Cues in 3D Subtitling", Online Journal of Art and Design, 1(3), 16–29, 2013.
23. C. Mangiron,"Reception of game subtitles: an empirical study", The Translator, Volume 22, Issue 1, 2016.
24. A. Brown, et al. "Dynamic Subtitles: The User Experience", ACM TVX'15, Brussels (Belgium), June 2015.
25. C. Hughes, M. Montagud, P. tho Pesch, "Disruptive Approaches for Subtitling in Immersive Environments", ACM TVX 2019, Manchester (UK), June 2019.
26. P. tho Pesch, R. Bouqueau, M. Montagud, "White Paper: Recommendations for immersive accessibility services" arXiv:2005.03756v1, May 2020.

27. Asociación Española de Normalización y Certificación (AENOR), “Norma UNE 153010: Subtitulado para personas sordas y personas con discapacidad auditiva. Subtitulado a través del teletexto”, 2003.
28. H. G. Debarba, M. Montagud, S. Chagué, J. Lajara, I. Lacosta, S. Fernandez, C. Charbonnier, “Content Format and Quality of Experience in Virtual Reality”, ArXiv, August 2020, https://arxiv.org/abs/2008.04511
29. F. De Simone, et al. “Watching videos together in social virtual reality: An experimental study on user’s QoE”, IEEE VR 2019, pp. 890–891, Osaka (Japan), March 2019.
30. R. S. Kennedy, N. E. Lane, K. S. Berbaum, Mi. G. Lilienthal, “Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness”, The International Journal of Aviation Psychology, 3:3, 203-220, 1993.
31. Igroup Presence Questionnaire (IPQ), http://www.igroup.org/pq/ipq/index.php, Last Access in January 2021.
32. R. Bernabé, P. Orero, “Easy to Read as Multimode Accessibility Service”, Hermeneus, 21, 53-74, 2019.
33. E. Oncins, M. Montagud, R. Bernabeu, “Accessible scenic arts and Virtual Reality: A pilot study in user preferences when reading subtitles in immersive environments”, MonTI, 2020.
34. M. Montagud, J. Segura, J.A. De Rus, R. Fayos, “Towards an Immersive and Accessible Virtual Reconstruction of Theaters from from the Early Modern”, ACM IMX 2020, Barcelona (Spain), 2020.
35. S. Rothe, A. Schmidt, M. Montagud, D. Buschek, H. Hußmann, “Social Viewing in Cinematic Virtual Reality: A Design Space for Social Movie Applications”, Virtual Reality (Springer), October 2020.

MARIO MONTAGUD CLIMENT was born in Montitxelvo, Spain. He received a BsC in Telecommunications Engineering in 2011, an MsC degree in “Telecommunication Technologies, Systems and Networks” in 2012 and a PhD degree in Telecommunications (Cum Laude Distinction) in 2015, all of them at the Polytechnic University of Valencia (UPV). He has experience as a postdoc researcher at UPV and at CWI (The National Research Institute for Mathematics and Computer Science in the Netherlands). He is currently a postdoc researcher at University of Valencia and i2CAT Foundation (Spain). His topics of interest include Computer Networks, Interactive and Immersive Media, Media Synchronization and QoE (Quality of Experience). Dr. Montagud is (co-)author of over 100 scientific and teaching publications, and has contributed to standardization. He is member of the Organization and Technical Committee of many international conferences, and of the Editorial Board of international journals. He is currently involved in many European, national and regional projects. Webpage: https://sites.google.com/site/mamontor/

CRISTIAN HURTADO SANJUÁN was born in Real, Spain. He is currently finishing the B.Sc. degree in Multimedia Engineering from the University of València (UV), at the Escola Tècnica Superior d'Enginyeria (ETSE), and works as Software Engineer at Capgemini (València). As part of his B.Sc. thesis, he is collaborating with the High Performance and Intelligent Systems (HiPIS) group at UV in the design, development and evaluation of accessibility and multi-modal interaction solutions for immersive media.

JUAN ANTONIO DE RUS ARANCE was born in Bailén (Jaén), Spain. He received the B.Sc. degree in Multimedia Engineering from the University of València (UV), at the Escola Tècnica Superior d'Enginyeria (ETSE), and is currently pursuing a Ms.C degree in Artificial Intelligence, Pattern Recognition and Digital Imaging at Universitat Politècnica de València (UPV), both universities located in València (Spain). Juan Antonio is a developer and junior researcher and developer with the High Performance and Intelligent Systems (HiPIS) at UV. His work focuses on the design, development and evaluation of accessibility and multi-modal interaction solutions for immersive media. In this period, he has participated in many research projects, and has been co-author of two conference papers.

 doi:10.20944/preprints202107.0621.v1

SERGI FERNÁNDEZ LANGA is a computer scientist and holds a MsC degree on Artificial Intelligence by the Universitat Politècnica de Catalunya (Spain). Since May 2008, he has given support to the coordination of i2CAT's Media & Internet Unit and led i2CAT's involvement in national and European projects. Since July 2012, he is the director of that unit. He has co-authored several articles in the field of Interactive and Immersive TV, and he has coordinated different European projects, like TV-Ring (2013-2016) ImmersiaTV (2016-2018), VR-Together (2017-2020), and ImAc (2017-2020).