Politecnico di Torino

Politecnico di Torino

Laurea Magistrale in Ingegneria del Cinema e dei Mezzi di Comunicazione

Sessione di Dicembre 2021

Immersive movies: The effect of point of view on narrative engagement

Supervisor

Prof. Fabrizio Lamberti

Candidate

Antonio Castiello

Co-supervisors

Dr. Alberto Cannavò

Prof. Tatiana Mazali

Dr. Filippo Gabriele Pratticò

1 Introduction **3**
1.1 Introduction to cinematic virtual reality 3
1.2 Point of views in CVR 7
1.3 Narrative engagement 9
1.4 Research introduction 10

2 State of the art **13**

3 Technologies **22**
3.1 Shooting and post production 22
3.2 Video player 27
3.3 Eye tracking 28

4 Design **31**
4.1 Material research 31
4.2 Study definition 34
4.3 Material design 35

5 Realization **37**
5.1 Material production 37
5.2 Measuring narrative engagement 48
5.3 Eye tracking system 51
5.4 Experiments 63

6 Results and discussion **67**
6.1 Oreste è ancora vivo 67
6.1.1 Subjective results 67
6.1.2 Objective results 69
6.2 Persone che potresti conoscere 79
6.2.1 Subjective results 79
6.2.2 Objective results 80
6.3 Other results 81

7 Conclusions and future work **85**

References **88**

Abstract

Cinematic Virtual Reality (Cinematic VR, or CVR) gives a wide range of possibilities to experiment new techniques regarding movie editing, narrative and shooting. This includes both live action and computer generated movies. Many filmmakers proposed different solutions to tackle the above domain, but still a few scientific studies were made to investigate their efficiency and to analyze how different approaches affect our experience. As for traditional cinema, one of the choices that a CVR director has to make is to select from which perspective the user will see the movie, the so-called point of view (POV). Taking in consideration the actual products related to CVR, the types of POVs can be roughly split in *first-person perspective* (1-PP) and *external observer*. In the 1-PP approach, the user is a diegetic element of the story, a character or even an object; conversely, in the external observer approach, the user watches the scene from a detached POV, untied from any element of the story.

The aim of this study is to understand to what extent, watching a VR movie from a specific POV could actually affect the *narrative engagement, i.e.* the user's sensation of being immersed in the movie environment and being connected with its characters and story. The main hypothesis of this thesis is that the 1-PP version might be more engaging, especially in terms of emotional engagement and narrative presence. In the interest of isolating the POV as variable, two live action 360° videos were produced in two different versions. A user study was conducted where the participants, after watching one version of each video in VR, had to fill in a questionnaire aimed to evaluate the narrative engagement. Additionally, eye tracking data were collected to correlate the subjective experience with objective observations.

1 Introduction

1.1 Introduction to cinematic virtual reality

Virtual reality (VR) has been defined as the *ultimate empathy machine* [1] due to its ability to recreate immersive scenarios. However Janet H. Murray's opinion [27] seems to take a different direction, considering that empathy is not just a direct outcome of wearing a headset but it is necessary to find the right mechanism in order to unlock specific feelings into the viewer. Besides considering VR an *empathy machine* may be controversial, due to our limitations to objectively evaluate empathy, it can be said that VR has increased its appeal in the last few years. Many of the most important film's competitions, such as the Venice Film Festival, have already made a section for VR movies. The reason why the continuous growth of this media may lay on its new opportunities and new language. The possibility of walking around the streets of Tokyo or on the moon while you are comfortably sitting in your room is something that fascinates both content creators and consumers. The VR pioneer Nonny de la Peña defines VR as an experience that we can remember not just with our mind but with our full body [28]. As a journalist, her mission is to build a new form of storytelling in order to engage the audience into the story. For example her work *Project Syria* (https://bit.ly/2YXchNP, 16-09-2021) represents the violence during the civil war in Syria and how it especially affects the children.

On the other hand, there is also part of the population that is strongly against the development of this media. They consider the possibility that VR may substitute our daily life and in the future we could live only in a digital world.

However, besides the opinion about this media that may sound too romantic or too apocalyptic, it is fundamental to understand its power and effectiveness through scientific experiments. These studies cannot avoid considering human behaviour and the way we

perceive the stimuli given by VR. As Murray underlines [27] it is fundamental to look for a new language for this media through the creation of new contents and the comprehension of what really works.

Nowadays there are several options to make and then watch a VR movie. In order to create a VR content one can choose between:

- Computer graphics (CG): generate the movie entirely using software (and its relative plugins) for animating 3D elements; some examples are Blender (open source), Unity and Cinema 4D.
- Live action: recording a real world seen by using a 360° camera such as Go Pro Max or Samsung Gear 360.
- Hybrid: mixing the two techniques mentioned above. The compositing (matching together CG elements and live action) can be done with several software like Adobe After Effects.

All those techniques can recreate monoscopic or stereoscopic images. The difference is that the monoscopic VR showsjust one image for both eyes while the stereoscopic VR presents one image for each eye (so two images of the same object). The latter allows the viewer to perceive the depth of the scene since our brain can process the differences between the two images.

In order to watch a VR movie there are several equipment and platforms available at accessible costs (or even for free). Regarding the equipment there are three possibilities:

- Laptop display: online platforms such as YouTube and offline software such as VLC allow people to enjoy VR by using the mouse to move around the environment.
- Smartphones: on the various stores (Google Play, Apple's App store, etc.) there are several apps that allow people to see VR movies (like the Guardian or even

YouTube); in this case users can explore the scene by touching the screen or even by moving their phone (taking advantage of the gyroscope and accelerometer).

- VR headset: this kind of equipment is often considered the most immersive also because it allows people to explore the movie in a more natural way such as turning the head or even walking in the room; it is also the only way one can appreciate stereoscopic videos.

Regarding the latter category, headsets available can be subdivided by their complexity and relative costs. There are three options:

- Mobile phone VR headsets. They are equipped with special lenses and a space where to insert the smartphone. This kind of headsets exploits the movements sensor within the modern smartphone. They are usually the most accessible in terms of costs and there are thousands of models available on the market. One of the most famous is the Google Cardboard (Figure 1).

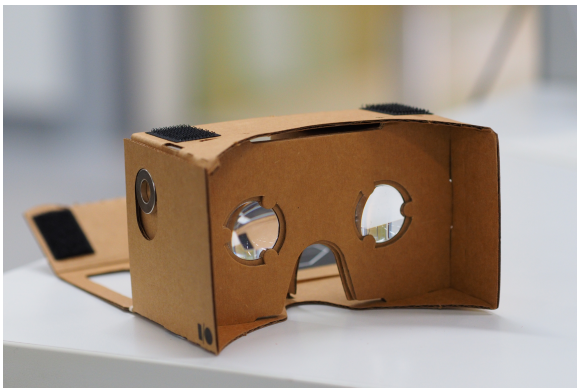


Fig 1: Google Cardboard

- PC VR headsets. This equipment needs to be connected to a computer (or to a game console) and they allow users to watch the VR content directly on the integrated display. An example is the HTC VIVE (Figure 2).

Fig 2: HTC VIVE

- Standalone VR headsets. Without the use of any additional equipment it is possible to upload and watch VR content or even take advantage of the gallery of videos that each model offers. An example is the Oculus Go (Figure 3).

Fig 3: Oculus Go

The last two types of headsets give additional immersive opportunities and consequently higher costs. In fact, PC and standalone headsets allow the viewer to enjoy six degrees of freedom (6DoF) experiences. This means that thanks to additional sensors and devices (camera and joypads), standalone and PC VR headsets can detect both rotational (head rotation) and translational movements (user walking into the room) of the viewer. Equipment such as cardboards allows only to detect the movements at 3DoF (mostly the rotational ones).

1.2 Point of views in CVR

As for many other media, such as novels and traditional movies, the POV represents one of the most important choices for content creators. It represents the perspective from which the viewer perceives the story and it may affect what the viewer experiences (sounds and images). As mentioned in [1], the viewer in CVR can be:

- an internal observer;
- an external observer.

The first kind of observer represents the view from a diegetic element of the environment. The camera is set at eye level of the character or on the position of the object into the scene: for example in movies such as *The party - A virtual experience of autism* (Figure 4) and *Car crash experience in VR: What is it like to be involved in a car accident in 360°* (Figure 5). In the first movie, the viewer can see how stressful a simple party is for those who are affected by autism. The viewer turns into a young girl and it can hear her thoughts and see how she deals with a crowd of people. In this case the viewer cannot see her body. In the second movie, the viewer can experience a car crash from a passenger's POV; the viewer can see the body of the protagonist and other characters interact with him or her.

The internal observer POV can be considered as a first-person perspective (1-PP) or even can be taken in consideration of the view from an object within the scene (like in movie *The harvest, https://bit.ly/3jx8rTG,* where the camera represents a diegetic element).

Fig.4: Screenshots from the movie *The party* available at https://bit.ly/3DAs0SL (19-07-21).

Fig.5 : Screenshots from the movie *Car crash experience in VR: What is it like to be involved in a car accident in 360°* available at https://bit.ly/3kEJy7F (19-07-21)

In many 3DoF 360° videos, the first person is used especially for simulation (like the car crash, Figure 5) or to try putting the viewer into the shoes of someone else (like a one year old baby or a person with autism, Figure 4).

The external observer POV still immerses the user in the environment, but the scene is completely watched by external eyes (basically the camera). The camera is settled in the position that the director considers the best for watching the movie. In this case, the camera can also make some movements already seen in traditional movies, such as dolly shots. The viewer can be considered as a non-diegetic element of the scene or even a third-person perspective (an example is the movie *HELP*, shown in Figure 6).

Fig.6: Screenshots from the movie *HELP* available at https://bit.ly/3BsFjDa (19-07-21)

The movie *Across the line* (Figure 6) uses both these kinds of POVs. In the first part, in fact, the viewer is settled between the two speakers (external observer); in the second part, instead, the camera follows the POV of the protagonist (internal observer).

Fig. 6: Screenshots from the movie *Across the line* available at https://bit.ly/2YeepjZ (31-08-21). In the first part (live action), on the left, the scene is presented with an external perspective. In the second part (CG), on the right, the scene was recorded with a first-person perspective.

The internal observer POV is often considered as a technique that gives more involvement. Even though this is a common practice for filmmakers to give a more immersive experience, it is not really clear how a different POV can affect the viewers' perception of the events in VR. Chapter 2 exposes previous studies and the actual research status regarding this issue.

1.3 Narrative engagement

Csikszentmihalyi defined narrative engagement as the sensation of flow that can arise from any activity in which people strongly focus their attention, lose awareness of themselves and perceive the activity to be part of themselves [20]. It represents an enjoyable experience triggered by any narrative content such as a book or a movie. Evaluating how different techniques influence users' engagement is important for the benefit that this state of flow has on their lives. Researchers confirm that engaging content can create benefits in routine and a temporary relief from daily life difficulties [21].

Busselle and Bilandzic [20] stated that engagement is the consequence of a mental representation of the story made by the viewer followed by a complete attentional focus on the story itself. This representation is made by recreating different aspects of a narrative experience, called mental models. Busselle and Bilandzic [20],[21] considered three mental models related to a story:

- the situation model which includes the story itself (plot) and all the connection between the actions and events of the story (the causal link);
- the character models which represent all the possible information regarding the protagonists of the story such form physical to psychological;
- the story world model which consists in all the information related to the narrative world (logic and rules included).

In order to comprehend a story and to get into the state of flow, viewers have to change their perspective and get into the mental model of the narrative environment [20].

1.4 Research introduction

The purpose of this work is to explore human behaviour during the view of an immersive movie. Specifically it focuses on how the camera perspective (the POV) affects the sense of transportation and absorption that an immersive movie generates into a person. These kinds of sensations have been synthesized in a metric called narrative engagement (Paragraph 1.3). Two different POVs were analysed as they were the most used in previous CVR contents: external observer and 1-PP (Paragraph 1.2). They were both used for the shooting of two immersive dialogue scenes so as to obtain two different versions of each scene. Once the two scenes and their two versions were produced, a 360° video player integrating an eye tracking

system was implemented using Unity. Finally, a group of people was recruited in order to participate in a user study. The test consisted in watching only one version of each scene and answering a 12-items questionnaire aimed to evaluate the narrative engagement. During each experience, eye tracking data were collected for further analysis related to the attentional focus. The data collected from the eye tracking system and the 12-items questionnaire were processed and statistically analysed in order to elaborate the results of this study. The results showed substantial differences between the 1-PP and the external observer POV; most of the users preferred the 1-PP since it really gives a new experience compared to watching a traditional movie. Furthermore, the 1-PP was able to create a greater sense of immersion and emotivity.

The general steps taken during this work are the following: material research, study definition, material design, test implementation and data analysis. The material research phase included an in-depth analysis of scientific works related to VR (Chapter 2) and the view of immersive movies produced by professional and amatorial authors (Chapter 4, Paragraph 4.1). The work done has been reported in the seven chapters of this document. In particular, Chapter 1 aims to contextualize the main elements that were tackled in this thesis. These elements include a general overview of CVR, the use of different POVs in immersive movies, and the narrative engagement (the main metric used in this work). Chapter 2 contains the state of art related to the research of human behaviour in immersive movies; it includes a report on various publications selected from notable scientific literature.

Chapter 4 shows the design process of the experiments related to this thesis. This process includes research regarding the actual productions of VR movies and the different platforms where it is possible to watch them; furthermore Chapter 4 explains how the idea of this study was developed and how the needed materials were designed.

Chapter 5 explains how this work was practically implemented; specifically, it contains the explanation of the material production process, how the narrative engagement was measured, and how the experiment was structured.

Chapter 6 presents the statistical analysis of data collected during the experiments and related results. Furthermore, a brief discussion with several considerations is given in order to compare the objective results of the experiments and the hypothesis made in this work.

Finally, Chapter 7 contains the conclusions of this work, including what was achieved, the limitations and possible future works.

2 State of the art

In the literature, a large part of the studies related to CVR have focused on how to guide the users' attention. It is fundamental to understand humans' attentional behaviours while watching a VR content for two reasons: to allow the content creator to make an enjoyable experience, and to reduce the *Fear of Missing Out* (FoMo) [2] which the user feels during the view. FoMo is defined *as a pervasive apprehension that others might be having rewarding experiences from which one is absent* [2]. This can cause frustration and negative feelings in the user that doesn't realise if something was missing or not. The study in [2] suggests the possibility to reduce the horizontal view from 360° to 180° or 225°. Even though this solution may help the user to get less stressed, the results confirm that a 360° view gives a higher feeling of immersion, reality and of being at the center of the scene. The same study suggests the possibility of limiting the view just for certain kinds of movies and for the kind of experience where rotating your head and your body isn't comfortable. The possibility to switch the kind of view and experience depending on "the mood" of the user and the "vision condition" could be additional features of VR movies in order to meet the needs of different users.

Even though focusing on the user's attention in VR is not the main purpose of this study, these works have been useful to set specific guidelines for the production of this study material and avoid any element that might affect its results. Furthermore, part of the collected material is linked with the attentional focus. Guidance methods can be divided into diegetic and non-diegetic [3]. The former are an integral part of the movie, for example moving characters, lights or sounds [3]. The study developed in [4] focused on these methods, specifically on spatial sound, lighted objects and moving elements. A movie with four different scenes was presented to a group of users and each scene had a specific kind of cues.

As a general outcome, objects and lights in movements seem to be effective guidance methods whereas static lights are not.

Regarding lights, it is important to also consider the source direction of these stimuli. In fact, the peripheral view is more sensitive to light stimuli whereas the central view (fovea) is more sensitive to color, movement and shape stimuli [5]. Furthermore, elements with sound, even if not spatial, showed higher efficiency to attract the attention as lighted elements. One of the limitations of this study was the approximation of the gaze eye to the head direction. The work in [6], in which more than 100 testers were involved and both gaze and head orientation were considered, confirmed one of the outcomes of [5]: the higher explorative behaviour at the beginning of each scene. Pillai produced for this study a 360° movie (3DoF) called "Dragonfly" and showed that the eye gaze of the character's in the movie can also be an effective cue for attention. However the diegetic cues have lower effects when the target (region of interest) is not in the user's field of view [3]. Non-diegetic cues in this case can be more effective; they usually include indicators, such as arrows, radars and outlines, or even image manipulations, like fade to black, desaturation and blur. In [10], a user study involving 30 participants showed the effectiveness of two types of indicators, i.e., bubbles and shadows. They are defined as *social indicators* since the area where they try to attract the user were processed in advance thanks to the gaze data of other viewers. The bubble indicator was a simple circle; if the gaze recorded had a low variance, the area of the circle was smaller in order to focus on a more specific part of the image. The shadow used to darken the edges of the image in case a viewer was looking at the opposite part of a popular focus area. This study showed the efficiency, but at the same time pointed out the importance of balancing between self-exploration and guiding. It suggested to estimate a specific threshold for the indicator's frequency in order not to make these methods too intrusive. Further studies must be done to understand how to estimate these frequency values. The study in [33] evaluated

additional guidance methods: an arrow, a radar (non-diegetic), and a butterfly (diegetic). All the three methods were effective in supporting the user's search of the salient point (especially the arrow), compared to watching the images without any guide. However this study was still focused on static images and the analysed methods limited the audience's freedom to explore the scene. In [42], other six kinds of non-diegetic cues were proposed to highlight certain elements of a given scene. These methods included label (simple text tag), box, mesh, outline, halo (additional lighting) and frame cues. These methods were explored with the additional possibility to include them as interactive opportunities for shot transitions; for example, whenever a user clicked on a specific label a new scene could start. The outline cue seemed the most efficient in terms of guiding the user's attention on static objects. However, these cues might be inadequate for movie genres that include more complex shots, actions and dialogues. All the mentioned cues were in fact investigated in static 360° images or basic 360° videos (such as street views).

An additional issue to the human behaviour in immersive movies concerns the watching condition and the audience range of motion. What if someone would like to enjoy an immersive movie while he or she is on a train? What if someone would like just to lay on the sofa and watch a VR movie? These situations represent different experiences from what one may usually think about VR and immersive experiences. However, the spreading of headsets and immersive contents at the mass-market level suggest considering these kinds of situations. Watching an immersive movie while standing or sitting, while travelling or just chilling on a sofa opens different scenarios that must be faced. Rotating the head while watching an immersive movie could be caused by the wish to explore the environment or by the need to feel more comfortable while sitting. The latter situation was the topic of one of the studies by Travis Stebbins and Eric D. Ragan [37]. Their work focused on how to recreate a valid method of auto re-alignment between the ROI of the scene and the user's view. This

approach would bring the focus of the scene to follow the user's head movement. The study purpose automated rotational adjustments of the view but still no solid values regarding an acceptable and not-intrusive rotation speed and neither specific range of thresholds (angle and delay) for when these adjustments may occur. As VR allows an interactive experience, another idea might be the production of movies that can dynamically change the turn of events depending on the user's behaviour. In Cinévoqué, for example, the storylines are triggered by the user interaction (mainly visual interaction) with certain points of the environment called hotspots [38]. Focusing on a different hotspot will create a different scenario to the user from a predefined set. A prototype has been realised but the system needs further developments and adequate audio-visual material to test it. Within the watching condition it must include the type of device immersive movies are watched through. In [39], it is explored the use of three different devices for watching 360° videos: a standard 16:9 TV, SurroundVideo+ and an HMD. The work reports on a user study involving 63 participants who watched four 360° videos using only one of the mentioned devices. The videos included a music clip, a documentary, a horror and a narrative film. Only the participants with the HMD display were allowed to explore the 360° panorama and decide which portion of the video they could watch. For the other two devices, the field of view was redirected in order to show the main element of the action in the center of the screen. According to the different metrics evaluated, the HMD showed important advantages in terms of spatial awareness and enjoyment. The type of device might also affect the sense of immersion and the emotional response. The 38 participants of the study in [40] were asked to watch a horror movie using only one device between an HMD and a traditional screen. The HMD group statistically demonstrated a higher level of immersion; for the emotional response, even though the HMD group showed better results than the no-HMD one, no statistical significance was shown.

This might suggest that the connection between immersiveness and emotional response should be explored with more different genres of movies.

Another condition that can affect a VR experience is the posture of the user. The two main different postures for watching a VR content are: sitting or seating. The work in [41] explores this issue from different aspects. In terms of safety, indeed a seated position can represent a better solution for the user as being seated reduces the risk of falling or even hitting an obstacle. A feeling of safety might implicate greater cognitive performance; since the user is less worried about his or her movements, a sitting posture could better allow focusing on cognitive tasks. However, there is no study that confirms this hypothesis. The choice of one posture should be dependent on the kind of VR experience; however the sitting position represents the most accessible option for a wide range of users (including people with certain disabilities).

The actual shooting methods in CVR are mainly based on the experience of VR content creators and further studies should be performed in order to prove their effects on the audience [7]. Even for traditional cinema studies concerning how the POV affects the narrative engagement are not so common. In [11], through an experiment with 126 users, it was observed how shot scale, shot length and perspective affect empathy, arousal and narrative engagement in violent movie scenes. The scenes were all taken from Quentin Tarantino's films (original audio with german subtitles). The internal observer perspective seemed to have a negative effect on narrative engagement, whereas the external observer perspective gave higher values of engagement. An opposite result was reported in [12] and [13], where subjective camera perspective gave a higher sense of presence and more frequent arousal response. However, in the latter cases the audio-video material presented to the user was sport represented by football matches. This result may suggest that the kind of material has a strong influence on the effect of the camera perspective. The evidence of the relation

between the kind of content and camera perspective effect was seen also in VR. In a survey conducted in [9] it was seen how most of the people preferred to watch VR video streams in the first-person perspective rather than in a third-person perspective (3-PP). The 3-PP is when the camera is settled behind the back of a character and follows all his or her movements; this setting is quite common in shooting games. However, the results discussed in [9] showed that the users' preference was affected by the kind of video game that was considered. In order to avoid such situations, in the study reported in this thesis there will be two versions of the same story. Furthermore, the focus will be on cinematic content and not on gaming.

Additional studies regarding 1-PP and 3-PP in VR videogames were presented in [15] and [16]. In [15], 24 subjects were presented with a simple game where a ball had to be moved through a specific path. The POVs used in this study were four: first-person, over the shoulder, behind the back and from top to bottom. When the players were watching the game through the 1-PP, they showed better scores; this finding seems to suggest that for guiding an object in a VR environment, the 1-PP may be more efficient. In [16], it was seen that the 1-PP provided better results regarding embodiment; in this work, subjects could control an avatar in an immersive environment through 1-PP and 3-PP. Both perspectives gave high results in terms of spatial presence; while 1-PP enabled a better experience in terms of interaction, 3-PP gave higher values of spatial awareness. These studies considered an immersive environment with 6 DoF, whereas the work in this thesis will consider only 3 DoF. Furthermore these studies were developed in a game VR environment which is basically experienced with a higher level of interaction than the movie that will be used in the hereby reported study. The work in [17] is related to the use of two POVs in CVR. Using the movie *UTurn*, a novel cinematic 360-degree video (http://www.uturnvr.com/) filmed from the perspective of two characters, this study focused on the viewers' motivations to select a POV, and their subsequent behavior. Viewers could switch between the two characters' POV just

by moving the gaze (the 360° environment is divided into halves). As first outcome of this work, it was shown that viewers' preference for a POV and reasons for their preference influenced how they watched and recalled content. Both perspectives in this case were 1-PP, but of two different characters, a male and a female. The subjects who chose to start to look from the POV of the male character and those who did not have any preferences switched between POVs with higher frequency. Subjects who selected the female POV at the beginning made less changes between POVs.

The work reported in this thesis will instead look at the differences between a first-person and an external observer perspective, focusing on the impact on the narrative engagement.

Another topic that took the interest of different researchers is related to editing techniques for VR movies. In 2017, it was published one of the first studies that looked into the perception of continuity in VR movies [24]. For the *event segmentation theory* [34], the human brain gives sense to what is continuously happening around by splitting it into discrete events. A new event is registered whenever something changes in action, time or space. This theory seems to be also connected with several editing techniques of traditional movies and the way humans perceive consecutive multiple shots. The study in [24] analysed different variables such as the type of edit, angular distance between the salient point of two adjacent shots, the number and the position of salient points in each shot. Its outcome implies that humans use to recognize different scenes in VR with the same cognitive process as we do in traditional movies. Those processes in VR mainly include discontinuity in actions, followed by temporal/space discontinuity. This means that whenever the viewer uses one of these types of discontinuity, he or she will perceive that the scene has changed. Regarding the misalignment between the focus point of two consecutive scenes as outcome, it seemed that a higher value would be an incentive for the users' exploration and help to recreate different feelings (such as suspense).

In [25], it was introduced the concept of *spaceline* as a new method of editing shots in immersive movies. The idea is to set the different cuts of the movie, not on a timeline (as it happens in traditional movies), but using specific areas of the space. Each scene will contain three different areas:

- out-region: point of each scene that indicates the end of a schot and the beginning of the transition to a new shot;
- in-region: starting point that the user sees after each shot transition;
- act-region: area of the shot where the user can look to have more details;

However there is no study that went through this method and analysed whether it is really valid.

A recent work [25] investigated the audience's attitude to different movie cuts in a professionally edited movie, The People's House, created by Felix & Paul Studios. The importance of this work is that it is one of the first studies that included a massive participation of users (more than 3000) and a professional 360° video content. A first outcome was that two consecutive cuts, where the first is more explorative (without any specific salient point) and the second has a well defined point of interest may help to guide the user's attention. The reason why this may occur is that also in traditional movie cuts it is common to have a first shot that includes the whole environment (so that to easily frame the situation) and then a cut where the main point of the situation is shown. Additional studies have to be carried out with different movie genres which include more dynamic scenes and dialogues.

Three different scene transitions techniques were studied in [36].

- Teleportation: the view changes instantly due to rotation and/or translation movements. The user does not see any movement in action. Two different variants

were analysed, one with a fade to black and one with an immediate change of the view.

- Animated interpolation: the view is slowly changed from a point to another in a way that the user can follow it. This type of transition was studied using three different speeds: slow (10 m/s), medium (25 m/s), and fast (50 m/s).
- Pulsed interpolation: the view changes are developed through a sequence of interposed viewpoints. This method was analysed with three different numbers of steps: 2 steps (low), 3 steps (medium), and 4 steps (high).

Each editing technique was analyzed in terms of spatial awareness, motion sickness and audience's satisfaction. Even though all the three methods did not show that much difference in terms of spatial awareness, the animated interpolation seemed to give the best results. Furthermore, the speed of view motion did not seem to affect spatial awareness; this statement requires further studies, according to the authors.In general, spatial awareness resulted lower when there were rotational and translational changes of the view. Even though animated interpolation gave the best results for spatial awareness, it was also the technique that had the worst consideration in terms of motion sickness. This work suggests that teleportation is the best solution in case of rotation of the view without any transition.

The effect of editing was studied in [18] using, among other metrics, the narrative engagement. In the study, it was evaluated the effect of different transitions (the montage between the shots) in VR movies. The considered techniques were: simple cut, fade and VR portal. The latter is a physical portal in the virtual environment: once the viewer steps in, he or she would see the next shot. This early study did not find any statistical evidence that the narrative engagement is affected by the transitions.

The work presented in this thesis is probably the first one to focus on the effect of camera perspective on narrative engagement for 360° videos with 3 DoF.

3 Technologies

In this Chapter is presented an overview regarding which tools, software and hardware, were used during the different phases of this work. For each tool is included a short description of its functioning and the possibilities of use.

3.1 Shooting and post production

The camera used to shoot the scenes was a GoPro Max equipped with two ultra wide-angle lenses and six microphones. This camera has the possibility to work using only one lens (HERO Video mode) or with two lenses (360° video mode). A small touch display is integrated and besides the preview of what is seen by the camera, it is useful to change different settings such as ISO, frame rate and exposition time. Using the 360° video mode, it is possible to shoot up to 6K of resolution and 30 frames per second (fps). However, after the stitching process, the resolution lowers to 5.6K since the field of view of each lens is partially overlapped. Thanks to a WiFi-Bluetooth connection it is possible to remotely control the camera with a smartphone and the GoPro app (Quick). Additional features are the functions Max HyperSmooth Video Stabilization and Horizon Leveling which help to delete small camera movements and align the horizontal axis of the camera with the horizon level.

As additional shooting equipment, two different GoPro supports were used: a head strap and a tripod. The latter is simply a stick of variable height (up to 1,75 meters) with three legs; the connection between the camera and the tripod was done thanks to a GoPro special adapter and its screw. The head strap (see Figure 7) is a set of elastic bands that permits placing the GoPro above the forehead of a person.

Fig. 7: GoPro head strap

A Zoom H5 handy recorder (Figure 8) was used for additional voice recording. This equipment includes two cardioid microphones combined using the X-Y technique; the two capsules are really close, forming an angle of 90° (directional microphone). This recorder permits various recording formats (stereo and multitrack mode) at different sampling rates (44.1/48/96 kHz). It is possible to add additional inputs (such as additional microphones) and to directly listen to the audio recorded thanks to a small speaker or using headphones as output. It has a small display (128×64 pixels) that allows checking the recording status and the various microphone settings.

Fig.8: Front view of the Zoom H5 handy recorder.

A Led Fresnel spotlight HWAMART FB-800G (Figure 9) was used to illuminate the set of the scene *Oreste è ancora vivo* (Chapter 5). This light has a maximum illuminance of 1200 lux and two possible colour temperatures 3200, and 5600K.

Fig.9: side view of the Led Fresnel spotlight HWAMART FB-800G

During the shootings, in order to document the backstage and all various settings, several photos were taken using a Sony Alpha a7 III digital camera with a 28-70mm zoom lens. This is a mirrorless camera with a 35mm full frame CMOS sensor.

The post production (audio and video) process was entirely done on a laptop Dell Precision 3551 with Windows 10 Pro 64 bit as operating system, an Intel(R) Core(TM) i7-10750H CPU, 16 GB of RAM and a Nvidia Quadro P620 GPU.

The software used in the post production phase were GoPro Player, for stiching and exporting the videos, Adobe Premiere 2020 to do the editing and Adobe After Effects to correct the images. GoPro Player is a free software created by GoPro which allows stitching, cutting and exporting 360° videos. The stitching is basically the process of merging together different images in order to obtain a unique image. In 360° videos (or even photos), it consists in merging together the two images captured by the two lenses and creating an equirectangular image (an image where the width is twice the height).

Adobe Premiere Pro 2020 is a professional video editing software, part of the Adobe Creative Suite. It is a timeline-based video editor, which means that every element (image, video and audio) is treated as part of a temporal line where the main unit is represented by the frame. It is possible to work on multiple channels both for audio and video; this means that at the same frame there can be two images on two different channels and the priority, in terms of visibility, is given to the one that is located on the higher channel (top-bottom priority). This software allows working with a large set of video formats in terms of resolution and codec. Besides the editing, Premiere Pro has sections and features for color correction and a wide library for visual and audio effects.

It is possible to edit 360° videos using the VR panel with a set of features and effects dedicated to immersive content (Figure 10).

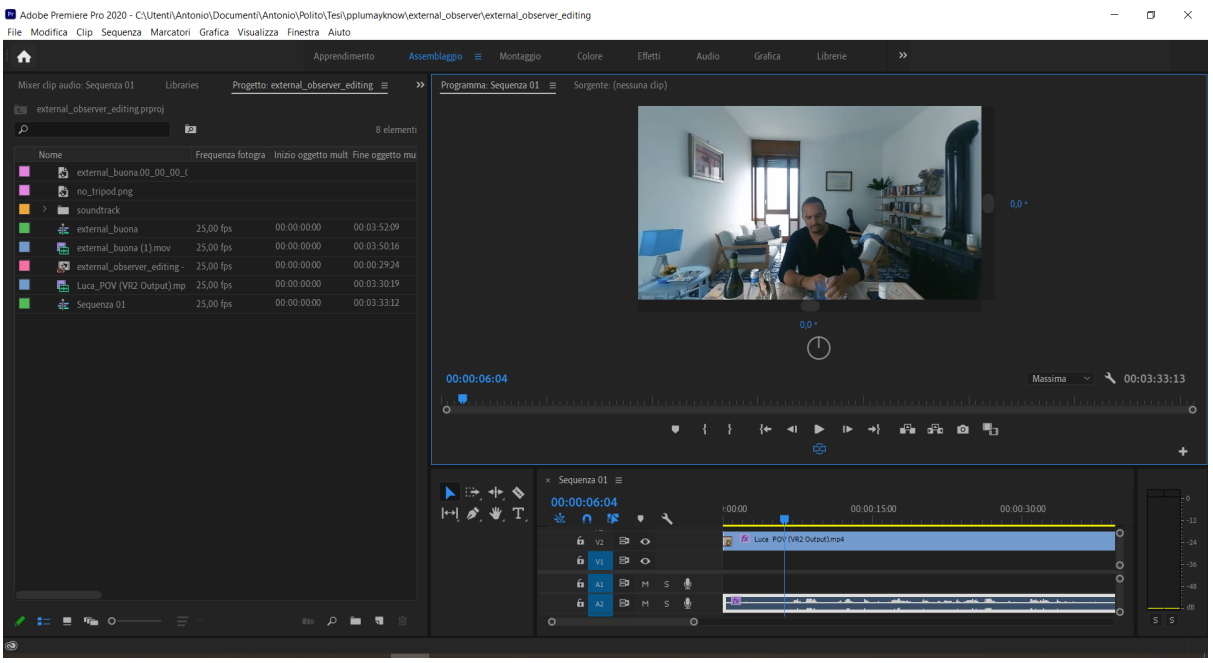


Fig.10: Screenshot from the editing process of a 360° video in Adobe Premiere Pro.

Adobe After Effects 2020 is another professional tool from the Adobe Creative Suite dedicated to visual effects, motion design and image compositing. As Premiere Pro, it is timeline-based and allows working with different video codecs and resolutions; furthermore,

it supports 2D and 3D assets in the same project. An After Effects plugin called VR Comp Editor permits to track the camera of a 360° video and re-orient the image on the image by controlling the rotation angle along the three axes. Furthermore, once the camera is tracked, the VR Comp Editor allows the stabilization of its movements. This process reduces abrupt movements that might occur when the camera, during the shooting, is not located upon a fixed support (such as when it is placed on the head of a character). Another possibility that After Effects offers is color correction and color grading. Color correction represents the general process of adjusting the colors, the withe balance and the exposure of the light for each frame and scene of a video. The color grading, instead, is an operation that takes place after the color correction and represents the process of editing the colors of an image in order to give a specific look. Thanks to After Effects it is possible to carry out both the processes using integrated tools or even different plugins.

Other tools from the Adobe Creative Suite used for this work were Adobe Illustrator 2021, Photoshop 2021 and Media Encoder 2020. Illustrator is a professional vector graphics software used for a wide variety of digital designs. Most of the graphic assets needed for this work were realised through this software. The second is a software for editing and manipulating digital images. For this work, Photoshop was used to create static image layers in order to cover unwanted parts of the video (such as the tripod's legs). Finally, the latter is a software used for encoding and compressing video files in different formats.

Audacity, a free and open source digital audio editor, was an additional tool used for audio post-processing. Besides supporting audio post processing, this software allows recording from multiple sources.

The screenplays of the two scenes were written using KIT Scenarist. This software supports the authors in the writing process of a script by formatting the text in accordance with the industry standards, creating statistics regarding the story (number of words, how many times

a character appears, etc.) and providing different export formats. Furthermore, it includes various tools which help the authors in their creative process. For example, the card module permits summarizing a story on a cardboard where each card represents a part of the story. The cards can be sorted and colored in different ways in order to give a proper visual impact.

3.2 Video player

A video player for 360° video was developed using Unity (version 2020.3.17f1) (Figure 11). Unity is a cross-platform game engine developed by Unity Technologies, launched in 2005. It permits to create and render 2D and 3D environments with the possibility to import assets from various softwares such as Blender and Maya. Besides enabling the production of video games for different platforms, in recent years it has become one of the most used tools for the realization of VR products. For API scripting, Unity uses, as primary programming language, C#.

The mechanism that permits the view of 360° videos in Unity can be assimilated to place on a sphere an animated texture and to position the viewer inside the sphere. In order to correctly project the texture (and so the video) on the sphere, a preliminary process called UV unwrapping is required. This process consists in creating a flat (2D) representation of a 3D model. Unity supports two ways to unwrap 360° videos: 6-frame (Cubemap) and equi-rectangular. Which one is to be used depends on how the video was shooted or rendered. In the case of this thesis, the unwrap format was the equi-rectangular one, since this is the format used by the GoPro Max camera.

In the Unity 3D project, three main elements were added to an empty scene in order to set up the 360° player: a video player, a render texture and a skybox material. The video player is an element that permits the visualization of video content onto a target; this target can be a camera, the background, or a render texture. The latter are types of textures used in Unity for

different purposes related to visual effects such as dynamic shadows and projectors. Finally, a skybox material represents a 6-side cube that is shown behind all graphics assets in the scene. Frame by frame the video player permits to project on the render texture the 360° video. Using this texture for the skybox material, it is possible to display the video as a panorama of the Unity scene. A detailed description of all the steps to realise a 360° video player with Unity is available on the learning section of the Unity website (https://bit.ly/360videoplayer) . Once the scene is played, in the Unity game section it is possible to watch the 360° video and move the camera according to input system settings (for example with a mouse or a VR headset).

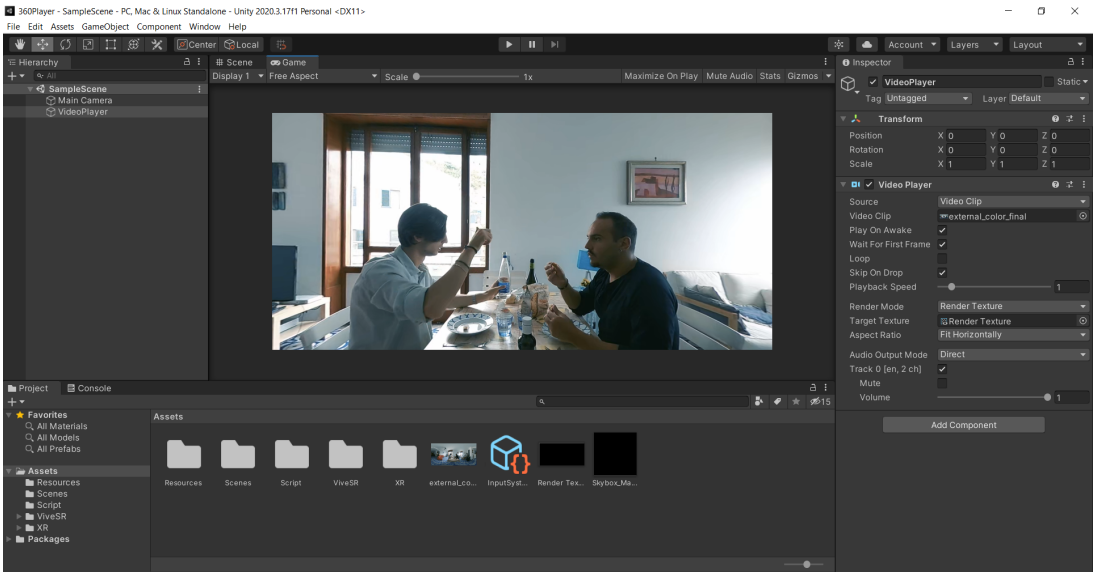


Fig.11: Screenshot of the 360° Video Player realised with Unity.

3.3 Eye tracking

In order to show the videos and record gaze data it has been used as hardware a HTC Vive Pro Eye (Figure 12). This VR headset has a Dual OLED screen (3.5” diagonal), a resolution of 1440×1600 pixels per eye (2880×1600 pixels combined), 90Hz as refresh rate and 110° field of view. For the audio, it also includes a removable set of headphones.

Fig.12: HTC Vive Pro Eye

For the position and rotation tracking of the user, this device includes different sensors such as a G-sensor (accelerometer), a gyroscope and a proximity sensor. Furthermore, thanks to the SteamVR tracking sensors and two additional base stations it is possible to track the position of the user relative to the room up to a distance of five meters (between the stations and the headset). For the eye tracking, the headset includes two cameras, one per eye. The headset is able to collect gaze data with a maximum frequency of 120 Hz in a tracking space of 110°; the accuracy is between 0,5° and 1,1°. The set of data available through the eye tracking system are: timestamp (device and system), gaze origin and direction, pupil size and position, and whether the eye is blinking or not. The collection of these data is possible through the HTC SRanipal SDK interface which is compatible with Unity. The latter was also the software part of the eye tracking system. An additional package available for Unity, the Tobii XR SDK, was used;, besides being compatible with all the functionalities of the SRanipal interface, this package provides additional APIs and functions for processing eye data. Specifically, it includes different tools: Tobii Ocumen, Tobii Ocumen studio and Tobii G2OM. For this study, it was used only the Tobii G2OM (Gaze-2-Object-Mapping), a

machine-learned selection algorithm able to accurately detect where the user is focusing. This was a useful tool to understand even at runtime (while the video was playing) which object took the visual attention of the users.

Another plugin from the Unity Asset Store called Record and Play (https://bit.ly/recordandplayEliDavis) was used in order to assure a stable sampling rate of the gaze data during the tests. At runtime, Unity does not guarantee a stable frequency of the different operations specified in the various scripts, hence the use of a plugin is necessary when it is fundamental to have a fixed sampling rate. This asset permits recording, with a given frame rate, rotational and positional movements of the Unity game objects while the scene is playing. Additionally, it allows saving custom events, setting further meta data and doing the playback of the recorded animation; lastly it is possible to export the saved data in different formats such as JSON and CSV.

Further details regarding how the eye tracking system was implemented are given in Chapter 5 (Paragraph 5.3).

4 Design

In Chapter 4 it is described the design phase which has followed a first research on the actual production of immersive movies; subsequently it was defined the purpose and the structure of the study related to this thesis and finally it was outlined how to produce the needed material.

4.1 Material research

In order to develop this study, the first step was to get familiar with CVR by watching pre-existing VR movies. This material was found on different platforms that offer the possibility to watch 360° videos at 3DoF; the main sources were YouTube (where it is also possible to find amateur materials), The Guardian App, Google Spotlights Stories and Within App. This section reports some examples which have been particularly interesting both for their quality and for the aim of this study.

Considering first animation movies, it is a must to cite those by Baobab Studios. Their short movie *Crow: The legend* (Figure 13) describes the story of a bird with a magnificent voice and colors, that will be sacrificed to save the rest of the animals from the cold winter (Figure 7). The user watches the movie as an external observer that follows the main action. In the movie *INVASION!* (Figure 14) instead, the user can be considered as an internal observer of the action; in fact, a little rabbit at the beginning of the movie comes close to the viewer and smells it. The rabbit role seems to be also to guide the user's attention with its own gaze.

Fig.13: Screenshots from the movie *Crow: The legend*.non-interactive version available at https://bit.ly/3gLD4mi (31-08-21).

Fig 14: Screenshots from the movie *INVASION! available at https://bit.ly/3DwrVzE (31-08-2021).*

The movies available on Google Spotlight Stories seemed to privilege the external observer perspective. In the sci-fi/action movie *HELP*, a mix of live action and CG generated content, (available at https://bit.ly/3mN9twB) an alien spreads panic in the middle of Los Angeles. The user can follow the action thanks to the continuous movement of the camera similar to a sequence shot. Another movie from Google Spotlight Stories is *Pearl* (Figure 15): it shows the relationship between a father and his daughter. The whole story takes place inside a car and the camera (as the observer) is positioned on the front seat, passenger side.

Fig.15: Screenshots from the movie Pearl available at https://bit.ly/2WAEwRh (31-08-2021).

The use of a 360° camera as a car cam has seen multiple usages also in live action VR movies. On YouTube there are many examples of VR experiences inside a car, as the one mentioned in Chapter 1 (*Car crash experience in VR: What is it like to be involved in a car accident in 360°*). Most of them privileges the 1-PP in order to let the viewer live the experience as one of the passengers; this kind of content may include driving tests or even car crash simulations. Other live action material, made with a 1-PP, include, for example, the movies *First impressions* (Figure 16) and *VR patient experience* (Figure 17). The former represents the simulation of how a baby perceives the world and how the other people interact with him or her. In the latter it is shown the POV of a patient in a hospital. In both the movies the user can see the other characters interact with him or her but only in the latter he or she can actually see the body of the protagonist (represented by a mannequin).

Fig.16: Screenshot from the movie First impressions available at https://bit.ly/3kBC4T3 (31-08-2021)

Fig.17: Screenshot from the movie VR patient experience available at https://bit.ly/38oJMdz (31-08-2021)

An internal observer POV and especially the 1-PP is often used in products that aim to elicit empathy or recreate a simulated experience. However, the 1-PP is not the only option used in

live action VR movies with a high emotional impact. In the VR drama *In the shadow* (available at https://www.youtube.com/watch?v=janQg_aulXc), the authors show the lives of human trafficking victims using the perspective of an external observer. Also the horror-drama *Dinner Party* (available at https://bit.ly/3gNDUPn), inspired by a true story, is presented with an external observer POV.

The mentioned material plus other content were a useful source in order to understand the actual methodology, languages and trends among CVR creators.

4.2 Study definition

As the study of human behaviour in viewing immersive movies gives a wide possibilities of research, it was necessary to define which path this work had to take. As said, he POV was chosen as the main variable for this work. A few studies related to this element and the way it affects the user experience in VR were found. Furthermore, as mentioned in Chapter 2, in [11], [12] and [13] it is seen that this element has an impact on the viewer's experience in traditional cinema. Since in these studies it was observed that the kind of movie may influence the impact of this element, it was chosen to focus on a specific type of scene: the dialogue scene. The reason is that dialogues are quite common in traditional and immersive movies. Once it was decided the main variable and the type of content, it was selected as the main metric of this study the narrative engagement (NE). The NE is considered a good evaluator of a narrative experience since it includes emotional, attentional and cognitive parameters. In addition, the NE was previously adopted as a metric in other studies (both immersive and traditional cinema, [11], [12], [13], [18]) and it has a clearer definition and more reliable evaluation methods [19] than other metrics such as empathy [22].

4.3 Material design

In order to isolate the perspective as a variable from the type of story viewed by the users, two scenes were designed and produced in two different versions: 1-PP and external observer. One represents the view of one on the character, the other the perspective of an observer that is not an element of the scene (Figure 18).

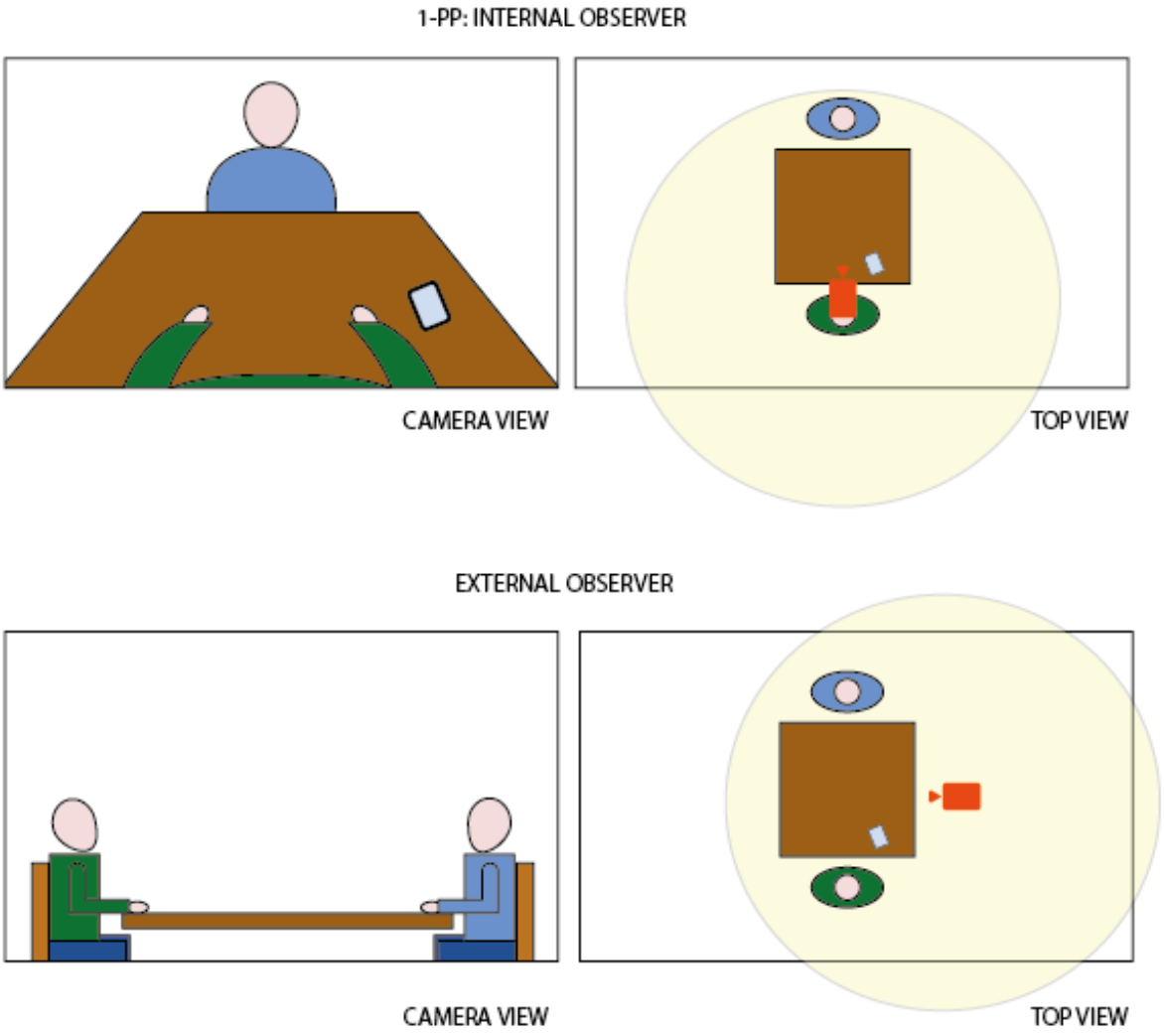


Fig.18: Representation of the camera view and floor plan for the 1-PP POV (top) and the external observer POV (bottom)

In this phase several aspects were considered which may affect the study's results.

- Actors' gaze and movements: these stimuli can be a method to guide the visual attention of the viewer [6]. In order to avoid possible ways to generate any change of

the point of interest (besides what comes out from the specific POV) there were no body and gaze movements outside the field of view where the characters are settled at the beginning.

- Lightning: the light can be an additional method to direct the user's gaze [4]. As mentioned above, to avoid as much as possible its influence, the light was used to mainly focus on the main action of each scene.
- Distance between the camera and the main action: it is important to empirically evaluate the depth of field (minimum and maximum distance to keep the focus) of the camera in order to have the action focused.

5 Realization

Chapter 5 is focused on the practical work that was done during this thesis. This work includes the production of two immersive movies, the selection of the metrics and the required system for the evaluation of these metrics.

5.1 Material production

For the implementation of this work, it was fundamental to have proper scripts in order to evaluate the narrative engagement and the differences between the two types of POV. As outcome of several brainstormings, discussions and reviews, two scripts have been realised and produced: *Persone che potresti conoscere* and *Oreste è ancora vivo*. Both scripts are in Italian.

The first script, *Persone che potresti conoscere*, represents a lunch between two brothers: Luca and Fabio. Scrolling on his phone, Luca found out on Facebook, in the section *People you may know,* the person who abused him. Shocked by this fact, Luca decides to confess everything to his brother. Since part of the narrative engagement metric relies on emotional engagement, this script was written with the intention of recreating in the viewer a high emotional burden. This script was inspired by a poem by Kevin Kantor, *People you may know* [26]. The two characters were played by two young actors who have already taken part in some professional productions. The location of the shooting was the living room of an apartment located in the small village of Cetara (Amalfi Coast, Province of Salerno, Italy).

For the 1-PP version several tests were made, taking in consideration previous work, in order to understand which was the best way to shoot it. In the end, it was decided to mount the camera on the head of one of the actors (who played Luca) with a GoPro head strap. It was stable enough (too much camera movements can create motion sickness), and permitted to partially see the body of the character that owns the POV. For the external observer version,

the camera was mounted on a tripod with an adapter for the GoPro. For this version, some tests were made to find the best distance between the camera and the actors in order to avoid the image distortion typical of the GoPro lenses when an object is too close to the camera (Figure 19).

Fig.19: Camera settings for the 1-PP version (left) and the external observer one (right). Photos taken during the shooting of Persone che potresti conoscere.

The various shots were then imported on a laptop and opened on GoPro Player. This software allows stitching images automatically, meaning that the two images taken from the two lenses are combined automatically. After this step, in order to preserve as much as possible the quality of the image, the shots were converted by GoPro Player using the CineForm codec. Subsequently, further editing was made using Adobe Premiere Pro 2020 to cut what was unneeded and add a soundtrack and ending titles. For the external observer version, a mask layer was added (created on Photoshop) in order to cover the tripod; for the 1-PP version, instead, the part where the GoPro support was visible was blurred (using the VR Blur effect). For both the versions, the audio was recorded with the six integrated microphones of the GoPro Max. The resolution of the camera was set at 5K which is its maximum value (5376×2688 pixels). The frame rate instead was set at 30 frames per second. The lighting of the scene was implemented by using only the sunlight coming from the window of the apartment.

The color grading was made through After Effect using the Effect Lumetri Color. The effect was not applied directly to the video images but to an adjustment layer, which is a special layer present in several tools of the Adobe Creative Suite that permits to keep separate the adjustments made to an image and the image itself. In this way, it is possible to have more flexibility and higher control on each editing without changing the pixels of an image. Besides some adjustments related to the tone of the image which included the intensity of the shadows, highlights, blacks and whites, it was added a preset look of the Lumetri Color called SL Blue Steel. This choice is merely artistic and was done in order to give to the image and higher feeling of melancholy. In Figure 20 it is possible to see the difference between two frames, before and after color grading.

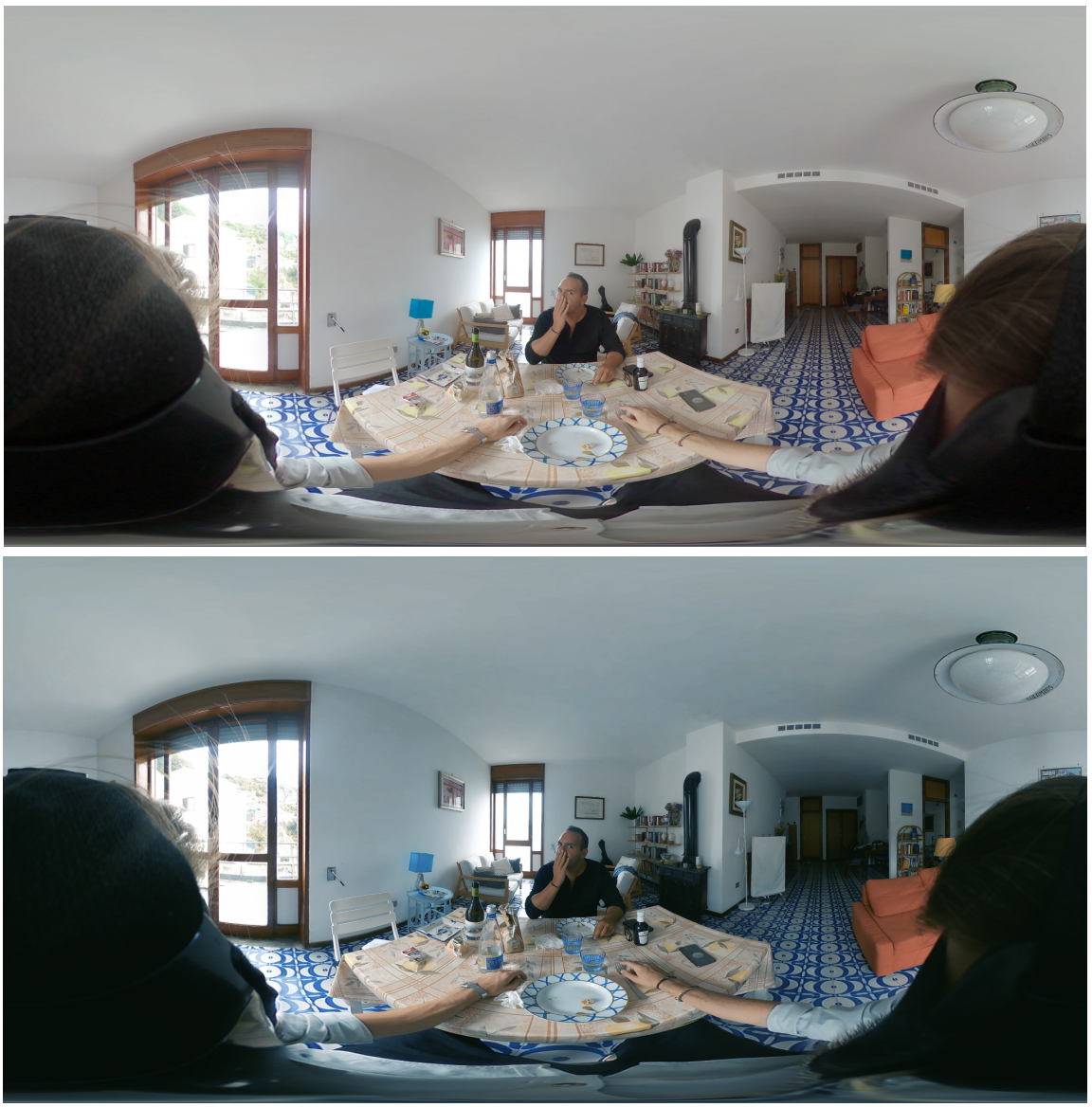


Fig. 20 : Two screenshots from the scene Persone che potresti conoscere. Before (top) and after (bottom) the process of color grading.

For the 1-PP version, Adobe After Effect 2021 was used also to stabilize the camera motion since the head of the actor made some small movements during the shooting. For the stabilization process, the VR Comp Editor was used.

Further material needed for the shooting were two images representing two Facebook's screens: one showing the section "People you may know" and the other the profile picture of Andrea Sala (the guy who violated Luca). All the names, including those that appear in these

images, are completely fictional and the profile pictures were generated by an online Face Generator (https://generated.photos/face-generator); this tool, thanks to Artificial Intelligence, creates unique faces of people that do not really exist. For final realization of the two images, Adobe Illustrator was used (Figure 21). In order to simplify the post-production and avoid any time consuming 360° tracking process, the two images were directly uploaded on the actor's smartphone (Luca) and then shooted as the character was actually using Facebook.

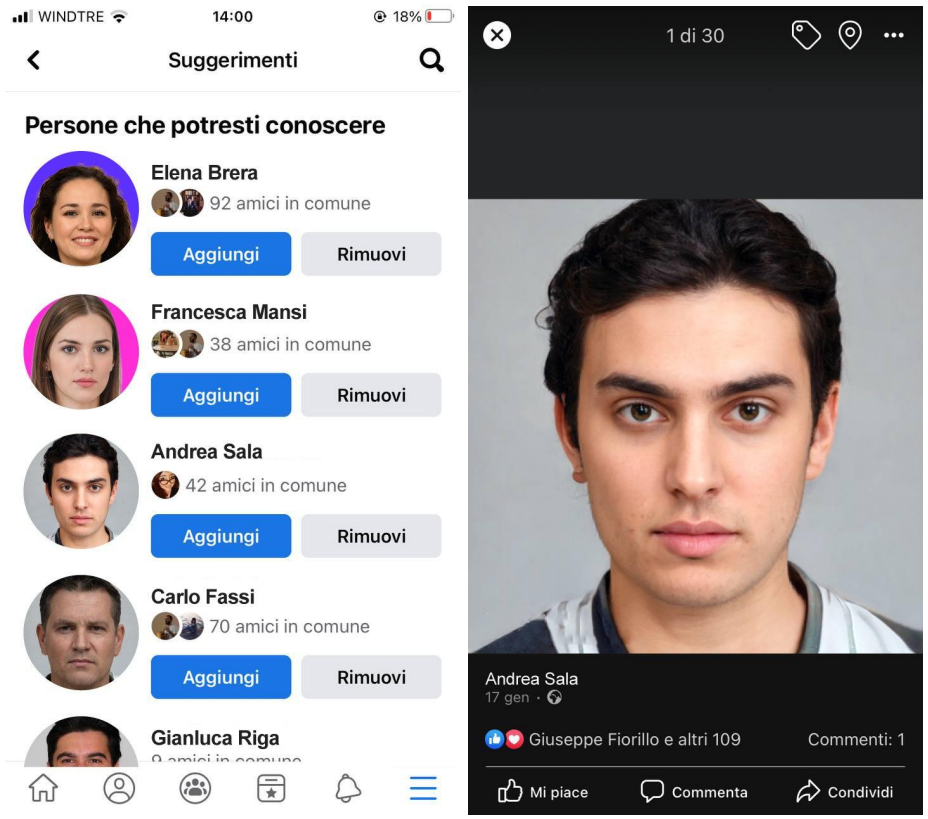


Fig.21: Facebook screens realised for the shooting of the video "Persone che potresti conoscere"

At the end of both versions it was added in Adobe Premiere Pro the title of the scene on a black background for five seconds. The videos were exported on Adobe Media Encoder using the coded HEVC (High Efficiency Video Codec, also known as h265), 25 fps and 5K resolution. Also the audio was exported with the same software as stereo audio (two channels, left and right) at 48 kHz. These export settings were used in order to keep both

good image quality and small size; in this way, it was possible to easily upload and watch the videos on different platforms such as YouTube and Unity.

The duration is 3 minutes and 33 seconds for the 1-PP version and 3 minutes and 52 seconds for the external observer one. The different duration is due to the fact that the actors were not able to perfectly replicate the same timing while shooting the two versions.

Once the two versions of this scene were realised, the outcome of a preliminary analysis was a concern related to the fact that both video were very engaging due to the great emotional impact; furthermore the two versions seemed to not have enough elements that could underline the differences between a 1-PP and a external observer POV. For such reasons, it was decided to add another scene that could better stress the NE. The second script, *Oreste è ancora vivo*, was designed trying to add elements that could better elicit the differences in NE between the 1-PP and the external observer versions (if any). The story, with a lower emotional impact, is about a dystopian future in which the government (called *La Rete,* i.e., The Network) constantly controls people by forcing them not to stay offline for more than 48 hours. Oreste, a mysterious man, is a wanted man since he has been offline for more than three days and his body was not found by the government's drones (which search for these subjects). The vicious head of the security Vadio Sersi begins to investigate and finds out that a young smartphone mender, Sena Diaz, might be involved. The scene shows the pressing interrogation made be Vadio to Sena. This script was inspired by two famous dystopian social science fiction novels: *Nineteen Eighty-Four* by George Orwell and *The Circle* by Dave Eggers.

The two main characters were played by young actors with theater and cinema experiences, whereas the third character (the security agent), who had no lines in the script, was played by a person without prior professional acting experience. The location for the shooting was the Visionary Lab at Politecnico di Torino.

The elements that differentiate the two versions of this scene were that in the 1-PP version the viewer could hear the thoughts of Sena, whereas in the external observer version only the dialogue between the two characters can be heard. Furthermore, the 1-PP viewers could have a detailed view of the objects that were shown to Sena by Vadio (a photo of Oreste, a message on a piece of paper ,the content of a box, and a video projected on Vadio's phone). An additional visual element was that in the 1-PP version, at the beginning of the scene, the view is covered by a black bag (which is on the head of Sena); only after the first line of Vadio the viewer can see the room; in the external observer version, the room is always visible (Figure 22).

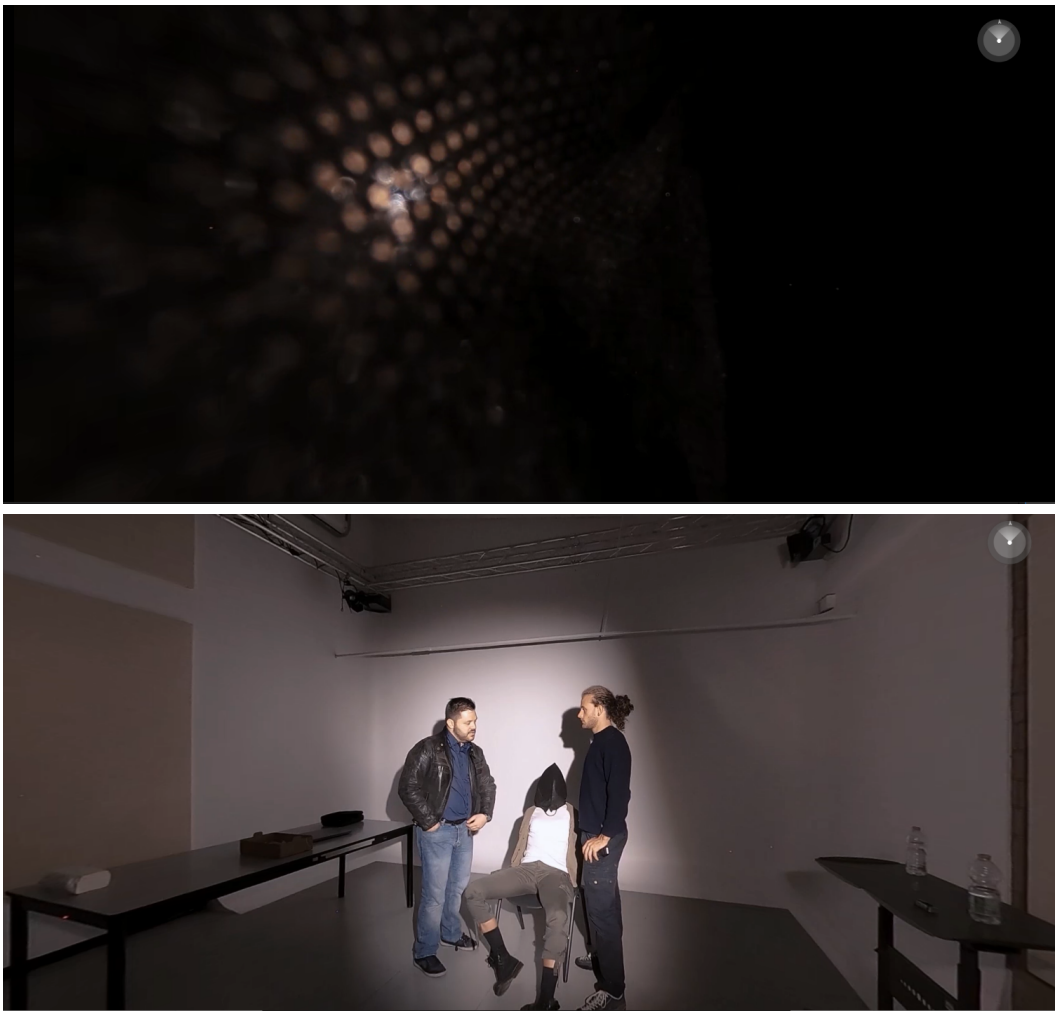


Fig. 22 : Two screenshots from the beginning of the scene Oreste è ancora vivo. 1-PP version (top) and external observer version (bottom)

The hypothesis was that these elements in the second scene could give a higher understanding of the characters (especially of Sena) and that the 1-PP, in particular, could give a higher emotional impact.

Before the shooting, a preliminary study was conducted on the second script in order to have a first feedback that may suggest its effectiveness in stressing the NE and indicate whether the two versions would actually translate into different experiences. Two versions of the second script were prepared aligned with the peculiar features of the 1-PP and external observer versions. The participants of the preliminary study were asked to read just one version of the script and answer to the narrative engagement form which was previously adapted for a reading experience. The questionnaire and the script were in Italian, and only native speakers participated in order to avoid any bias generated by linguistic issues. Overall, 21 subjects were involved (15 males and 6 females, 26.3 years old on average), 9 for the 1-PP version and 12 for the external observer one. Since the results were promising and in line with the above-mentioned hypothesis, it was decided to proceed with the shooting of the scene.

The camera settings for this scene were basically the same as the previous scene in terms of resolution and frame rate. For the 1-PP version, the camera was mounted on the head of the main actress (the one who was playing Sena), whereas for the external observer version the camera was settled on a tripod approximately 1,5m high (Figure 23).

Fig.23: Camera settings for the 1-PP version (left) and the external observer one (right). Photos taken during the shooting of Oreste è ancora vivo.

The lighting of this scene was made by using only the Led Fresnel spotlight which was pointing on the face of the actress. The intention was to create an atmosphere similar to the typical interrogation of crime movies. The audio dialogue of this scene was recorded using the microphones of the camera, whereas the thoughts of Sena were captured in a second moment with the ZOOM H5 recorder (Figure 24).

Fig.24: Photo taken during the recording of the audio of Oreste è ancora vivo.

The process of post-production mainly followed the same steps of the previous scene, and the same software was used (After Effects, Premiere Pro and Photoshop). For the audio recorded with the ZOOM H5, which included the thoughts of Sena, a reverb effect was added though

the Audacity software. Reverb an audio effect applied to a sound signal to simulate reverberation which is how long a sound persists after its production. Reverberation usually depends on how the reflection is caused by the space where the sound is produced. This effect is commonly used in cinema to simulate an audio that is reproduced in the head of a character such as his or her thoughts. Audacity permits to control different parameters such as the pre-delay, the room size and the percentage of reverb. After testing different presets, it was chosen the one that simulates audio recorded in a large room (Figure 25). This preset permitted to have good intelligibility of the audio and to diversify the thoughts of Sena from her dialogue lines.

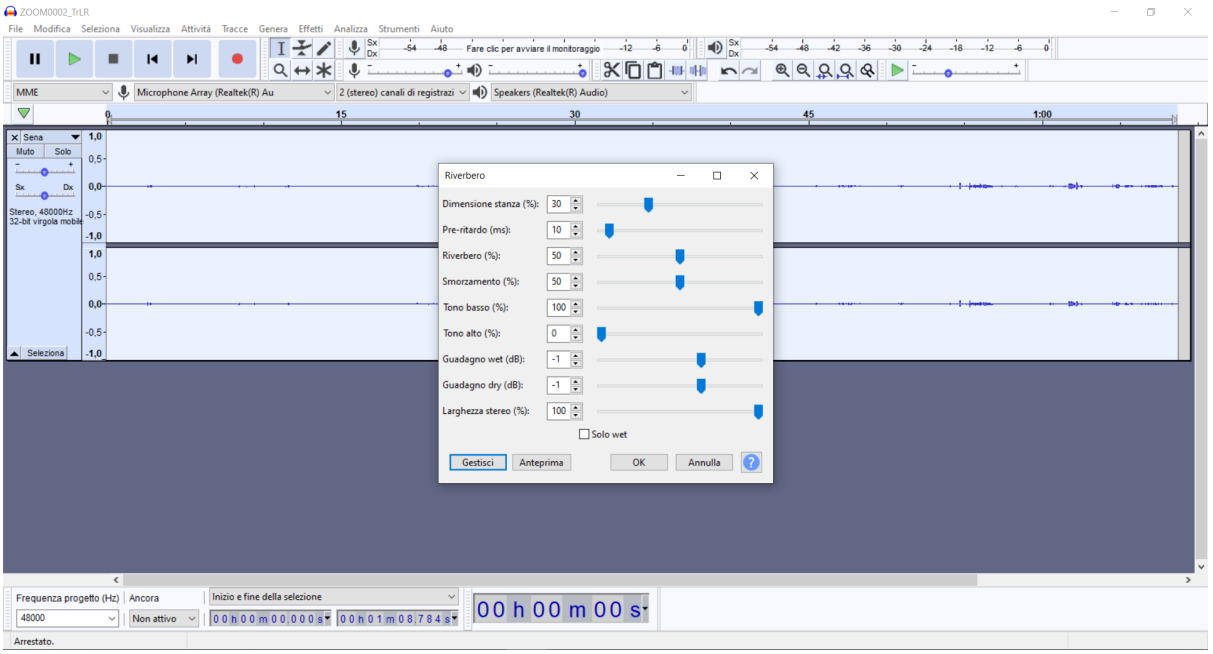


Fig.25: Screenshot from the Audacity software showing the reverb effect window and its parameters.

In this case, since there was less light compared with the other scene, the color grading was minimal. In After Effects, using an adjustment layer and the effect Lumetri Color, it was selected the preset CineSpace2383. Furthemore, in order to avoid orange halos on the side walls of the room, the white balance was set at the temperature of -20°. The combination of

these choices gave to the image a cold look and a high concentration of the light in the characters' position (Figure 26).

Fig. 26 : Two screenshots from the scene Oreste è ancora vivo. Before (top) and after (bottom) the process of color grading.

The end title of the scene was realised with Adobe Illustrator and added for five seconds at the end of both versions. Further graphic materials include: the photo of Oreste (generated by the online tool Face Generator) and a 30 seconds video that shows the time that Sena has been offline (Figure 27). This video was realised using Adobe Illustrator for the design and then animated in Adobe After Effects.

The exporting process used the same software (Media Encoder) and the same settings as for the first scene. The duration of the 1-PP version is 4 minutes and 37 seconds, whereas for the external observer version is 4 minutes and 31 seconds.

Fig.27: Two assets realised for the shooting of Oreste è ancora vivo. The photo of Oreste (left) and a screenshot of the video that shows the time that Sena has been offline.

5.2 Measuring narrative engagement

In order to measure the NE,the scale presented in [19] by Rick Busselle & Helena Bilandzic was used. The NE scale (NES) consists of a 12-items questionnaire which evaluates this element considering the four factors reported below.

- Narrative understanding (NU): the user's level of understanding in relation to the elements of a story (the plot, the characters, the events, etc.). When a user finds it hard

to understand a narrative content, this might cause confusion or even frustration which leads to lower levels of engagement.

- Attentional focus (AF): how difficult it is for the user to focus on what is happening in the story. A narrative content (such as a novel or a movie) is considered engaging also for its capacity to keep the concentration of the users on its main action. The reasons for distraction in a movie can be an uninteresting storyline (a bored person more easily starts to wonder about something else) or even an element of the film itself that is not perfectly consistent with the story (i.e., a background sound that does not match the environment where the action takes place).
- Narrative presence (NP): the feeling of being transported from the real world to the story world. An activity (such watching a movie) can be considered engaging when a person tends to lose track of the space and of the time.
- Emotional engagement (EE): the emotions that viewers have with respect to characters, either feeling the characters' emotions (empathy), or feeling for them (sympathy).

The outcome of each experiment was collected and later analyzed in order to evaluate if the two versions with an internal observer POV show higher significant results than the one with an external observer POV. The use of this questionnaire had two motivations:

- its effectiveness was largely validated in previous studies [19][11];
- it is easily adaptable for short stories [19].

Statistically, as said in Chapter 2, the studies in [18] did not show significant results regarding the effect of scene's transitions on the narrative engagement. The explanation provided by the authors was that that might have been caused by a low number of participants or by the fact that different transitions do not actually affect the narrative engagement in relation to the

above mentioned parameters. The aim of the study carried out in this thesis is to discover if the use of different POVs might lead to statistically significant differences, assuming that different POVs may have a stronger narrative impact than transitions. The hypothesis of this study is that the 1-PP version might give higher emotional engagement and narrative presence, since the user is settled at the center of the story as one of the protagonists. Furthermore, the scene *Oreste è ancora vivo* was specifically designed to additionally stress the factors of attentional focus and narrative understanding (as mentioned in Chapter 5). The hypothesis is that, for this scene, also the attentional focus and narrative understanding can be higher in the 1-PP version than in the external observer version; in fact, the users who watch this version can hear the thoughts of the protagonist (Sena) and they are able to watch further elements of the story that can support its comprehension. For example, when the head of the security Vadio shows to Sena the box with several evidences, the 1-PP users can hear *I have nothing left, besides this hope of revolution.* This Sena's thought does not add anything to the plot of the story but it might help the user to better understand the character of Sena and her commitment in a future revolution. Furthermore, another example is at the end of the scene when Vadio tells Sena that the drones are coming to take her. He explains that she had been offline for almost 48 hours (against the rules of their society),then he shows her a countdown on his phone. This dialogue part is the same for both the versions, but in the 1-PP version the users are able to watch the countdown which might give an additional cue towards narrative comprehension and emotional engagement.

For each item of the questionnaire the participants should express how much they agree using a Likert Scale from 1 (strongly disagree) to 7 (strongly agree). The 12 items of the questionnaire are related to the four factors mentioned above (three items per factor). All the items related to the attentional focus and narrative understanding have a negative impact on the overall score due to the fact that they are formulated with a negative connotation. An

example of statement related to the evaluation of the narrative understanding is: *At points, I had a hard time making sense of what was going on in the video.* For the attentional focus: *While the video was on I found myself thinking about other things*. The items related to the narrative presence and the emotional engagement, instead, have a positive impact on the overall score of the narrative engagement. For the narrative presence, an example is: *During the video, my body was in the room, but my mind was inside the world created by the story.* For the emotional engagement: *I felt sorry for some of the characters in the video.*

The items were randomly sorted in order to avoid possible biases. The questionnaire was translated to the Italian language, and for some of the items were added specific examples in order to make it easier for the users to understand the meaning of the statement. For instance, for the the item *My understanding of the characters is unclear* (narrative understanding), the relationship between the characters was included as an example.

5.3 Eye tracking system

In order to obtain objective data from each user and compare them with the attentional focus factor, raw gaze data were collected during the view of each video. These raw data included the head pose (where the camera is pointing), the combined gaze direction (an interpolation of the direction where each eye is pointing), and the timestamp related to the time instant at which these data were recorded.

Data were collected in order to compute several metrics that were selected for this study, which are described below.

- PercFixInside: the number of fixations within a specific region of interest (ROI) related to the total amount of fixations. A fixation represents a set of gaze points close in time and range, which occurs when the eyes are fixed at a particular object in a

stimulus. Typically, it has a minimum duration of 50 ms and an average duration of 200 ms [23]. The percFixInside is a metric introduced in [29] and used in several other studies related to VR such as [2]. This value represents an indicator of the viewer's interest in a specific ROI. In the scene *Oreste è ancora vivo*, the ROIs represented by the Vadio's phone, the box with the papers and the letter, contain more information and details in the 1-PP version. This aspect might be a reason for the 1-PP viewers to focus more on these ROIs than the external observer viewers; this fact, consequently, might be expressed by a higher value of the percFixInside for these ROI in the 1-PP viewers.

- nFix: the total number of fixations compared to the number of collected gaze points. This metric was used in [2] and [29], and represents how many saccades and fixations the viewers execute; a lower value indicates more saccades movements, which might represent a higher explorative behaviour. As in the external observer version the scene actions are more concentrated in a smaller portion of the image, it is expected to have a higher value for this metric so less saccade movements.
- Heatmap: a visual attention map that gives a representation of where the viewer's gaze has focused on. Usually, the red color represents a hot spot, an area with higher interest (more gaze points registered in that area). It is one of the most used metrics in research with eye tracking [23].
- Experiential fidelity with Intended POV and ROI: introduced in [6], this metric indicates the percentage of viewers who have watched a ROI considered as intended in a certain interval of time. For example, in *Oreste è ancora vivo*, while Vadio is showing Sena his phone, this is considered the intended ROI. As the 1-PP version gives additional details, the hypothesis is that further information stimulates more viewers to watch the intended ROI.

- Gaze path and head path: the distances traveled by the gaze and the head during the view of a video. A greater value of the head path means that the user has changed more times his or her field of view (the visible portion of the 360° video); this aspect might indicate a more explorative behaviour that is expected to occur during the view of the 1-PP version. A greater value of the gaze path instead, could indicate that the user has been moving his or her eyes (pupils) more frequently even without changing field of view.

As mentioned in Chapter 3 (Paragraph 3.3), the eye tracking system was developed in Unity using C# for the scripts. The system saves head poses and gaze points with a sampling rate of 50 Hz; this value permits to obtain two samples per frame, since the produced videos for this work are at 25 fps. Furthermore, this sampling rate is inline with other studies related to VR which included the collection of eye tracking data such as [29] and [24]. The data represent the head and gaze rotations along the longitudinal and latitudinal axes considering the head of the viewer at the origin of the coordinate system. The gaze data is registered through the API TobiiXR.GetEyeTrackingData(TobiiXR_TrackingSpace) available with the Tobii XR SDK. The TobiiXR_TrackingSpace represent the coordinate system used to track the gaze data and it has two possible values:

- World: the data are registered considering, as tracking space, the world space of the Unity scene; this is the tracking space used in this study since it was recommended by the Tobii documentation as the best solution to get what the user is looking at in the scene;
- Local: the tracking space has the same origin of the XR camera; data collected in this space is unaffected by head movements.

TobiiXR.GetEyeTrackingData(TobiiXR_TrackingSpace) returns a data set called TobiiXR_EyeTrackingData which contains the timestamp (when the data was received, measured in seconds) and an element called TobiiXR_GazeRay. The latter includes origin and direction of the gaze ray and a boolean value (IsValid) that indicates if the gaze data is valid or not; whenever this value is false (i.e., while the viewer is blinking) the data is not collected. For the head movements it is simply registered the rotation of the Unity camera which follows the head orientation. Finally, the system records latitude and longitude for the gaze and the head, so four angles in total. The latitude represents the angle along the x-axis which is the direction parallel to the pavement and it has a range between -90° and 90°. The longitude instead represents the angle along the y-axis which is the direction perpendicular to the pavement and it has a range between -180° and 180°. These angles are saved firstly in quaternions (which is the measurement unit for angles in Unity) and then converted as Euler angles. Once the angles from head and gaze are combined a function converts the latitude and longitude into values of the image coordinate system. The image coordinate system consists in a 2D system where the x-axis corresponds to the pixels along the width of the image and the y-axis to the pixels along the height of the image (Figure 29). The origin is settled at the left-bottom of the image and the values in pixels are obtained using the following formula:

$$x = (5376 \div 360) \times (longitude \times 180)$$

$$y = (2688 \div 180) \times (-latitude \times 90)$$

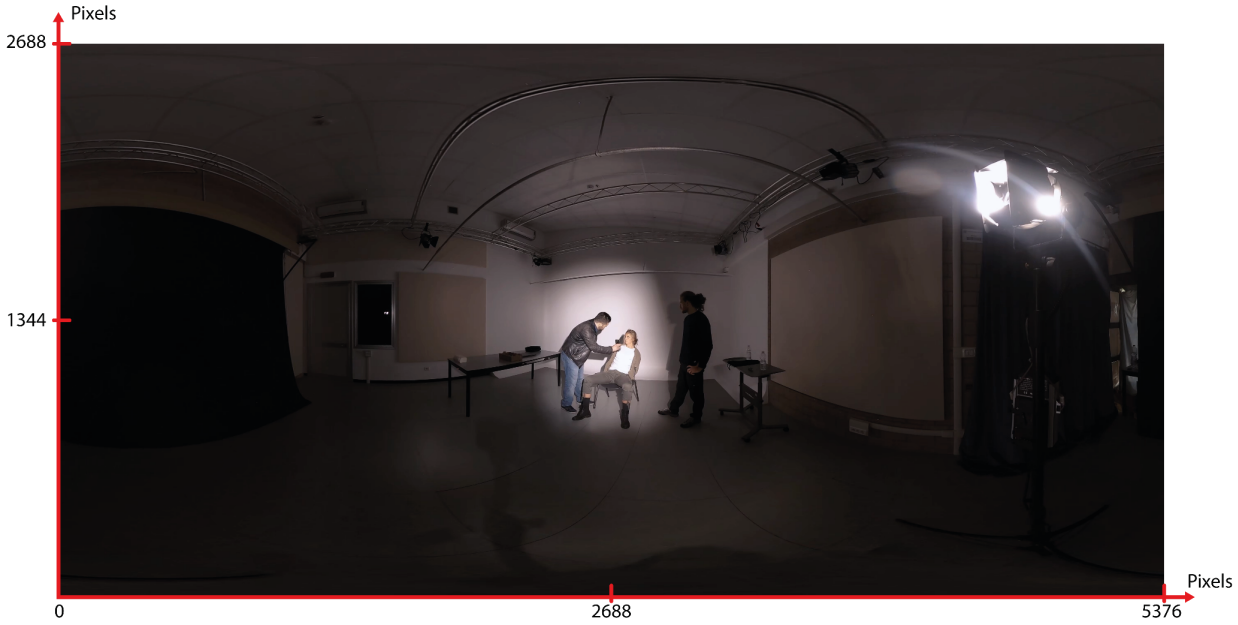


Fig.29: Image coordinate system.

The value 5376 represents the width in pixels of each frame in the video and 2688 its height. The initial orientation of the camera is set at (0°,0°), this value corresponds to the center of the image at (2688; 1344) in the image coordinate system.

Once the data are collected for the full length of each video, they are stored in a CSV file. In order to calculate the fixation-metrics (PercFixInside and nFix), it has been created a script that reads each gaze data from the file and evaluates whether it can be considered a fixation or not. For the fixation detection, the script implements the identification by dispersion-threshold (I-DT) algorithm introduced by Salvucci and Goldberg in [30]. This algorithm basically considers fixation a set of gaze data whose distance is below a certain dispersion threshold with a duration higher than a minimum duration threshold. The dispersion threshold was settled at 1° and the minimum duration at 100 ms following the recommendations in [30] and [31]. If the maximum angular distance between a set of gaze points was below 1° and the difference between the timestamp of the last gaze point and the first gaze point of this set was greater than 100 ms, a fixation was detected. The script used

for the fixation detection is part of the open source tool PyGazeAnalyser [32]. This tool, available on Github (https://bit.ly/pygazeanalyser), is originally written in Python. For further adaptation to the needs and specifications of this work the script was converted in C#.

Each scene has several ROIs that represent a peculiar element of the story. For the scene *Persone che potresti conoscere* the ROIs are:

- Luca (character), only in the external observer version is considered as ROI; in the 1-PP version is the POV;
- Fabio;
- Luca's phone.

In the scene *Oreste è ancora vivo*, the ROIs are:

- Sena, only in the external observer version;
- Vadio;
- the agent;
- Vadio's phone;
- the letter (a paper where is written that Oreste is still alive);
- the box (with several letters found at Sena's home).

For the *Experiential fidelity with Intended POV and ROI* metric, it was established in which time window a ROI was valuable in relation to the story. For the scene *Oreste è ancora vivo* it was established that the Vadio's phone, the letter and the box full of papers were valuable elements of the story. An element is valuable when it can be considered important to recognize the tread of the story. Specifically, for the phone, two intervals of time were considered: when at the beginning Vadio shows the picture of Oreste and when, at the end, he shows the amount of time that Sena has left. For the letter, it was selected the time window when Vadio exposes the letter to Sena; for the box instead, the time when Vadio leaves it on

Sena's legs. Considering the scene *Persone che potresti conoscere,* the only ROI that was considered was the phone of Luca, in the time window when he checks his Facebook feed.

To estimate the experiential fidelity, instead of the number of users, it was calculated the percentage of time that each user watched the above mentioned ROI related to the time window this element is considered valuable.

The script related to the fixation detection, besides the starting time, ending time, position and duration of each fixation, evaluates which ROI was the target of each fixation. To simplify the position tracking process, for each ROI it was defined a dynamic bounding box. Each bounding box was represented by a cube (Unity GameObject) and the properties of position, rotation and scale were animated in Unity so as to keep, for every frame, the ROI within the surface of its bounding box. The script basically compares the position of each fixation with the ones of the bounding box of each ROI. In Figure 29 and Figure 30 it is possible to see how the bounding boxes were settled, respectively for *Persone che potresti conoscere* and *Oreste è ancora vivo*.

Fig.29: Examples of the bounding box for each ROI in the scene Persone che potresti conoscere. On top the 1-PP version and at the bottom the external observer one.

Fig.30: Examples of the bounding box for each ROI in the scene Oreste è ancora vivo. On top the 1-PP version and at the bottom the external observer one.

For the realisation of the heatmaps it has been developed on Unity a dynamic heatmap generator starting from a template available on Github called UnityHeatmapShader (https://github.com/ericalbers/UnityHeatmapShader). This system includes a display and a section of input commands (Figure 30). The display is made with a Quad (Unity game object) that has been scaled in a way that its width is twice its height (like an equi-rectangular image). The display has attached a video player that projects on a render texture the scene. The texture (sized 5376×2688 pixels) is attached to a Unity material which includes a shader.

The latter represents a small script that contains the mathematical calculations and algorithms for elaborating the colour of each pixel rendered. The shader is written in a language called ShadersLab.

Considering a time instant of the scene and a temporal window, a C# script reads the gaze data from the CSV file and filters them considering only the gaze points with a timestamp greater than the difference between the intended time instant and the temporal window and less or equal to the time instant itself. In case the difference between the time instant and the temporal window is lower than zero, the gaze points are selected starting from the beginning of the video. Once the gaze points are filtered, it is calculated the number of times that each gaze point has been recorded in the considered time window. The resulting values are divided by the number of different gaze points recorded in the given time window and considered as intensity values for each related gaze point. Once obtained the coordinates of each gaze point and their intensity values, the shader script is able to calculate the color of each pixel and generate the heatmap. Taken a single pixel it is calculated the contribution (weight) that every gaze point has on it, considering the formula:

$$weight = intensity(GP(x,y)) * distsq(P(u,v),\ GP(x,y)) \div NumbeOfUsers$$

In the formula, *GP(x,y)* is the gaze point at the (*x*,*y*) coordinates and *intensity(GP(x,y))* represents its intensity. *P(u,v)* is the pixel at the (*u*,*v*) coordinates. All the coordinates are considered in pixels. *NumberOfUsers* represents the number of users that took part in the data collection. The factor $distsq(P(u,v),\ GP(x,y))$ depends on the distance between *P(u,v)* and *GP(x,y)*. Basically, a higher distance implies a lower contribution on the total weight. Its formula is:

$$distsq(P(u,v),\ GP(x,y)) = max\{0\ ,\ 1 - [distance((u,v),\ (x,y)) \div A]\}^2$$

The A value is an additional input parameter that can be considered as the area of interest of a gaze point. A greater value of this area means that each gaze point has a larger influence on the pixels around its position (Figure 31). This parameter is controlled through a slider and its value can be between 0,01 and 1.

The contributions of each gaze point are summarized in order to calculate the total weight of a pixel. According to the total weight, an algorithm establishes the color of the pixel. The red color corresponds to a total weight greater than or equal to 1. No color is assigned if the total weight is lower or equal to 0. For intermediate weight values, the color takes a value between a gradient:

- transparent-green gradient if the total weight is between 0 and 0.25;
- green-yellow gradient if the total weight is between 0.25 and 0.5;
- yellow-orange gradient if the total weight is between 0.5 and 0.75;
- orange-red gradient if the total weight is between 0.75 and 1.

The main input parameters, the time instant and the window length, are controllable through the input section. With a slider it is possible to change the time instant and with an input text field the window length. The heatmap is generated immediately whenever one of the input values changes. Furthermore, by clicking “S” on the keyboard it is possible to save a screenshot of the heatmap as a PNG file.

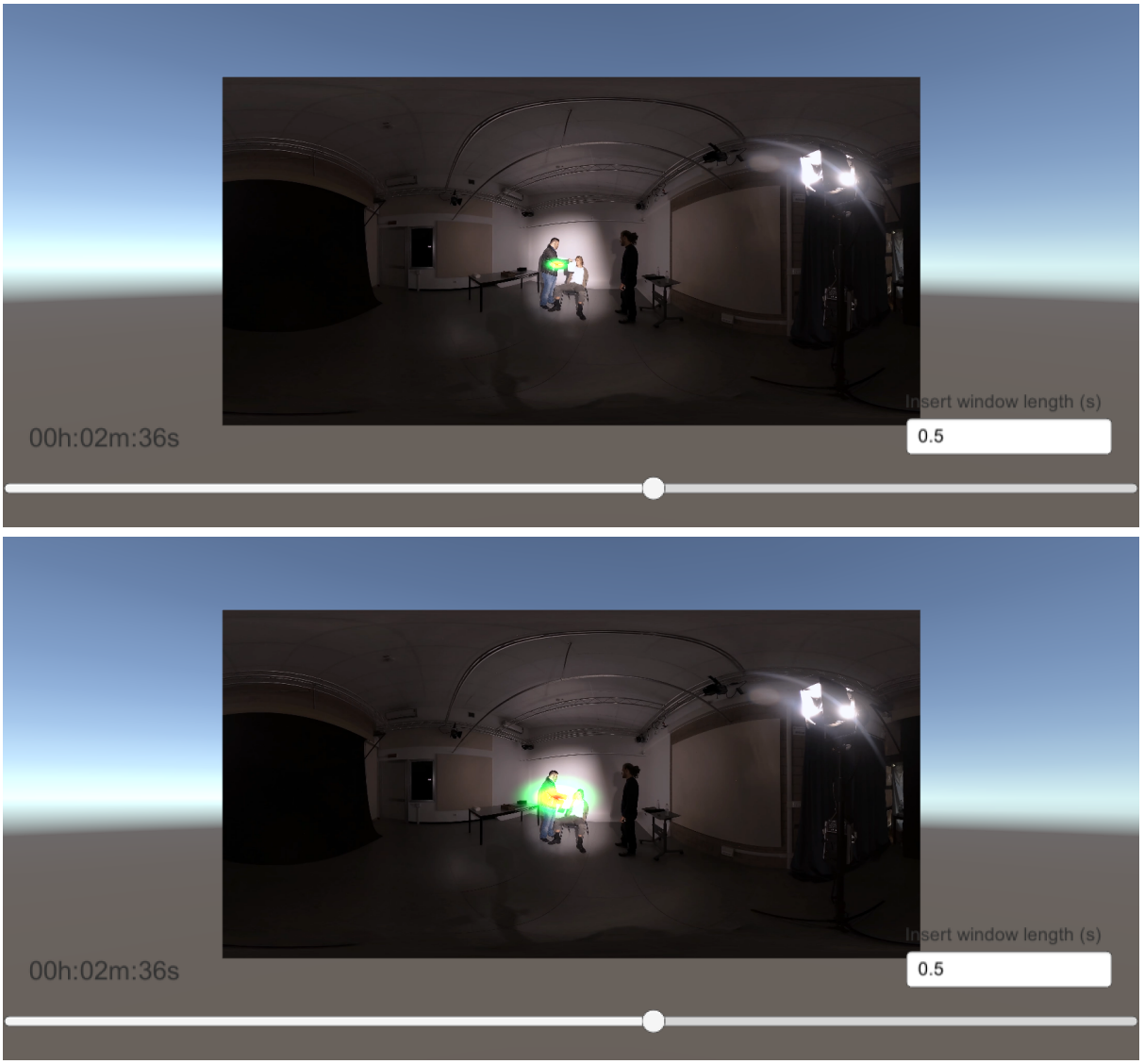


Fig.31: Two examples of heat maps generated using the data collected during the experiments. In the top image, the parameter A is set at the value 0,1; in the bottom image, A is equal to 0,25.

For the gaze path, once all the points relative to the gaze position were stored, it was calculated the total traveled distance. The same was done for the head path, but considering the points relative to the head position. Both distances have been calculated in pixels and divided by the video length (seconds) since the two versions of each scene were not perfectly matched in terms of duration.

5.4 Experiments

A group of 32 participants was recruited to participate in a user study. The users were mainly students and they were all volunteering for the study. The recruiting process was made through personal and academic connections.

Each experiment followed the protocol reported below.

- Welcoming and task explanations. In this first phase, it was told the participants that they had to see two different scenes. It was specified that each scene was shooted using two different POVs: 1-PP and external observer. No other specific task was asked besides not to move too far from the SteamVR base station (in order not to lose the user's location in the room). In case the user was not familiar with the VR headset, it was shown how to wear it and how to regulate its settings;
- Calibration of the headset. Once the user wore the headset, he or she was asked to perform the eye calibration provided by the SRanipal SDK. It consists in a three step calibration where the user has to adjust the interpupillary distance (IPD, distance between the center of the eyes) and the distance between the lenses and the eyes. Furthermore there is a short test where the user is asked to follow five dots with the gaze.
- Audio and video test. In order to avoid possible biases caused by vision problems, it was asked to the users if they could easily read what was written on the SteamVR dashboard (Figure 32). For an audio check, it was tested if the users could hear properly a test audio played on YouTube. In case there was any problem, further controls and adjustments to the headset were made.

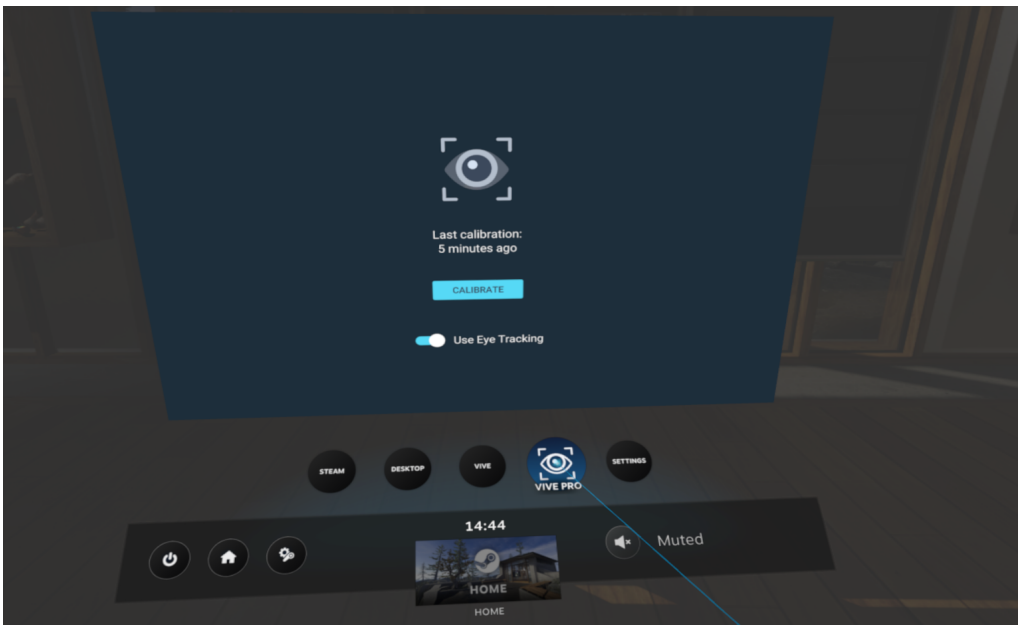


Fig.32: A screenshot of the StreamVR dashboard.

- View one version of the scene *Persone che potresti conoscere.*
- Answer to the NES questionnaire related to the scene *Persone che potresti conoscere* on a laptop.
- View one version of the scene *Oreste è ancora vivo.*
- Answer to the NES questionnaire related to the scene *Oreste è ancora vivo* on a laptop.

All the scenes were played on Unity and showed through the integrated display of the HTC Vive Pro Eye. Thanks to the messages on the Unity's console, it was possible to check if the eye tracking data were correctly stored during each session. The audio was played directly through the headphones of the HTC Vive Pro Eye. Regarding which posture to have during the vision, it was up to the user to decide for each video between sitting on a swivel chair or simply standing. The sitting posture was the one suggested for the 1-PP versions and the standing one for the external observer videos. These suggestions came from some first users that took part in the test sessions; they noticed that matching the position in height between the virtual camera and the user helped to better immerse in the story. However, since not all

the first users expressed this issue related to the height of the camera, it was said to the users to use the posture more comfortable for them during each video even though it was different from what suggested.

At the beginning of each scene, the orientation was set in a way that the user's gaze was pointing at the center of the image, i.e., to coordinates (2688,1344).

The users were equally divided in two groups, so that a group watched first the 1-PP version of *Persone che potresti conoscere* and then the external observer version of *Oreste è ancora vivo*. The other group instead watched first the external observer version of *Persone che potresti conoscere* and later the 1-PP version of *Oreste è ancora vivo*. The first scene was mainly used as a training session in order to prepare the user for the view of the scene *Oreste è ancora vivo*. Furthermore the first scene was useful for the users to familiarise with the VR cinema and the use of a headset. *Oreste è ancora vivo* was selected as the main scene for the experiment since, as said, it contains more elements that might stress the factors of narrative engagements and better underlines the differences between 1-PP and external observer POVs. Notwithstanding, objective and subjective data were stored for both the scenes so that it would be later possible to make an analysis between their respective results and see how much the kind of scene influences the various metrics.

Besides the 12-items questionnaire related to the evaluation of the narrative engagement, additional questions were asked to the participants. Three items were related to how often they were used to watch VR and traditional movies. Specifically:

- How often do you use devices related to VR (cardboards, headsets, etc.)?
- How often do you watch immersive movies?
- How often do you watch traditional audio-visual contents (movies, series, cartoons)?

For these items, the users had to answer by selecting a value between 1 and 5; 1 means never, 2 sometimes, 3 once per month, 4 once per week and 5 every day.

Finally, at the end of both visions and both questionnaires, the users were asked which POV they generally appreciated more (1-PP or external observer), their reasons, and if they had additional comments related to the experience. All the questions and their answers were written and stored on an Excel file, using a laptop.

During all the sessions, eye tracking data were collected for the whole duration of each video, except for the end titles where no data was collected. The eye tracking system automatically stored the gaze data on the last frame before the beginning of the titles. In Figure 33 it is possible to see an example of how the test environment was configured.

Fig.33: Photos taken from two of the experiment sessions.

6 Results and discussion

The group of 32 participants included 25 males and 7 females. The mean (M) age of the participants was 25,06 with a standard deviation (SD) equal to 3,90. On the one hand, most of the users had little experience with devices (M=1,94 and SD=1,47) and immersive cinema (M=1,25 and SD=0,43). On the other hand, the majority of the participants watched traditional cinema content every day (M=4,81 and SD=0,46).

6.1 Oreste è ancora vivo

The first results presented in this chapter are related to the scene *Oreste è ancora vivo*; as said, this scene was the one specifically designed to stress the differences between the two POVs.

6.1.1 Subjective results

The data collected for each metric (NES and eye-tracking) were analysed through a Shapiro-Wilk test in order to evaluate if distributions were normal. Since none of the distributions was normal and the data collected for the two POVs were independent (due to the fact that they were collected from two different groups of users), a Mann-Whitney test (two tails) was implemented to check if there was any statistically significant difference between the 1-PP and the external observer data. Two distributions were considered different when the *p*-value (abbreviated *p*) was lower than 0,05. Furthermore, in order to evaluate how strong the effect of the POV was, the effect size *d* was calculated (using Cohen's formula) comparing the two distributions (1-PP and external observer) for each metric. As suggested in [42], the effect of the POV was considered:

- Small when $|d| \geq 0,2$;

- Medium when $|d| \geq 0,5$;
- Large when $|d| \geq 0,8$;
- Very large when $|d| \geq 1$.

Table 1 presents the numerical results for the NES, considering for the two POVs the mean value (M), the standard deviation (SD), the p-value (p) and the effect size (d). Wherever $p < 005$, the p field is filled in green.

Oreste è ancora vivo						
Metric	M		SD		*p*	*d*
	1-PP	Ex	1-PP	Ex		
NU(-)	**4,4667**	7,5333	1,4996	3,8274	0,0218	1,055
AF(-)	**4,8667**	7,5333	2,0613	3,2836	0,0109	0,9727
NP	**15,6**	10,933	4,5869	3,3757	0,0035	-1,1588
EE	**14,8**	8,7333	4,847	3,316	0,0017	-1,4609
NE	**21,067**	4,6	8,7519	10,8	0,0004	-1,6752

Tab.1: Results of the NES for the 1-PP and external observer (abbreviated Ex) versions of Oreste è ancora vivo. The presence of (-) indicates reverse coded aspects.

As highlighted in Table 1, all the results given by the NES showed significant differences ($p < 005$). Generally the users who watched the 1-PP version experienced a greater level of narrative engagement (1-PP: $M = 20.87, SD = 8,81$; External observer: $M = 5,20, SD = 10,32; p = 0,0005; d = -1,6327$). Furthermore, the effect size between the two POVs in relation to the NE showed a really strong effect ($|d| \geq 1$) on narrative engagement. This result seems to confirm the main hypothesis of this work, which considered the 1-PP a better technique for creating interest and involvement in 360° scenes. The users who experienced the external observer version showed major difficulties to

comprehend the plot and the characters of the story (1-PP: $M = 4,4667$, $SD = 7,5333$; External observer: $M = 1,4996$, $SD = 3,8724$; $p = 0,0218$; $d = 1,055$). This result was predicted by the hypothesis that giving additional audio and visual elements enables an easier understanding of a movie. Also for the NU, it was found a very large difference between the two POVs ($|d| \geq 1$). Similar outcomes were observed for the AF factor; in fact, the external observer subjects showed higher difficulties in terms of concentration on the main action of the scene (1-PP: $M = 4,8667$, $SD = 7,5333$; External observer: $M = 2,0613$, $SD = 3,2836$; $p = 0,0109$; $d = 0,9727$). As hypothesized, this result might be due to the fact that virtual interactions (i.e. the characters directly speaking to the camera) and the additional details (i.e. the phone's display) in the 1-PP version represent an additional stimuli to the users' attention. However, the effect of the POV on this factor appeared to be slightly lower than on the other factors ($|d| \geq 0,8$).

As hypothesized in Chapter 5, among the four factors of the NE, the effect of POV was more evident in the EE and NP ones. The users who watched the scene through the eyes of the protagonist showed greater empathy with the character and greater emotional involvement with the story (1-PP: $M = 14,8$ $SD = 8,7333$; External observer: $M = 4,847$ $SD = 3,316$; $p = 0,0017$; $d = -1,4609$). Regarding the immersivity and the sense of transportation (NP), also in this case the 1-PP version showed substantially better results (1-PP: $M = 15,6$ $SD = 10,933$; External observer: $M = 4,5869$ $SD = 3,3757$; $p = 0,0017$; $d = -1,6752$).

6.1.2 Objective results

In Table 2 it is possible to see the results related to the eye tracking data for the scene *Oreste è ancora vivo*. The significant differences are highlighted in the *p* column with a green color.

Oreste è ancora vivo						
Metric	M		SD		*p*	*d*
	1-PP	Ex	1-PP	Ex		
nFix	0,8602	**0,9937**	0,1711	0,289	0,1985	0,5618
PercFixInside Phone	**13,293**	1,6653	2,3595	1,0184	0,0001	-6,3989
PercFixInside Agent	**15,213**	7,8947	5,3783	4,5195	0,0009	-1,4732
PercFixInside Vadio	**55,014**	53,567	8,7183	9,7723	0,74	-0,1563
PercFixInside Letter	**2,3887**	1,5547	1,1765	0,5369	0,0344	-0,9121
PercFixInside Box	**3,5747**	2,1107	1,6307	1,0614	0,0089	-1,0641
Head path	**216,11**	138,84	86,188	56,462	0,031	-1,0606
Gaze path	3272,6	**4974,8**	1029	1844,4	0,0279	1,1399
Experiential Fidelity Phone (time 1)	**49,953**	22,532	16,55	27,265	0,0251	-1,2159
Experiential Fidelity Phone (time 2)	**38,658**	8,2313	11,928	12,404	0,0001	-2,5005
Experiential Fidelity Letter	**44,83**	38,945	9,388	14,9	0,1985	-0,4726
Experiential Fidelity Box	42,482	34,657	15,894	23,473	0,229	-0,3904

Tab.2: Results of the eye tracking data collected for the scene Oreste è ancora vivo.

On average, the percentage of fixations (*nFix*) was slightly higher for the users who watched the external observer versions (1-PP: $M = 0,8602\%\ SD = 0,1711$; External observer: $M = 0,9937\% SD = 0,289$; $p = 0,1985$; $d = 0,5618$); however no significant difference was found for this metric ($p > 0,05$); hence, based on this information it cannot be assured that the 1-PP supports a more explorative behaviour (more saccades movements).

Regarding the percentage of fixations inside a specific ROI (*PercFixInside*), the results presented substantial differences for the ROI labeled as Phone, Letter, Box and Agent. All of them presented significantly higher values of *PercFixInside* in the 1-PP version. These results were expected for the phone, the letter and the box; the additional information they presented into the 1-PP version could have made them more attractive and interesting for the users. The result for the *ROI Agent* could be related to the fact that in the external observer version, the user is more focused on the characters that are speaking (Sena and Vadio). Always in relation to the *PercFixInside*, a very large difference was observed for the phone (1-PP: $M = 13,293\ SD = 2,3595$; External observer: $M = 1,6653\, SD = 1,0184$; $p = 0,0001$; $d = -6,3989$); besides the amount of information, in this case the POV influenced particularly also the scale of this object as presented to the users. In the 1-PP version, the phone covers a larger portion of the field of view. The scale difference between the phone in the two versions was probably an additional variable that affected this result. This assumption is supported by the results of the *Experiential fidelity with Intended POV and ROI*. In fact, in this case the only ROI that showed significant differences is the phone (for both time intervals where it was considered as intended); for the box and the letter, the differences were not statistically significant considering the *Experiential fidelity*.

Interesting results were found for the metrics related to the *head path* and *gaze path*. For both the metrics the data showed statistically significant differences; however, the *head path* indicated higher values for the 1-PP version whereas the *gaze path* showed the opposite

trend. These results suggest that the 1-PP (as expected) stimulated the users to change their field of view more than the external observer POV. On the other hand, the same users covered a smaller distance with their gaze (*gaze path*). This result might be due to the fact that in the 1-PP version, the users had to redirect their head in order to make certain elements of the story visible (e.g., the users had to keep their head down to watch the content of the box). The 1-PP users were more prone to change their field of view (redirecting their head orientation) and then keep their gazes focused on the element considered as ROI; this assumption seems to be coherent with the result of the *PercFixInside* metric. On the other hand, the external observer POV allowed the users to see a wider portion of the 360° environment, without any change of the field of view; therefore, the users that watched the external observer version were more prone to keep stable their field of view and focus on the various elements of the story by redirecting their gaze.

In order to evaluate how the heatmap evolved for each version of the scene *Oreste è ancora vivo*, the gaze points collected from each user were gathered into two CSV files: *gaze_points_external_observer*, for the ones who watched the external observer version and *gaze_points_1pp*, for the users who watched the 1-PP version. Both files were imported in the Unity project dedicated to the creation of heatmaps (Paragraph 5.3), using a time window equal to the duration of one frame ($t = 0,04\ s$). For every frame of each version the relative heatmap was created and stored. Once the two sequences of heatmaps were obtained, they were edited together in Adobe Premiere Pro and exported as a video at the same frame rate of the original scenes (25 fps). The results are two videos where it is possible to understand, for each instant, where the users were mainly watching. Both the videos are available at this link: https://bit.ly/heatmaps-oreste.

Figure 34 represents three different frames extracted from the heatmap sequences of both the versions. The frames labeled as *Time 0*, *Time 1* and *Time 2* are all relative to the moment

when Vadio shows the photo of Oreste on his phone's display to Sena. The time interval between each frame is equal to 1 second. Consistently with the result of the *External Fidelity* metric, it can be seen how most of the users who watched the 1-PP versions kept their visual attention on the phone. In fact, for all the three frames of the 1-PP version, a concentration of gaze points (red area) on the phone's display can be seen; for the external observer version, it is visible a larger dispersion of gaze points around the image, especially on frames *Time 1* and *Time 2* (more green areas). This result indicates that most of the users who watched the external observer version, at the moment Vadio presented the photo to Sena (*Time 0*), were following with their gaze the phone's position but then their visual attention went soon to other parts of the image. At *Time 1* and *Time 2*, for the external observer version, basically there is no "hot point" (red area), which means that the users were watching all in different directions.

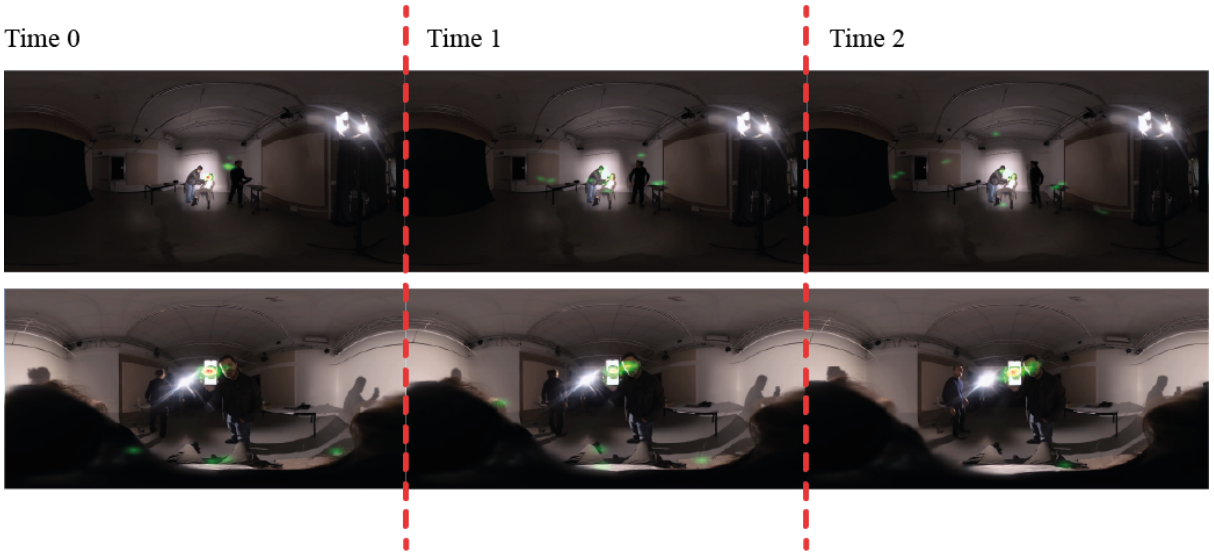


Fig.34: Three frames from the scene Oreste è ancora vivo, representing the moment when Vadio shows for the first time the photo of Oreste to Sena. The heatmaps show the concentration of gaze points recorded while watching the external observer version (up) and the 1-PP version (bottom). A red area indicates that the majority of the users were looking in the same direction at the same time.

Similar results are shown in Figure 35, where the frames are related to the moment when Vadio reveals, at the end of the scene, how much time Sena has left. Also in this case the

users who watched the external observer version were less prone to focus on the phone as seen in the results for the *Experiential Fidelity*.

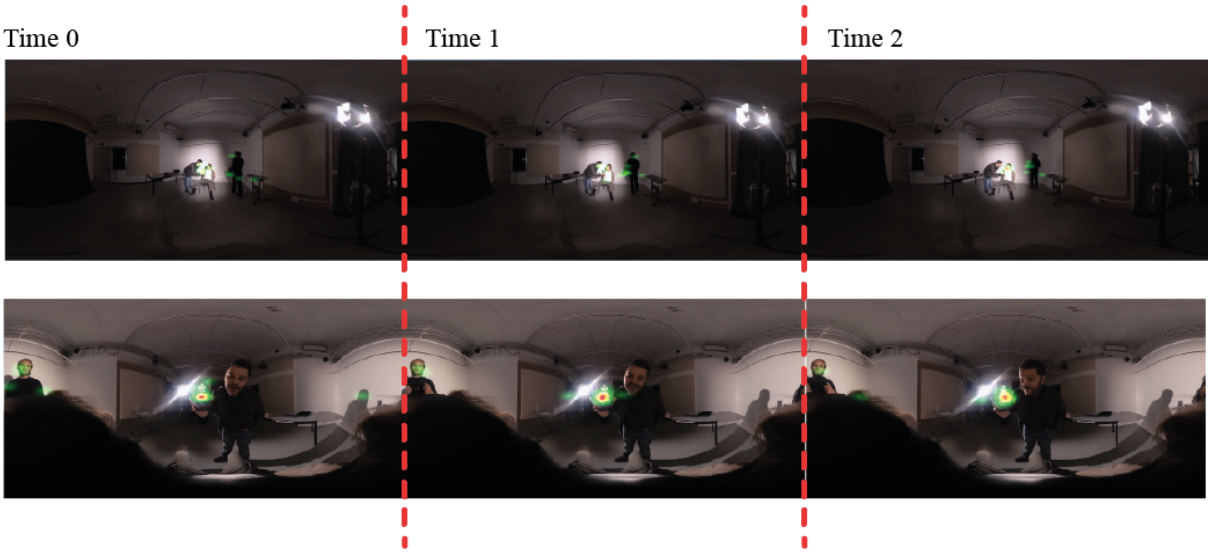


Fig.35: Three frames from the scene Oreste è ancora vivo, representing the moment when Vadio shows the time that Sena has been offline. The heatmaps show the concentration of gaze points recorded while watching the external observer version (up) and the 1-PP version (bottom). A red area indicates that the majority of the users were looking in the same direction at the same time.

In Figure 36, where the frames show the moment when Vadio presents the letter to Sena, in both the versions the gaze points are concentrated next to the letter. For the external observer version, however, the red area is closer to Sena's eyes while she is reading. The heatmaps indicate that the visual attention of the 1-PP users focused more specifically on the letter (in accordance to the results concerning *PercFixInside*; however the *Experiential Fidelity* did not show statistically significant differences since both the groups of users where mainly watching in that direction while the letter was showed.

In Figure 37, the frames represent the moments when Vadio reveals the box to Sena and leaves it on her legs. At the same time, the agent comes closer with his gun pointing to Sena's head. This element partially reduced the interest of the users into the box, especially for those who watched the external observer version.

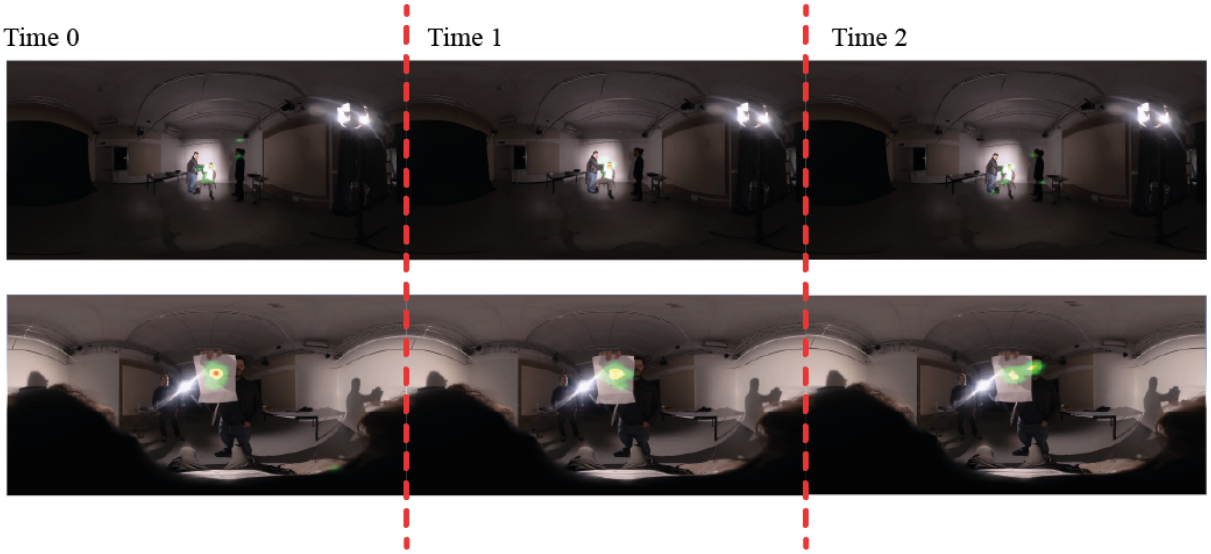


Fig.36: Three frames from the scene Oreste è ancora vivo, representing the moment when Vadio shows a letter where it is stated that Oreste is still alive. The heatmaps show the concentration of gaze points recorded while watching the external observer version (up) and the 1-PP version (bottom). A red area indicates that the majority of the users were looking in the same direction at the same time.

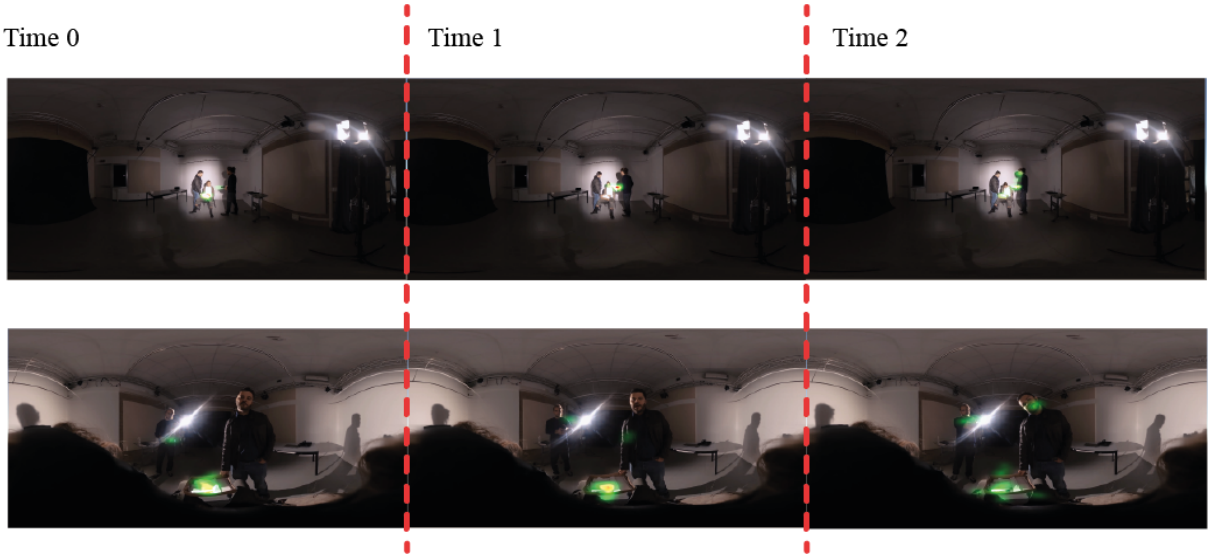


Fig.37: Three frames from the scene Oreste è ancora vivo, representing the moment when Vadio shows the box full of letters found in Sena's house. The heat maps show the concentration of gaze points recorded while watching the external observer version (up) and the 1-PP version (bottom). A red area indicates that the majority of the users were looking in the same direction at the same time.

Looking at how the heatmaps evolve, some additional considerations, in line with previous studies, can be made:

- The light and the actors' gaze seems to be an efficient method for attention guidance as seen respectively in [4] and [6]. The position of the Agent for most of the video

duration is on a side of the room; for the external observer users, this position is outside the cone of light projected by the spotlight in the room. Looking at the heatmaps, the visual attention of the majority of the external observer users is redirected to the Agent by two events: the Agent comes inside the cone of light (Figure 38) and Vadio looks at the Agent (Figure 39).

Fig.38: Frames from the external observer version of Oreste ancora vivo. Both moments show the Agent inside the cone of light and as a target of the majority of the users' gazes (red area).

Fig.39: Frame from the external observer version of Oreste ancora vivo. In this image, Vadio is looking at the Agent who also becomes the target of the majority of the users' gazes (red area).

In the 1-PP version, the POV enables that the Agent position is next to the light source; this might have affected the results relative to the *PercFixInside* discussed above. The influence of Vadio's gaze has been seen also for the 1-PP version (Figure 40);

- The red area of the heatmaps are most of the time in correspondence with the faces of the characters. Specifically, in the external observer version, most of the users were following the dialogue redirecting their gazes to the character who was speaking (Figure 41); for the 1-PP version instead, the users were mostly focusing on Vadio's faces since he was the only character they could see speaking.

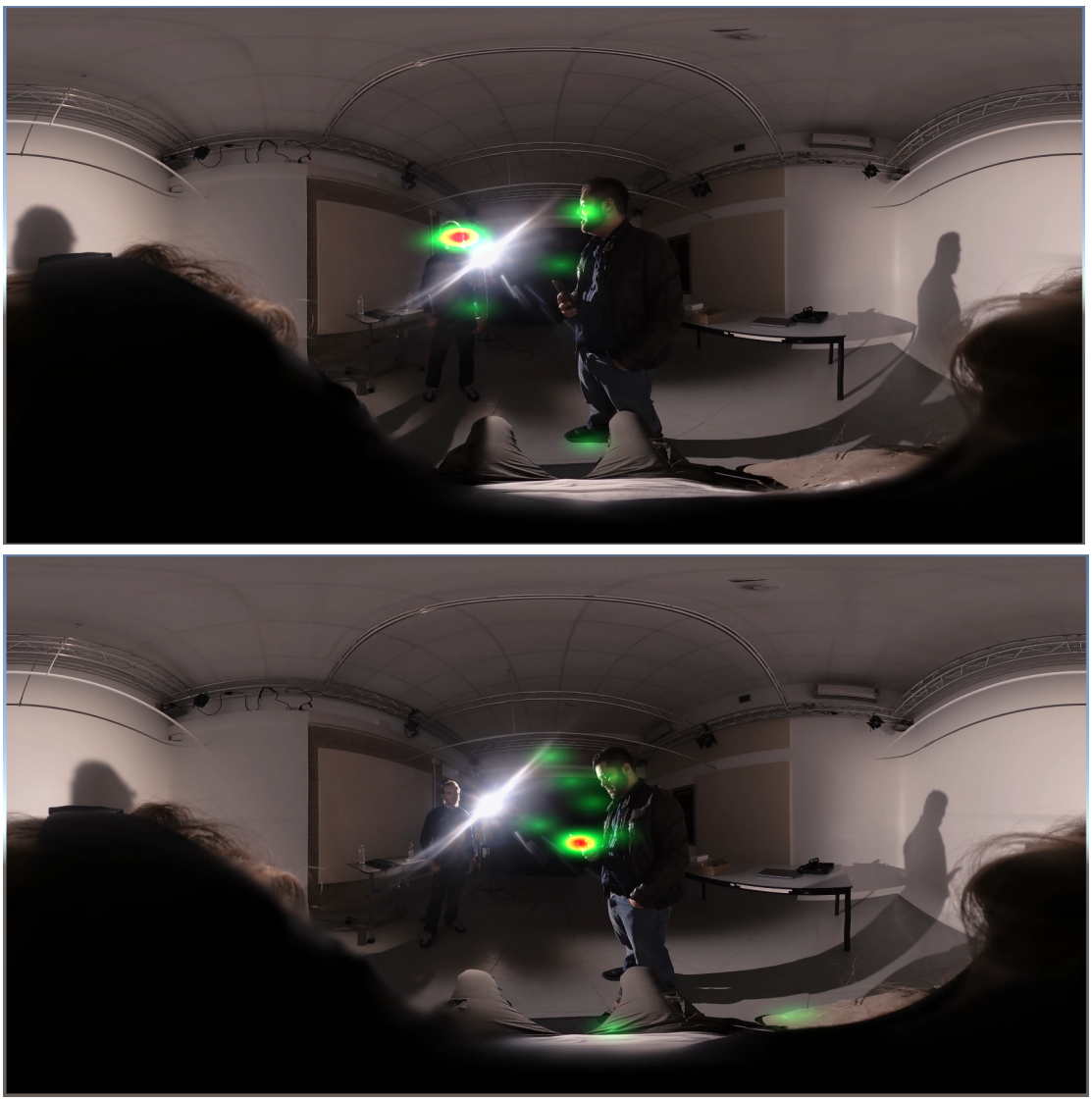


Fig.40: Frames from the 1-PP version of Oreste ancora vivo. (Top) The target of the majority of the users' gazes (red area) is the Agent, also target of Vadio's gaze. (Bottom) Vadio looking at his phone which is also the target of the majority of the users' gazes (red area).

Fig.41: Frame from the external observer version of Oreste ancora vivo. The red area moves from Vadio to Sena during a part of their dialogue.

6.2 Persone che potresti conoscere

The scene used as training, *Persone che potresti conoscere*, also showed interesting results that partially confirm those observed in the main scene of this work.

6.2.1 Subjective results

As shown in Table 3, even though the scene was not particularly designed to stress the differences between the 1-PP and external POV, the NP and in general the NE showed relevant greater values for the 1-PP version. It can be assumed that also for this scene the users who watched the 1-PP version felt more immersed (NP) and generally engaged with the story (NE) than those who watched the external observer version.

Persone che potresti conoscere						
Metric	M		SD		*p*	*d*
	1-PP	Ex	1-PP	Ex		
NU(-)	**3,4**	3,6667	1,2	1,1926	0,7530	0,2229
AF(-)	**6,8667**	7,7333	2,6043	3,2551	0,4363	0,2940
NP	**16,2**	11,9333	2,104	3,5678	0,0003	-1,4568
EE	**18**	15,5333	2,3381	4,6024	0,1607	-0,6758
NE	**23,9333**	16,0667	5,6032	9,0146	0,0113	-1,0482

Tab.3: Results of the NES for the 1-PP and external observer versions of Persone che potresti conoscere. The (-) indicates reverse coded aspects.

These results suggest that the NP factor may be the most sensible to the effect of the POV among the four factors of the NE. For the other three factors, even though the 1-PP version showed on average better results, no significant difference was observed. One of the reasons could be that these factors might be particularly influenced by the kind of story that is considered. In fact, this scene was intended to have a strong emotional impact without

particular attention on the differences between the two POVs. Unlike the main scene, *Persone che potresti conoscere* had no visual or even audio element that could create stress differences among all the factors of the NE.

6.2.2 Objective results

The objective data for to the scene *Persone che potresti conoscere* are presented in Table 4. As said, statistically significant differences are highlighted by filling in green the *p*-value field. Looking at the percentage of fixations (*nFix*), it can be said that for this scene the users who watched the 1-PP version showed a more explorative behaviour ($p < 0,05$).

Persone che potresti conoscere						
Metric	M		SD		*p*	*d*
	1-PP	Ex	1-PP	Ex		
nFix	0,4988	**1,0377**	0,1570	0,1948	0,0001	3,0453
PercFixInside Phone	**5,78533**	3,3633	4,1960	1,6707	0,1485	-0,7584
PercFixInside Fabio	**60,11**	15,1707	23,6982	8,4351	0,0001	-2,5265
Head path	212,6193	**249,3275**	95,4156	90,6126	0,2017	0,3945
Gaze path	2103,747	**3296,651**	1024,563	887,9636	0,0043	1,2443
Experiential Fidelity Phone	**24,6807**	12,8347	18,1665	20,723	0,0329	-0,6079

Tab.4: Results of the eye tracking data collected for the scene Persone che potresti conoscere.

Regarding the percentage of fixations inside a specific ROI, as predictable, the 1-PP version showed significantly greater values for the ROI labeled as Fabio ($p < 0,05$); in fact, Fabio is the character seated in front of the protagonist (Luca) and in the 1-PP version he is directly speaking to the camera. In the external observer instead, the viewer had the opportunity to have the two main characters, Sena and Vadio, in the same field of view; for that reason, the user could easily follow the dialogue just by switching the gaze between the characters. As seen in the main scene, also in this case the *gaze path metric* showed significantly greater values in the external observer subjects. As mentioned before, the reason might be that in this version the users were switching their gaze between the characters to better follow the dialogue. No statistically significant differences were shown for the *head path* this time, probably because there were not enough elements to stimulate the head re-orientation. Regarding the *Experiential Fidelity*, also for this scene the 1-PP users showed substantial better results ($p < 0,05$) for the ROI labeled as Phone (the only one that was considered valuable for the story at a certain time). As said for the previous scene, the larger scale of the phone in the 1-PP version may have played an important role in this result.

6.3 Other results

An additional analysis was performed comparing the same version of the two different scenes. The comparison of the results for the 1-PP versions of both the scenes is presented in Table 5. Looking at the statistically significant differences ($p < 0,05$), it can be said that the main scene (*Oreste è ancora vivo*) created more problems of comprehension (NU) whereas the other scene (*Persone che potresti conoscere*) generated more problems focusing on the main action (AF). The first result could be attributed to the fact that the main story had a more complex plot and it was set in a dystopian future, pretty far from reality. This result seems to be confirmed also in the external observer versions (Table 6). The difference related

to the AF should requires further analysis; the reason for this result might be attributed to the richest environment (in terms of quantity of lights and objects) in the scene *Persone che potresti conoscere*, which could have made the users more prone to wander on it. However, this result was not confirmed in the comparison between the external observer versions.

1-PP						
Metric	M		SD		*p*	*d*
	Oreste	Persone	Oreste	Persone		
NU(-)	4,4667	**3,4**	1,49963	1,2	0,0476	0,746
AF(-)	**4,8667**	6,8667	2,06128	2,6043	0,0095	-0,905
NP	15,6	**16,2**	4,58693	2,104	0,8091	-0,1595
EE	14,8	**18**	4,84699	2,3381	0,0731	-0,6297
NE	21,0667	**23,9333**	8,75189	5,6032	0,5391	-0,2881

Tab.5: Results of the NES for the 1-PP versions of the scenes Oreste è ancora vivo (Oreste) and Persone che potresti conoscere (Persone). The (-) indicates reverse coded aspects.

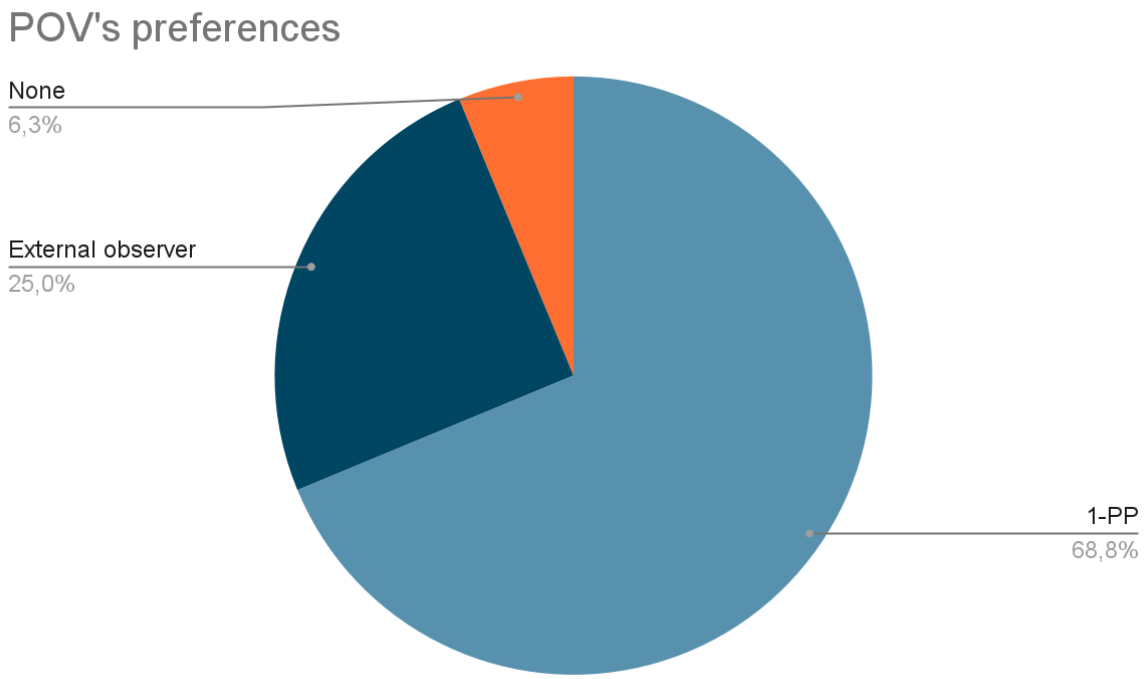
External Observer						
Metric	M		SD		*p*	*d*
	Oreste	Persone	Oreste	Persone		
NU	7,266667	**3,8**	3,714237	1,22202	0,000668	1,253831
AF	**7,266667**	7,8	3,395422	3,187475	0,832985	-0,16196
NP	10,93333	**11,86667**	3,37573	3,537733	0,323736	-0,26993
EE	8,8	**15,13333**	3,208323	4,800926	0,000446	-1,55114
NE	5,2	**15,4**	10,32279	8,845338	0,005859	-1,06112

Tab.6: Results of the NES for the 1-PP versions of the scenes Oreste è ancora vivo (Oreste) and Persone che potresti conoscere (Persone). The (-) indicates reverse coded aspects.

In Table 6, besides the significant difference shown for the NU factor, it is possible to highlight two further statistically significant differences between the two scenes (external observer versions): the EE and NE. The scene *Persone che potresti conoscere* generally created more emotional involvement with the character and the story; this might be attributed to the fact that the story was quite close to the reality and more dramatic. One of the users in fact commented: "I felt more sad in the first scene due to the fact that I could more easily imagine it as a real story."

The results relative to the NE are direct consequences of the differences in the NU and NE factors.

Finally, regarding the preference of the type of POV, the 1-PP was chosen by the majority of the users (Graph 1).



Graph 1: Preferences of the participants of the user study in relation to the POV.

Specifically, 22 participants preferred the 1-PP as it helped them more to immerse in the story and to create empathy with the characters. These reasons are coherent with the results

observed for the NP and EE factors. The interaction, even though not direct, between the characters and the users seemed to help the sense of immersion and transportation. Additionally, most of the users who preferred the 1-PP argued that the external observer POV does not give a different experience from watching a traditional movie.

One participant said “this is the potential of virtual reality, you are at the center of the story.” Eight participants preferred the external observer POV; most of them considered it hard or even “weird" to impersonate a character with a totally different voice, body and personal background. “It bothered me that the character could move and I couldn’t”, was one of the comments made by a user. Also some of the participants who chose the 1-PP indicated problems related to the embodiment, especially because of the different height of the actors.

Only two participants had no preference. They considered the POV strictly dependent on the kind of story that one wants to represent. One of the participants said “I think it depends on the story; the 1-PP helps to get more into the drama of the characters. You can really feel every movement of the actor, like his hand-shaking. However the external observer POV gave me more sense of the environment around me.”

7 Conclusions and future work

This work explored the impact of point of view in 360° video scenes, considering as the main metric the NE. The main hypothesis was confirmed thanks to the results collected through a user study: the users who watched the 1-PP versions showed higher values of engagement for both the scenes that were produced in this work. The objective data showed that the 1-PP version provided better results in terms of gaze fixation on valuable elements of the story. Summarising the results of this work, it can be stated that:

- the 1-PP seems to be the most valid POV to enable a general involvement in the story;
- the sense of immersiveness in the story and its environment (NP) seems to be more easily recreated by the 1-PP;
- the emotional engagement, the narrative understanding and the attentional focus seem to be more affected by the kind of story than to the kind of POV;
- the 1-PP appears to better guide the attention of the users on the elements that are considered important at a certain time of the story.

There were several limitations to this study. The two videos of each scene are not perfectly equal and synchronized. Using real actors makes it impossible to have two perfect scenes with exact timing, gestures and moves. Two possible solutions to this issue can be using scenes created in CG or shooting the scene with two 360° cameras at the same time (one for each POV). For the latter option, there would be the problem of how to make the camera used for the 1-PP invisible to the one used for the external observer. Using small transparent supports and a special case (such as a green colored case) for the camera could help to tackle this problem with additional post-production processes (such as chroma key and mask layers).

Furthermore, the user study included participants with similar backgrounds in relation to CVR and within a narrow age group. Future studies may include users with a wider variety of backgrounds (such as VR content creators or even VR gamers) and different age ranges. There is a wide range of possible future works related to this study and its results. One of the options can be considering scenes of different genres and with a wider range of actions.

The dialogue scenes presented in this work were mainly static with a small range of movements for the actors; also the characters representing the POV in the 1-PP versions kept their sitting posture for the whole duration of the scene (same as the user). A study with more dynamic elements and actions may be implemented to check whether it brings different results. For example, it can be interesting to use scenes where the character that detains the POV in the 1-PP starts to walk or even run. In this case, the difference between the motion of the character and of the users may also influence the level of engagement. The user could feel less involved in case the character's actions are extremely incompatible with his or her posture. Additionally, the movements of the camera can cause motion sickness, which is also a factor that may influence the enjoyment of a cinematic experience.

In both the scenes, the action was concentrated in a limited portion of the 360° area and the rest of the environment did not have any particular visual and audio elements that could distract the user's attention. Exploring the influence of the POV in scenes such as a battle where each portion of the 360° environment presents different actions may lead to different results. In that case, having a bigger portion of the environment visible (like in the external observer POV) could support a better understanding of the scene.

Furthermore, the scenes used in this work were entirely produced in an academic environment; further studies might use professional content.

Another possibility might be giving the choice to the user to switch dynamically between POVs while watching the movie. This choice might include the POVs of different characters

of switching between internal or external observers. Introducing additional elements of interaction for the user (such as 6 DoF for the movements) can be another variable that might affect the incidence of the POV. In this work, it was studied from the first-person perspective when the user can only control the orientation of the camera. Some users, while watching the 1-PP version, struggled to embody a character that had a completely different body and did not follow their moves. With the aim to bring the 1-PP to the next level, it might be fascinating to study human behaviours in movies where the user is able to control the view and the body of a character (such as an avatar) and even to use his or her own voice. However, this solution might bring the CVR in a direction closer to VRy video games.

The scenes of this study had only stereo audio (two channels); additional comments regarding the audio were made by two users who suggested that spatialized audio might help an additional feeling of immersion. Future studies might also include this element.

Another possibility for future study could be considering new metrics; alternatively, since the NE includes different elements, future works might focus on specific factors. For example it would be interesting to further investigate the effect of POV on the sense of presence; this factor, in fact, appeared to be the most affected during this study for both the scenes.

The fact that many users looked for a totally new experience from traditional cinema could be an important perspective in relation to the creation of future VR movies. The users are not looking for what they have already watched on their TVs or laptops; CVR is a new media and, consequently, needs new methods to properly transmit its story and emotions. This work can be considered as an additional step towards the formation process of the immersive cinema language. The hope is that it can be useful for supporting future VR directors in their decisions regarding which POV to select for their works. Further steps must be done in order to make this new media engaging and attractive for a large audience.

References

[1] How virtual reality can create the ultimate empathy machine, Chris Milk, TED2015

[2] Sylvia Rothe,Lang Zhao, Arne Fahrenwalde, Heinrich Hußmann. How to Reduce the Effort: Comfortable Watching Techniques for Cinematic Virtual Reality. Augmented Reality, Virtual Reality, and Computer Graphics pp 3-21. 2020.

[3] Rothe, Sylvia & Buschek, Daniel & Hussmann, Heinrich. (2019). Guidance in Cinematic Virtual Reality-Taxonomy, Research Status and Challenges. Multimodal Technologies and Interaction. 3. 19. 10.3390/mti3010019.

[4] Rothe S., Hußmann H. (2018) Guiding the Viewer in Cinematic Virtual Reality by Diegetic Cues. In: De Paolis L., Bourdot P. (eds) Augmented Reality, Virtual Reality, and Computer Graphics. AVR 2018. Lecture Notes in Computer Science, vol 10850. Springer, Cham.

[5] A. Schmitz, A. MacQuarrie, S. Julier, N. Binetti and A. Steed, "Directing versus Attracting Attention: Exploring the Effectiveness of Central and Peripheral Cues in Panoramic Videos," 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 2020, pp. 63-72, doi: 10.1109/VR46266.2020.00024.

[6] Jayesh S. Pillai and Manvi Verma. 2019. Grammar of VR Storytelling: Analysis of Perceptual Cues in VR Cinema. In European Conference on Visual Media Production (CVMP '19). Association for Computing Machinery, New York, NY, USA, Article 10, 1–10. DOI:https://doi.org/10.1145/3359998.3369402

[7] Gödde, Michael & Gabler, Frank & Siegmund, Dirk & Braun, Andreas. (2018). Cinematic Narration in VR – Rethinking Film Conventions for 360 Degrees. 184-201. 10.1007/978-3-319-91584-5_15.

[8] Gorisse G, Christmann O, Amato EA and Richir S (2017) First- and Third-Person Perspectives in Immersive Virtual Environments: Presence and Performance Analysis of Embodied Users. Front. Robot. AI 4:33. doi: 10.3389/frobt.2017.00033

[9] Katharina Emmerich, Andrey Krekhov, Sebastian Cmentowski, and Jens Krueger. 2021. Streaming VR Games to the Broad Audience: A Comparison of the First-Person and Third-Person Perspectives. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, Article 445, 1–14. DOI:https://doi.org/10.1145/3411764.3445515

[10] Ville Mäkelä, Tuuli Keskinen, John Mäkelä, Pekka Kallioniemi, Jussi Karhu, Kimmo Ronkainen, Alisa Burova, Jaakko Hakulinen, and Markku Turunen. 2019. What Are Others Looking at? Exploring 360° Videos on HMDs with Visual Cues about Other Viewers. In Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '19). Association for Computing Machinery, New York, NY, USA, 13–24.

[11] Balint, Katalin & Schoft, Chantal & Rooney, Brendan. (2017). Depicting Violence: The Effect of Shot Scale, Shot Length and Camera Perspective on Narrative Engagement with Violent Films.

[12] Cummins, R. G. (2009). The Effects of Subjective Camera and Fanship on Viewers' Experience of Presence and Perception of Play in Sports Telecasts. Journal of Applied Communication Research, 37(4), 374–396. https://doi.org/10.1080/00909880903233192

[13] Cummins, R. G., Keene, J. R., & Nutting, B. H. (2012). The Impact of Subjective Camera in Sports on Arousal and Enjoyment. Mass Communication and Society, 15(1), 74–97. https://doi.org/10.1080/15205436.2011.558805

[14] Andrew K. Przybylski, Kou Murayama, Cody R. DeHaan, Valerie Gladwell,Motivational, emotional, and behavioral correlates of fear of missing

out,Computers in Human Behavior,Volume 29, Issue 4,2013,Pages 1841-1848, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2013.02.014.

[15] Arbresh Ujkani, Jan Willms, Lezgin Turgut, and Katrin Wolf. 2019. The Effect of Camera Perspectives on Locomotion Accuracy in Virtual Reality. In Proceedings of Mensch und Computer 2019 (MuC'19). Association for Computing Machinery, New York, NY, USA, 835–838. DOI:https://doi.org/10.1145/3340764.3344918

[16] Gorisse G, Christmann O, Amato EA and Richir S (2017) First- and Third-Person Perspectives in Immersive Virtual Environments: Presence and Performance Analysis of Embodied Users. Front. Robot. AI 4:33. doi: 10.3389/frobt.2017.00033

[17] A. S. Won, T. Aitamurto, B. Kim, S. Sakshuwong, C. Kircos and Y. Sadeghi, "Motivation to Select Point of View in Cinematic Virtual Reality," 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 2019, pp. 1233-1234, doi: 10.1109/VR.2019.8798184.

[18] R. Cao, J. Walsh, A. Cunningham, C. Reichherze, S. Dey and B. Thomas, "A Preliminary Exploration of Montage Transitions in Cinematic Virtual Reality" 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2019, pp. 65-70, doi: 10.1109/ISMAR-Adjunct.2019.00031

[19] Rick Busselle & Helena Bilandzic (2009) Measuring Narrative Engagement, Media Psychology, 12:4, 321-347, DOI: 10.1080/15213260903287259

[20] Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. New York, NY: Harper & Row.

[21] Bilandzic, H., Sukalla, F., Schnell, C., Hastall, M. R., & Busselle, R. W. (2019). The Narrative Engageability Scale: A multidimensional trait measure for the propensity to become engaged in a story. International Journal of Communication, 13, 801–832.

[22] Kate Carey, Emily Saltz, Jacob Rosenbloom, Mark Micheli, Judeth Oden Choi, and Jessica Hammer. 2017. Toward Measuring Empathy in Virtual Reality. In Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play(CHI PLAY '17 Extended Abstracts). Association for Computing Machinery, New York, NY, USA, 551–559. DOI:https://doi.org/10.1145/3130859.3131325

[23] P. A. Punde, M. E. Jadhav and R. R. Manza, "A study of eye tracking technology and its applications," 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), 2017, pp. 86-90, doi: 10.1109/ICISIM.2017.8122153.

[24] Ana Serrano, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, Belen Masia. Movie Editing and Cognitive Event Segmentation in Virtual Reality Video. ACM Transactions on Graphics, Vol. 36, No. 4, Article 47. 2017.

[25] C. Marañes, D. Gutierrez and A. Serrano, "Exploring the impact of 360° movie cuts in users' attention," 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 2020, pp. 73-82, doi: 10.1109/VR46266.2020.00025

[26] Kevin Kantor, "Please come off-book",2021, https://buttonpoetry.com/product/please-come-off-book/

[27] Janet H. Murray, "Not a Film and Not an Empathy Machine", 2016, https://immerse.news/not-a-film-and-not-an-empathy-machine-48b63b0eda93

[28] Nonny de la Peña (2015). The future of new? Virtual reality.TEDWomen 2015.

[29] V. Sitzmann et al., "Saliency in VR: How Do People Explore Virtual Environments?," in IEEE Transactions on Visualization and Computer Graphics, vol. 24, no. 4, pp. 1633-1642, April 2018, doi: 10.1109/TVCG.2018.2793599.

[30] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 symposium on Eye tracking research

& applications (ETRA '00). Association for Computing Machinery, New York, NY, USA, 71–78. DOI:https://doi.org/10.1145/355017.355028

[31] Blignaut, Pieter. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. Attention, perception & psychophysics. 71. 881-95. 10.3758/APP.71.4.881.

[32] Dalmaijer, E.S., Mathôt, S., & Van der Stigchel, S. (2013). PyGaze: an open-source, cross-platform toolbox for minimal-effort programming of eye tracking experiments. Behaviour Research Methods. doi:10.3758/s13428-013-0422-2

[33] J. O. Wallgrün, M. M. Bagher, P. Sajjadi and A. Klippel, "A Comparison of Visual Attention Guiding Approaches for 360° Image-Based VR Tours," 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 2020, pp. 83-91, doi: 10.1109/VR46266.2020.00026.

[34] Kurby CA, Zacks JM. Segmentation in the perception and memory of events. Trends Cogn Sci. 2008 Feb;12(2):72-9. doi: 10.1016/j.tics.2007.11.004. PMID: 18178125; PMCID: PMC2263140.

[35] Sylvia Rothe, Harald Brunner, Daniel Buschek and Heinrich Hußmann. 2018. Spaceline: A Way of Interaction in Cinematic Virtual Reality. In Proceedings of ACM SUI'18, Berlin, Germany, 1 page, https://doi.org/10.1145/3267782.3274675

[36] K.Rahimi, C. Banigan and E. D. Ragan, "Scene Transitions and Teleportation in Virtual Reality and the Implications for Spatial Awareness and Sickness," in IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 6, pp. 2273-2287, 1 June 2020, doi: 10.1109/TVCG.2018.2884468.

[37] T. Stebbins and E. D. Ragan, "Redirecting View Rotation in Immersive Movies with Washout Filters," 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 2019, pp. 377-385, doi: 10.1109/VR.2019.8797994.

[38] Amarnath Murugan, Jayesh S. Pillai, and Amal Dev. 2019. Cinévoqué: Development of a Passively Responsive Framework for Seamless Evolution of Experiences in Immersive Live-Action Movies. In 25th ACM Symposium on Virtual Reality Software and Technology (VRST '19). Association for Computing Machinery, New York, NY, USA, Article 59, 1–2

[39] A. MacQuarrie and A. Steed, "Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video," 2017 IEEE Virtual Reality (VR), 2017, pp. 45-54, doi: 10.1109/VR.2017.7892230.

[40] A. Kim, M. Chang, Y. Choi, S. Jeon and K. Lee, "The Effect of Immersion on Emotional Responses to Film Viewing in a Virtual Environment," 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 2018, pp. 601-602, doi: 10.1109/VR.2018.8446046.

[41] Zielasko, Daniel, and Bernhard E. Riecke 2021. "To Sit or Not to Sit in VR: Analyzing Influences and (Dis)Advantages of Posture and Embodied Interaction" Computers 10, no. 6: 73. https://doi.org/10.3390/computers10060073

[42] Sawilowsky, Shlomo S. (2009) "New Effect Size Rules of Thumb," Journal of Modern Applied Statistical Methods: Vol. 8 : Iss. 2 , Article 26. DOI: 10.22237/jmasm/1257035100 Available at: http://digitalcommons.wayne.edu/jmasm/vol8/iss2/26

Acknowledgements

First of all, I would like to thank the people who gave me the possibility to realise this work and supported me during all the its process: my supervisor, Prof. Fabrizio Lamberti, always available to share his experience and to clarify my doubts; my co-supervisor Dr. Alberto Cannavò, constantly present in every single step of this research; Dr. Gabriele Filippo Pratticò, a fundamental support for my lack of technical knowledge; Prof. Tatiana Mazali who introduced me to the world of cinematic VR.

I want to thank my family, without them I could not live this amazing experience as a student of the Polytechnic of Turin. To Tullio, Dino and Peppe the worst group of best friends I could ever have; without you I might have been a better student but for sure not a better human being. To Monika, in these two years I could feel her support even at 1500 km of distance. To Via Baretti 45, the best house in Turin. Thanks to all my university colleagues, facing all the projects and exams has been easier with your company. A special thanks to the actors that have played in my two immersive movies: Francesco, Gabriele, Angelica and Nicolò, I wish you the best future as professional actors. Thanks to all the people who volunteered for my test sessions.