Immersive Movies: The Effect of Point of View on Narrative Engagement

AI & SOCIETY
https://doi.org/10.1007/s00146-022-01622-9

OPEN FORUM

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

Alberto Cannavò[1] · Antonio Castiello[1] · F. Gabriele Pratticò[1] · Tatiana Mazali[2] · Fabrizio Lamberti[1]

Received: 28 July 2022 / Accepted: 21 December 2022

Abstract
Cinematic virtual reality (CVR) offers filmmakers a wide range of possibilities to explore new techniques regarding movie scripting, shooting and editing. Despite the many experiments performed so far both with both live action and computer-generated movies, just a few studies focused on analyzing how the various techniques actually affect the viewers' experience. Like in traditional cinema, a key step for CVR screenwriters and directors is to choose from which perspective the viewers will see the scene, the so-called point of view (POV). The aim of this paper is to understand to what extent watching an immersive movie from a specific POV could impact the narrative engagement (NE), i.e., the viewers' sensation of being immersed in the movie environment and being connected with its characters and story. Two POVs that are typically used in CVR, i.e., first-person perspective (1-PP) and external perspective (EP), are investigated through a user study in which both objective and subjective metrics were collected. The user study was carried out by leveraging two live action 360° short films with distinct scripts. The results suggest that the 1-PP experience could be more pleasant than the EP one in terms of overall NE and narrative presence, or even for all the NE dimensions if the potential of that POV is specifically exploited.

Keywords Omnidirectional · 360° · Immersive videos · Cinematic VR · First-person perspective · External perspective · User study

1 Introduction

The release of an ever-growing number of commercial head-mounted displays (HMDs) like the Oculus Quest 2 and the HTC Vive Pro, together with the development of solutions enabling affordable virtual experiences like the Google Cardboard, is promoting the interest in VR for home entertainment (Moghadam and Ragan 2017). The popularity of this medium has stimulated the growth of new interactive narratives for entertainment purposes (Stebbins and Ragan 2019). More and more immersive cinematic experiences are produced by VR companies (like, e.g., Baobab Studios[1] and Immersive Studios[2]) as short stories or movies, opening new opportunities to experiment with alternative approaches to storytelling and user interaction within the developed experiences (Stebbins and Ragan 2019).

Immersive movies started to be produced in Computer Graphics by making use of animation suites (such as Blender and Maya) or game engines (like Unity or Unreal Engine), and as live actions, i.e., as recordings of real-world scenes captured with 360° cameras like the GoPro Max or the Samsung Gear 360. However, there is a lack in the research literature and among practitioners for what it concerns the experience and/or the expectation of the users when they watch immersive movies (Marañes et al. 2020).

✉ Alberto Cannavò
alberto.cannavo@polito.it

✉ Fabrizio Lamberti
fabrizio.lamberti@polito.it

Antonio Castiello
antonio.castiello@studenti.polito.it

F. Gabriele Pratticò
filippogabriele.prattico@polito.it

Tatiana Mazali
tatiana.mazali@polito.it

[1] Politecnico di Torino, Dipartimento di Automatica e Informatica, Corso Duca Degli Abruzzi 24, 10129 Turin, Italy

[2] Politecnico di Torino, Dipartimento Interateneo di Scienze, Progetto e Politiche del Territorio, Viale Pier Andrea Mattioli, 39, 10125 Turin, Italy

[1] Baobab Studios: https://www.baobabstudios.com/.

[2] Immersive Studios: https://weareimmersive.co.uk/.

Published online: 12 January 2023

Differently than with traditional movies in which a well-established cinematographic language was developed over a century of continuous developments, the production of immersive movies is still undergoing an initial phase made up of experimentations (Marañes et al. 2020). For this reason, both researchers and content creators are still working on developing a new narrative language that is able to improve the effectiveness of the VR medium to leverage its full potential (Marañes et al. 2020; Stebbins and Ragan 2019; Sitzmann 2018; Xue et al. 2021).

Among the various research directions to be explored in this context, the point of view (POV) is becoming of paramount importance, since it represents the perspective from which the viewers perceive the story, and it can affect what they actually experience (in terms of both images and sounds). In traditional cinema, the director is in charge of defining the scenes' POV by choosing the positions of the camera during the shots (Marañes et al. 2020). However, in VR, the cameras are controlled/worn by the viewers, who can freely decide where to look in the 360° scene. As a result, the viewers may not look in the direction where the main narrative actions are taking place.

According to the taxonomy reported in (Ruscella and Obeid 2021), it is possible to identify two main POVs in CVR, which the authors refer to as two levels of embodiment, i.e., the perceived distance between the viewer (the VR user, in this case), and the experience. The first level, in this paper referred to as external perspective, or EP in short, refers to a sort of detached view, in which the viewer observes the scene from a disembodied POV. The viewer feels as part of the experience, but he or she is like an external observer of the actions happening in the environment. In this case, the camera is settled in the position that the director considers the best one for watching the movie, and it is allowed to make some smooth movements (like in traditional movies). An example of the EP view setup is provided by the movie "Help".[3] The second level of embodiment offered by immersive movies relies on a first-person view of the scene (in the following referred to as 1-PP, or first-person perspective). In this case, the viewer observes the scene from a diegetic element of the environment, i.e., a character or an object of the story. To record live actions with this level of embodiment, cameras are worn by the actors at eye level or are mounted onto an object in the scene. Examples of using the 1-PP are, e.g., the movies "The party—A virtual experience of autism"[4] and "Car crash experience in VR".[5]

Despite the numerous examples of immersive movies exploiting these POVs (in some cases also combined), evidence or studies that show how the POV can affect the viewers' experience in CVR are still scarce. This paper tries to take some steps forward in this direction by focusing, in particular, on the impact on narrative engagement (NE).

In (Busselle and Bilandzic 2009), the NE is described as the consequence of a mental representation, the mental model, of the story created by the viewers, who are completely posing their attentional focus on the story itself. According to (Busselle and Bilandzic 2009), it is possible to identify three mental models, all relevant for understanding a story: the situation model, which includes the story itself (plot) and all the connections between the actions and events of the story (the causal link); the character model(s), which represent all the possible information regarding the character(s) of the story, like physical and psychological factors; the story world model, which consists of all the information related to the narrative world (logic).

One of the reasons for choosing to focus on NE is the comprehensiveness of this metric, which encompasses emotional, attentional, and cognitive factors (Busselle and Bilandzic 2009). Moreover, existing literature already showed its reliability for the analysis of both traditional and immersive movies (Cummins 2009; Cummins et al. 2012; Cao et al. 2019). Lastly, the NE has a clearer definition and more reliable evaluation methods than other "general-purpose" metrics used in this context like, e.g., enjoyment or empathy (Busselle and Bilandzic 2009; Carey et al. 2017).

By moving from the above considerations, this paper reports on a user study aimed to investigate the influence that different POVs can have on the viewers' NE while watching immersive movies. To this aim, two scripts have been designed and produced in order to stress, in different ways, key aspects of the NE. The designed experiments were conducted by involving 32 participants. Both subjective and objective measurements based on standard questionnaires and state-of-the-art metrics were used to analyze the viewers' experience. The obtained results indicate that, overall, the 1-PP can improve the viewers' NE, by offering them, in particular, a higher degree of narrative presence compared to the EP. Furthermore, if the potential of 1-PP is exploited in the scripting and shooting phases, levels of narrative understanding, attentional focus, and emotional engagement higher than with the EP can also be observed.

2 Related work

CVR grammar is still in its infancy compared to that of traditional filmmaking (Dooley 2020). The authors of (Gödde et al. 2018) identified six key directions that shall be explored to expand the language of this new media: leading the viewer's attention towards salient story elements, setting the mise-en-scene, i.e., how to place actions and story

[3] Help: https://bit.ly/3BsFjDa.

[4] The party: http://youtu.be/OtwOz1GVkDg.

[5] Car crash experience in VR: http://youtu.be/aBiNNgfB7jI.

elements, rethinking framing and editing, balancing the spatial and temporal story density, and defining the viewer's role.

So far, the former is certainly the most explored direction. As a matter of fact, a decent amount of studies focused on proposing guidance methods that can be used to lead the viewer's attention towards a specific part of the scene by transposing approaches from traditional filmmaking. These methods can rely, e.g., on visual cues, either in the form of diegetic or non-diegetic scene elements (Rothe et al. 2019), on static and dynamic lights (Schmitz et al. 2020), on blurring techniques that relate to the conventional focus pull–push (depth separation) (Bender 2019), or on the use of spatialized 3D audio (Rothe and Hußmann 2018).

The remaining five directions are considerably less investigated (Dahl et al. 2021; Bender 2019; Tong et al. 2021). This consideration is especially true for what it concerns understanding the impact of the viewer's role in CVR. In fact, in the language of traditional cinema, the viewer's role is usually not seen as a fixed filming choice, but rather as an opportunity to stimulate different emotional responses by rotating among the various POVs (Branigan 1975; Carroll 1993). A few studies were conducted in the context of traditional telecasting, e.g., considering the broadcasting of sport events (Cummins 2009; Cummins et al. 2012); in this specific context, results showed that the 1-PP can offer a significantly higher sense of presence, emotional arousal, and NE compared to the EP, though not in terms of enjoyment, which was found to depend more on the actual gameplay than on the POV.

It should be noted that the features of CVR actually bring considerations regarding the choice of the POV closer to the those made for written narrative than for traditional cinema. In the former domain, there is consensus about the fact that the narrative perspective (1-PP or EP) has an impact in terms of agency and identification with the story characters (Zhou 2017; Hoeken et al. 2016), and it has been proved that the identification process has a direct influence on the NE (de Graaf et al. 2012). Collectively, the studies in this field identify the 1-PP as superior in terms of the above metric. This aspect has been successfully leveraged, e.g., in works aimed to study behavioral changes in educational contexts like (Chen et al. 2015; Lipsey et al. 2020; Nan et al. 2015), which found that the 1-PP was able to better involve the learner in the process. Nevertheless, the EP was deemed valuable when the narrative message is targeted to an entire audience, rather than to an individual (Dahlstrom and Rosenthal 2018).

Regarding immersive VR, the studies that dealt with POV selection mostly focused on comparing the 1-PP against the 3-PP (third-person perspective), which received attention in particular in the domain of interactive VR experiences. In (Gorisse et al. 2017) no significant differences were found in terms of presence and agency between the two POVs, whereas the 1-PP stimulated higher embodiment and body-ownership compared to the 3-PP. In (Emmerich et al. 2021), the impact of 1-PP and 3-PP was investigated considering the use case of a spectator of immersive VR gaming. However, the viewers were not provided with an immersive view of the VR environment but watched the gameplay on a separate screen.

It should be noted that, while in immersive VR, the 3-PP is considered to be at the opposite end of the POVs spectrum with respect to 1-PP (Hoppe et al. 2022), this may not be also true for CVR. In immersive VR, the implied interactivity of the experience poses a constraint on the fact that the character shall not entirely disappear from the user's sight even though the perspective can be placed out of the user-controlled character's POV (Hoppe et al. 2022). Although this kind of 3-PP is also conceivable in CVR, for instance in footage shot with handheld action cameras in which the perspective is out of the main character's POV but follows it, in the CVR literature it is rather commonplace to use the 3-PP term for referring to any kind of detached view (Gödde et al. 2018; Ruscella and Obeid 2021), regardless of the fact that it is a follow-me, out-of-the-character view or a "fly on the wall" one (Ruscella and Obeid 2021). As a matter of example, works exploiting the "fly on the wall" view are (Cummins 2009; Cummins et al. 2012; Bender 2019; van den Boom et al. 2015). However, it should be considered that, in CVR, the latter can also enable narrative opportunities, e.g., in terms of the so-called Joy of Missing Out (JOMO) (Tanja et al. 2021), like in a footage shot in a room in which the main character can exit the room and disappear from the viewer's sight for a given amount of time for narrative purposes. With the aim to better refer to these two different kinds views, in this paper the term 3-PP is used to indicate the 3-PP in the context of immersive VR, whereas the term EP is proposed to indicate the perspective that, in the literature, is typically referred to as spectator/objective/passive observer or "fly on the wall" (Dooley 2021).

Regardless of the differences between 3-PP and EP, results about the impact of POV in immersive VR are not immediately applicable to CVR since removing the interactive component could also lead to a less engaging experience, with a detrimental effect that could even make the CVR content worse than traditionally filmed content (Wu et al. 2021). A comparison between two experiences at the opposite range of the interactivity continuum (Tong et al. 2021) also showed that removing interaction has a negative impact in terms of NE (Christopher 2020). Conceptually, it is possible to envisage the implementation of the EP also in CVR, e.g., in the case of multi-POV storylines in which the viewer can engage with the story from multiple viewpoints. Although multi-POVs CVR experiences open interesting creative opportunities for which the mise-en-scene and editing (Pillai and Verma

2019) should be specifically investigated, they are not considered in this paper since it would be difficult to isolate the effect of the POV from other factors.

This is confirmed by the fact that the few works that studied the impact of POV in the context of CVR focused on the two extremes of the POVs spectrum (Ruscella and Obeid 2021), comparing the 1-PP against the EP. One example is represented by (Bender 2019). In their study, the authors used two immersive movies with the same script but with differences in the shooting aimed to stress the differences of the two POVs. It was found that both the perspectives were equally able to stimulate attentional synchrony (the collective behavior of viewers whose gaze predominantly clusters onto salient features) and keep the viewers engaged, but the 1-PP was able to stimulate a higher sense of presence and identification with the character compared to the EP. The authors of (Boom et al. 2015), in turn, compared the two POVs without introducing any modification neither to the script nor to the shot. Regarding presence and identification, the results were in line with those of (Bender 2019), confirming the capability of the 1-PP to stimulate these two dimensions more than EP. No significant differences for the two POVs were observed, instead, for engagement, naturalness, and enjoyment. Despite the promising results obtained in (Bender 2019; Boom et al. 2015), more investigations are still needed in this field. In fact, on the one hand, (Bender 2019) specifically targeted only attentional synchrony; on the other hand, (Boom et al. 2015) focused only on presence and enjoyment, partially disregarding emotional, attentional, and cognitive factors that are part of the NE.

This paper aims to widen the current knowledge in this field by investigating the differences between 1-PP and EP through a metric, the NE, which can provide a more comprehensive picture of the viewers' experience while watching immersive movies. Differently than works seen so far, the proposed study takes into account both subjective and objective metrics, using the latter to validate the former. Moreover, compared to (Bender 2019; Boom et al. 2015), the reported study is based on two different scripts, each filmed with the two considered POVs.

3 Methodology work

As said, this paper reports on a user study aimed to analyze the viewers' experience with immersive content. This section provides details on the different steps of the experiments, from the study definition to material design and production, and evaluation tools.

3.1 Material design

To extend the studies in (Bender 2019; Boom et al. 2015) and with the aim to isolate the contribution of the POV, two scenes with different narrative elements were designed and produced, each one in two different versions, 1-PP and EP.

Details of the produced scenes will be provided in Sect. 3.2. As said, in the 1-PP version, the viewers observe the scene from the eyes of the main character. The EP version, in turn, makes the viewers see the scene from a perspective that is not embedded into any of its elements. For the design of the two scenes, the following factors that could have an influence on the study results were considered:

Characters' gaze and movements: these stimuli are known to be capable of guiding the visual attention of the viewer (Pillai and Verma 2019). To avoid possible distractions that could influence the change of the point of interest, no body or gaze movements were performed by the actors that did not point at the salient aspects of the scene or the story.

Lightning: the lights can be used to direct the viewer's gaze (Rothe and Hußmann 2018). To avoid as much as possible their influence, the lights in the set were used to invite the viewer to focus on the main action of each scene.

Distance between the camera, i.e., the viewer, and the main action: to let the viewers see the actions happening in the story and easily recognize the cues offered by the objects in the environment (to understand the events), the shots were recorded by using an empirically defined distance that ensures the focus on the actions. This distance was obtained through a trial-and-error process, based on moving the camera among the trials and requesting few users (who already knew the scripts) to express their opinion regarding the size of the relevant objects when framing the scene from a given camera position. In future work, the subjectivity of this process could be reduced, e.g., by combining the users' feedback with eye-tracking logs.

3.2 Shooting and post-production

Like in (Pillai and Verma 2019), new immersive videos were created in order to study the viewers' experience under controlled conditions. It was chosen to produce new content since the use of existing movies could influence the NE. In fact, the viewers could link their experiences with memories or positive/negative impressions they had when watching those movies for the first time. Moreover, producing new content makes it possible to realize two versions (1-PP and EP) of the same scene, which should make the comparison fairer. To get inspiration for the creation of the immersive movies to be used in the experiments, the following resources, i.e., web platforms and apps that contain professional and amateur videos, were considered: YouTube

VR Channel,[6] Google Spotlights Stories,[7] The Guardian,[8] and Within.[9] Available videos were analyzed to understand the main trends regarding methodologies and languages adopted by content creators. Based on the outcomes of this analysis and under the supervision of a university professor with expertise in immersive movies, two scripts were designed and produced: "Persons you may know" ("Persone che potresti conoscere") and "Oreste is still alive" ("Oreste è ancora vivo"). Both scripts are in Italian language and feature a dialog among several characters. All the names and events described in the two stories are completely fictional. The videos are available as supplemental material.

3.2.1 First script: "Persons you may know"

The first script was inspired by the poem by Kevin Kantor, "People you may know". The story presents a lunch between two brothers: Luca and Fabio. Scrolling on his phone, Luca finds 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. The two characters were played by two young actors who had already participated in some professional productions. The camera used to shoot the scene was a GoPro Max equipped with two ultra-wide-angle lenses. To realize the 1-PP version, the camera was mounted on the head of the main character (i.e., the actor who played Luca) with a GoPro head strap. For the EP version, the camera was mounted on a tripod. Like in (Boom et al. 2015), this was the only intentional difference between the two versions.

Software for post-production (i.e., Adobe After Effect 2021) was used to automatically stitch images gathered by the camera (to obtain a fully 360° movie), edit and assemble recorded shots, add the title and soundtrack, add a mask layer to remove/cover the tripod and the camera support from the images, and stabilize the camera motion since the head of the actor made some small movements during the shooting of the 1-PP version. The resolution of the camera was set to 5 K (5376 × 2688 pixels). The frame rate was set to 30 fps. The audio was recorded with the six integrated microphones of the GoPro Max. The lighting of the scene was implemented by using only the sunlight coming from the window in the room.

The duration of the produced video is 3 min and 33 s for the 1-PP version, 3 min and 52 s for the EP version. 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.

3.2.2 Second script: "Oreste is still alive"

The second script was inspired by two famous dystopian social science fiction novels: "Nineteen eighty-four" by George Orwell, and "The circle" by Dave Eggers. The story deals with a dystopian future in which the Government (called "La Rete", in Italian language, i.e., "The Network") constantly controls people by forcing them not to stay offline for more than 48 h. The vicious head of the security, named 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 by Vadio to Sena. A second security agent is also present.

Differently than in the first script, in this case some specific narrative choices were made, as done in (Bender 2019), to improve the user's 1-PP viewing experience by leveraging the potential of this type of embodiment. This choice was made to investigate the possible impact on the NE. The strategy to differentiate the two versions (detailed in the following) is based on changing some visual and audio elements. Figure 1 provides a comparison of the two versions, by illustrating how the key elements of the story are shown to the viewers in the 1-PP and EP. At the beginning of the scene, the details of the surrounding environment are not visible in the 1-PP version, since Sena's head (holding the camera) is covered by a black hood (Fig. 1a). After a few words spoken by Vadio, the hood is removed, and the viewer can observe the room (Fig. 1b). Furthermore, in the 1-PP version, the viewer can have a detailed view of the elements that are shown to Sena by Vadio, i.e., the content displayed on Vadio's mobile phone representing the face of Oreste (Fig. 1c) and a countdown for Sena's execution (Fig. 1d), a message on a letter telling that Oreste is still alive (Fig. 1e), and the inside of a box containing possible proofs of the fact that Sena is involved (Fig. 1f). Finally, in the 1-PP version, it is possible to hear the thoughts of Sena with some details on the story. The two main characters, Sena and Vadio, were played by young actors with theater and cinema experiences, whereas the third character (the second security agent), who had no lines in the script, was played by a person without prior professional acting experience. The camera settings for this scene were the same used for the previous scene in terms of resolution and frame rate.

For the 1-PP version, the GoPro Max was mounted on the head of the main actress (who played Sena), whereas for the EP version the camera was settled on a tripod. The lighting of this scene was made by using only a 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

[6] YouTube VR Channel: http://tiny.cc/ddzouz.

[7] Google Spotlights Stories: http://tiny.cc/gdzouz.

[8] The Guardian: http://tiny.cc/hdzouz.

[9] Within: https://www.with.in/.

Fig. 1 Key elements in the second script as viewed in the 1-PP version (top) and the EP version (bottom)

of the dialogs was recorded using the microphones of the camera. Sena's thoughts were captured in a dedicated recording session using a ZOOM H5 handy recorder. The movie underwent the same post-production steps described for the first script. In addition, to differentiate the spoken parts and Sena's thoughts, a reverb effect was added to the recorded audio of the thoughts using the Audacity software. The duration of the 1-PP version is 4 min and 37 s, whereas that of the EP version is 4 min and 31 s.

3.3 Metrics

3.3.1 Subjective metrics

Subjective measurements were collected through a questionnaire, in Italian (available as supplemental material), which included three sections. The first section was aimed to evaluate the participants' familiarity with the technology/content related to the experiments. The second section measured the NE by making use of the NE scale (NES) proposed in (Busselle and Bilandzic 2009). In particular, the NES consists of a 12-item questionnaire that evaluates the NE by considering four dimensions, i.e., narrative understanding (NU), attentional focus (AF), narrative presence (NP), and emotional engagement (EE).

For each item, the participants have to tell how much they agree with the provided statement on a Likert Scale from 1 (strongly disagree) to 7 (strongly agree). The 12 items are related to the four dimensions above (three items per dimension). All the items related to the AF and NU have a negative impact on the overall score since they are formulated with a negative connotation. This section had to be completed for each immersive content viewed. For some of the items, examples were provided to make it easier for the participants to understand the meaning of the statement.

Finally, in the last section, the participants were asked about which POV (1-PP or EP) they appreciated more overall, their motivations, and whether they had additional comments on the experience.

3.3.2 Objective metrics

Data regarding the gaze and eyes movements were logged during the experiments with a sampling rate of 50 Hz. The sampling rate was selected by considering relevant studies that include eye-tracking data analysis, like (Sitzmann 2018; Serrano et al. 2017). Raw data were used to compute the following metrics:

nFix: a percentage value representing the ratio between the number of fixations identified in the data over the total number of collected gaze points. A fixation corresponds to a set of gaze points close in time and range, which occurs when the eyes are fixed on a particular element. Typically, it has a minimum duration of 50 ms and an average duration of 200 ms (Punde et al. 2017). This metric was used in (Rothe et al. 2020 and Sitzmann 2018), and describes how many saccades and fixations the viewers execute; a high value is an indication of many fixations (or few saccade movements), which might represent an indication of the fact that the viewer preferred to concentrate, e.g., on a detail of a specific scene element—once reached the element with the gaze—rather than spanning its whole surface.

PercFixInside: a percentage value representing the number of fixations within a specific Region of Interest (ROI) over the total number of fixations. This metric was introduced in (Sitzmann 2018) and used in some other studies related to VR, like (Rothe et al. 2020). It offers an indication of the viewer's interest in a specific ROI.

Experiential fidelity with Intended POV and ROI: an extension of the metric introduced in (Pillai and Verma 2019), which indicates the percentage of time spent by the viewer in observing a given ROI over the time window in which the given element is considered valuable. ROIs are considered valuable in a specific time interval when they contain elements of the scene that are fundamental for the story understanding (for instance, an important object that is visible only for few time instants, or an object that provides crucial information about the story's characters and events).

Head and Gaze paths: the distances traveled by the viewer's head and gaze while watching the immersive content. Paths were computed in pixel/second by dividing the total traveled distance of the pointer representing the head and the gaze in a given frame (in pixels) by the video length (in seconds). The average speed was computed since, as said, the 1-PP and EP versions of the two scenes were not perfectly aligned in terms of duration. A high value of the head path means that the viewer has changed many times his or her field of view (the visible portion of the 360° video); this outcome might indicate an exploratory behavior. A high value of the gaze path, instead, could indicate that the viewer has been moving his or her eyes (pupils) frequently even without changing the field of view.

Head and gaze tracking data used in the computation of the metrics were collected as latitude values (i.e., in terms of the Euler angle along the view direction parallel to the ground, ranging from − 90° to 90°) and longitude values (i.e., in terms of the Euler angle along the view direction perpendicular to the ground, ranging from − 180° to 180°). Euler angles were converted to pixels in the 2D image coordinate system, where the image is a frame of the 360° video. In this system, the x axis corresponds to the pixels along the width of the image, whereas the y axis is mapped onto the pixels along the height of the image. The origin is in the bottom-left corner. Besides enabling the computation of the objective metrics, tracking data can also be used to visually observe the movements of the participants' gaze and head during the experience. This is possible, e.g., through the constructions of interactive heatmaps, which are tools commonly adopted in the literature to support findings in this field (Marañes et al. 2020). Videos showing the computed heatmaps, as well as discussion supporting the obtained results are available in the Appendix, provided as supplemental material. Some sample frames are provided in Fig. 2.

Fixation was detected using the Identification by Dispersion-Threshold (I-DT) algorithm introduced in

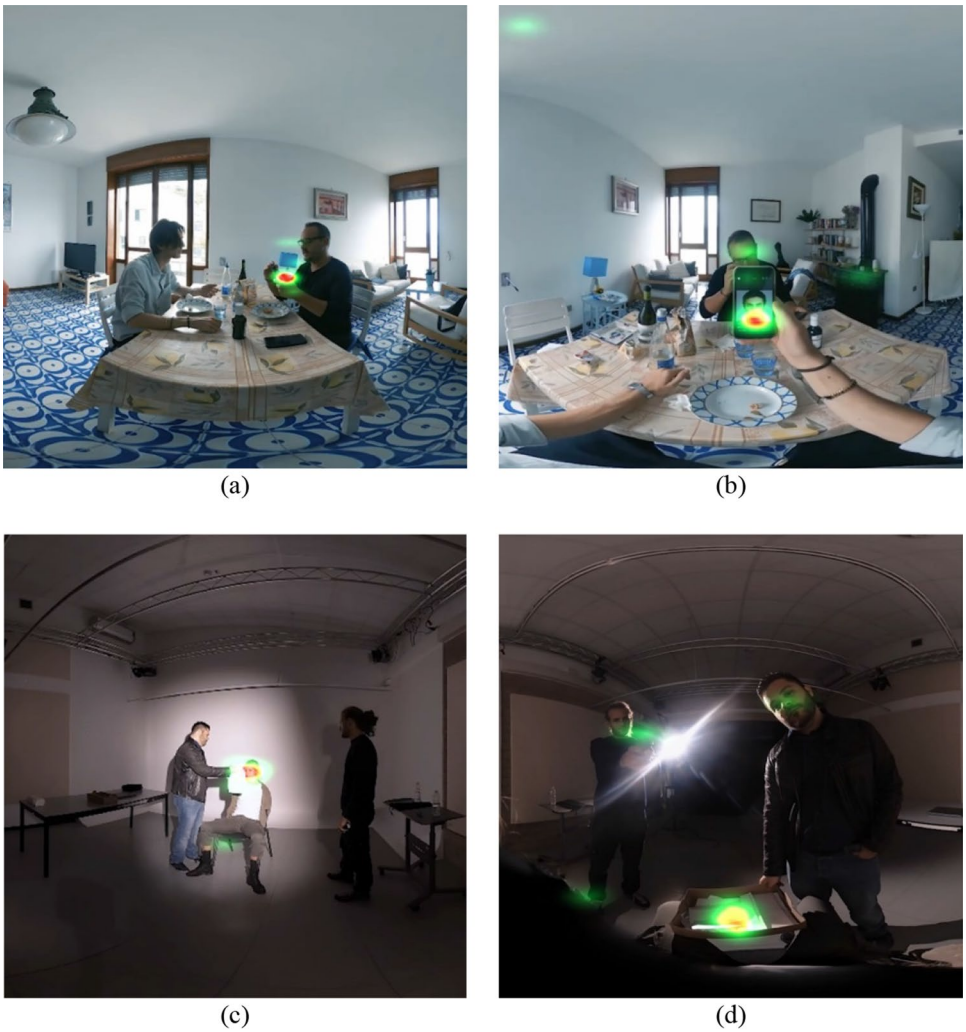

Fig. 2 Examples frames of the heatmaps: **a**, **b** "Persons you may know", and **c**, **d** "Oreste is still alive"

(Salvucci and Goldberg 2000). The algorithm considers as a fixation a set of gaze data whose distance is below a certain dispersion threshold with a duration higher than a minimum duration threshold. According to the recommendations given in (Blignaut 2009; Salvucci and Goldberg 2000), in this work the dispersion and the minimum duration thresholds were set to 1° and 100 ms, respectively.

To compute the metrics based on ROIs, a number of peculiar elements of the story were identified as fundamental for its understanding. In particular, for the first script the identified ROIs are the two characters (Luca and Fabio), and Luca's phone. In the second script, the ROIs were associated to the three characters (Vadio, Sena, and the second agent), to Vadio's phone, to the letter, and to the box containing several letters found at Sena's home). Figure 3 shows the bounding boxes of the identified ROIs in the two scripts.

4 Experimental evaluation

4.1 Participants

The designed user study was conducted by involving 32 participants (25 male and seven female) aged between 22 and 45 ($\mu = 25.06, \sigma = 3.90$) who were recruited among students and staff at the authors' university. Participants were all volunteers, and no rewards were given. Most of the participants had limited knowledge of VR systems as they never (62.50%) or only rarely (15.63%) experienced VR (everyday 15.63%, once a week 3.13%, once a month 3.13%). Likewise, 75.00% of the participants had never watched immersive movies and only 25.00% stated they had watched VR content sometimes. Conversely, the majority of participants said to watch traditional cinema content every day (84.38%) or at least once a week (12.50%) or a month (3.12%).

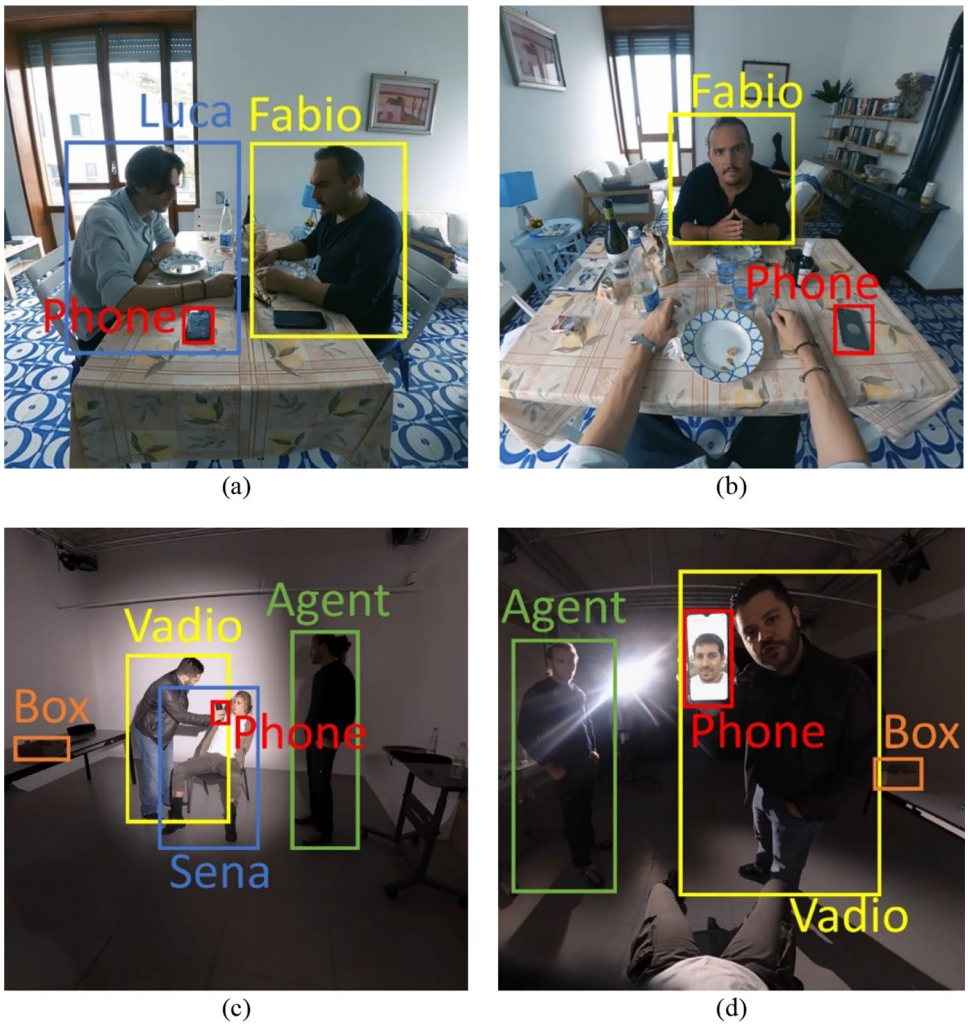

Fig. 3 Examples of the bounding boxes for each ROI in the scene: **a**, **b** "Persons you may know", and **c**, **d** "Oreste is still alive"

4.2 Procedure

All the study participants were requested to undergo the following steps. First, there was a welcoming and task explanation step, in which they were introduced to the experimental procedure. More specifically, the participants were informed about the two scenes they were going to watch, and the differences in terms of POV. In case the participants were not familiar with VR and the used headset, they were explained how to wear and regulate it in order to guarantee the most comfortable experience. Then, the eye-tracking calibration procedure was performed for each participant, checking that it was working as expected by means of a simple gaze-pointing game. After the calibration step, to avoid possible biases caused by vision or hearing issues, participants were asked to judge the image and sound quality by watching a demo video selected from the YouTube VR Channel. In case of issues, the participants were given time to adjust the equipment.

Then, the study followed a within-subjects approach. In particular, the participants were equally split in two groups, so that a group watched first one of the two versions, i.e., 1-PP or EP, of the script "Persons you may know" and then the alternative version, i.e., EP or 1-PP, of the second script "Oreste is still alive", whereas the other group did the opposite. Before starting the videos playback, the participants were asked to point their gaze at the center of the scene to guarantee similar starting conditions among them.

After watching each video, the participants were asked to answer the 12 statements of the NES. In addition, at the end of the experiment, they were requested to express their preferences on the experienced POVs, rate the quality of the produced videos, and leave additional comments, if any.

All the videos were watched via an HTC Vive Pro Eye headset. Audio content were delivered through the integrated headphones. Similarly, eye-tracking data were gathered through the embedded hardware. The participants were allowed to choose whether to watch the videos by either

Table 1 NES results for the 1-PP and EP versions the script "Persons you may know"

Metric	1-PP	EP	*p*	*d*
NU(–)	3.40 *(1.20)*	3.67 *(1.19)*	0.7530	0.2229
AF(–)	6.87 *(2.60)*	7.73 *(3.25)*	0.4363	0.2940
NP	**16.20** *(2.10)*	11.93 *(3.57)*	0.0003	–1.4568
EE	18.00 *(2.34)*	15.53 *(4.60)*	0.1607	–0.6758
NE	**23.93** *(5.60)*	16.07 *(9.01)*	0.0113	–1.0482

Bold font is used to highlight the significant differences and best results, whereas the italic font indicates the standard deviations

The (–) symbol indicates reverse coded dimensions

Table 2 NES results for the 1-PP and EP versions the script "Oreste is still alive"

Metric	1-PP	EP	*p*	*d*
NU(–)	**4.47** *(1.50)*	7.53 *(3.83)*	0.0218	1.0550
AF(–)	**4.87** *(2.06)*	7.53 *(3.28)*	0.0109	0.9727
NP	**15.60** *(4.59)*	10.93 *(3.37)*	0.0035	–1.1588
EE	**14.80** *(4.85)*	8.73 *(3.32)*	0.0017	–1.4609
NE	**21.07** *(8.76)*	4.60 *(10.80)*	0.0004	–1.6752

Bold font is used to highlight the significant differences and best results, whereas the italic font indicates the standard deviations

The (–) symbol indicates reverse coded dimensions

seating on a swivel chair or simply standing. It was decided not to fix a posture, since previous works—e.g., (Bellgardt et al. 2017; Zielasko and Riecke 2021)—indicate that the sense of immersion can be reduced if the users experience a mismatch between their height and the camera position. Considering the height of the camera in the two versions, the sitting posture was suggested for the 1-PP versions, whereas the standing posture was suggested for the EP versions. The participants were not forced to assume the suggested postures, since, e.g., the standing posture may be considered as not comfortable, resulting in a negative impact on the overall experience. The participants were allowed to terminate the experiment (withdraw) in case of motion sickness.

5 Results

5.1 Subjective results

The statistical significance of subjective results was studied by using Mann–Whitney (two tails) test. Moreover, to evaluate the effect size, Cohen's d measure was computed. According to (Sawilowsky 2009), the effect of the POV on the NE was considered as small ($|d| \geq 0.2$), medium ($|d| \geq 0.5$), large ($|d| \geq 0.8$), and very large ($|d| \geq 1.0$).

Table 1 presents the results of the NES for the first script, by reporting mean values (bold font is used to highlight the significant differences and best results), standard deviations (in italics), *p* values and effect sizes. Although, in this script, the potential of the 1-PP was not specifically exploited, the participants who watched this version experienced a greater level of NE than those who watched the EP version (23.93 vs 16.07, $p = 0.0113$, $d = -1.0482$). This result indicates that the 1-PP can be considered as a better POV, overall, for creating interest and involvement in 360° videos compared to the EP. Analyzing more in detail each dimension of the NE, it can be noticed that statistically significant differences were observed only for the NP, indicating that the participants who watched the 1-PP version felt more immersed (NP 16.20 vs 11.93, $p = 0.0003$, $d = -1.4568$) than those who watched the EP version. These results suggest that the NP could be the dimension that is more sensible to the effects of the POV. For the other three dimensions, no significant differences were observed between the two versions. Possible reasons could be the influence that the highly touching subject of the story has on these dimensions, as the strong emotional impact could have prevented the participants to concentrate on other aspects. This assumption seems to be confirmed by the high scores obtained by both the versions for the EE dimension.

The results obtained with the second script are reported in Table 2. Also for this script, the participants experienced a greater level of NE with the 1-PP version than with the EP version (NP 16.20 vs 11.93, $p = 0.0003$, $d = -1.4568$). Considering the four dimensions of the NE starting from the NU, it can be observed that the participants in the EP group showed more difficulties in understanding the plot and the characters of the story than the participants in the 1-PP group (7.53 vs 4.47, $p = 0.0218$, $d = 1.0550$). Similar findings were obtained for the AF dimension, since the participants who watched the EP version showed higher difficulties in terms of concentration on the main actions of the scene (4.87 vs 7.53, $p = 0.0109$, $d = 0.9727$). This outcome might be due to the fact that virtual interactions (e.g., actors speaking by facing the camera which is controlled by the participant's movements) and details (e.g., information shown on the phone's display) included in the 1-PP version represent additional stimuli for the participants' attention. Therefore, it could be hypothesized that a different placement of the camera in the EP, for instance one in which the resulting perspective allows the viewer to look at the phone display, may lead to different results even considering the same script. Anyhow, the effects of POV were even more evident considering the NP and EE dimensions. In fact, the participants who watched the scene through the eyes of the main character (Sena) showed a greater empathy with her and greater emotional involvement with the story than those who watched it from a detached POV (NP

Table 3 Objective results for the 1-PP and EP versions of the script "Persons you may know"

Metric	1-PP	EP	*p*	*d*
nFix [%]	0.49 *(0.16)*	**1.04** *(0.19)*	0.0001	3.0453
PercFixInside (Phone) [%]	5.78 *(4.20)*	3.36 *(1.67)*	0.1485	– 0.7587
PercFixInside (Fabio) [%]	**60.11** *(23.70)*	15.17 *(8.43)*	0.0001	– 2.5265
Exp. Fidelity (Phone)[%]	**24.68** *(18.16)*	12.83 *(20.72)*	0.0329	– 0.6079
Head path [px/sec]	212.62 *(95.41)*	249.33 *(90.61)*	0.2017	0.3945
Gaze path [px/sec]	2103.75 *(1024.56)*	**3296.65** *(887.96)*	0.0043	1.2443

Bold font is used to highlight the significant differences and best results, whereas the italic font indicates the standard deviations

The (–) symbol indicates reverse coded dimensions

15.60 vs 10.93, $p = 0.0035$, $d = -1.1588$, EE 14.80 vs 8.73, $p = 0.0017$, $d = -1.4609$). Interestingly, in this case, all the analyzed dimensions were in favor of the 1-PP, indicating how exploiting the potential of the 1-PP in the scripting and the shooting phases can actually have an impact of the resulting NE. It is unsure, however, how these metrics would have been affected by a script conceived to be more targeted to the EP, e.g., by having Vadio and the Agent wandering in the room space and moving closer and farther to the camera.

Interesting insights can also be obtained by comparing the results obtained for the same version (1-PP or EP) of the two scripts. Starting with statistically significant differences observed for the 1-PP version, it can be noticed that the second script generated more issues in the comprehension of the story with respect to the first script (NU 4.47 vs 3.40, $p = 0.0476$, $d = 0.7460$), whereas in the first script the participants reported more difficulties in focusing on the main action (AF 4.86 vs 6.87, $p = 0.0095$, $d = -0.9050$). The result regarding the NU dimension could be related to the fact that the story of the second script had a more complex plot and was set in a dystopian future, pretty far from reality. This motivation appears to be confirmed also when comparing the results obtained with the EP versions (NU 7.53 vs 3.67, $p = 0.0006$, $d = 1.2538$). The outcome related to the AF, instead, requires further analysis. The observed difference could be attributed to the richer environment (in terms of elements) of the second scene, which could have made the participants more prone to wander on it; however, this motivation was not confirmed by the comparison of the EP versions. From the analysis on the EP versions, besides the differences in terms of NU two other significant results can be observed. The first one regards the EE dimension, in which the first script obtained much higher scores (8.73 vs 15.33, $p = 0.0004$, $d = -1.5511$) since, as said, it probably created more emotional involvement with the character and the story, which is quite close to the reality and more dramatic. The second difference concerns the overall NE, which was much lower in the second script (4.60 vs 16.07, $p = 0.0058$, $d = -1.0611$): this result was due to the lower contribution made by the EE (as discussed above) and the NU (probably because of the lack of elements, in the second script, that can help the participants to understand a story which is distant from reality).

Finally, regarding participants' preferences, 65.63% of them preferred the 1-PP version, 25.00% the EP version, and 9.37% did not express any preference. These results are in line with the findings regarding the NE. The simulated interaction, although it was not an active process but only a passive role played by the participants in observing the characters, apparently made them perceive an improved sense of immersion and transportation. Furthermore, most of the participants who preferred the 1-PP version indicated that the EP version failed to satisfy their expectations of watching an immersive movie, since it was more similar to a traditional cinema experience. The participants who preferred the EP version supported their choice by stating that it was hard or even "weird" to impersonate a character with a totally different voice, body and personal background. Finally, no preferences were expressed by the participants who considered the POV to be dependent on the type of story or situation the movie actually represents.

5.2 Objective results

To study the statistical significance of the results, differences and effect sizes were analyzed through the Mann–Whitney (two tails) test and Cohen's d measure, respectively.

Table 3 reports the results regarding the head and eye-tracking data for the first script. Bold fonts indicate statistically significant differences. Focusing on significant differences, it can be observed that the participants showed a more exploratory behavior in the EP version than in the 1-PP version, as confirmed by the higher values of the Gaze path metric (3296.65 vs 2013.75, $p = 0.0043$, $d = 1.2443$). Interestingly, the values of the nFix metric are higher in the EP version than the 1-PP version (1.04 vs 0.49, $p = 0.0001$, $d = 3.0453$). This result is probably related to the limited size of the scene elements (e.g., the phone) when observed from the EP, which made the participants more prone to keep their gaze fixed on a specific region of the environment.

Table 4 Objective results for the 1-PP and EP versions of the script "Oreste is still alive"

Metric	1-PP	EP	*p*	*d*
nFix [%]	0.86 *(0.17)*	0.99 *(0.29)*	0.1985	0.5618
PercFixInside (Phone) [%]	**13.29** *(2.36)*	1.66 *(1.02)*	0.0001	– 6.3989
PercFixInside (Agent) [%]	**15.21** *(5.38)*	7.89 *(4.52)*	0.0009	– 1.4732
PercFixInside (Vadio) [%]	55.01 *(8.72)*	53.57 *(9.77)*	0.7400	– 0.1565
PercFixInside (Letter) [%]	**2.39** *(1.18)*	1.55 *(0.54)*	0.0344	– 0.9121
PercFixInside (Box) [%]	**3.57** *(1.14)*	2.11 *(1.06)*	0.0089	– 1.0641
Exp. Fidelity (Phone 1) [%]	**49.95** *(16.55)*	22.53 *(27.26)*	0.0251	– 1.2159
Exp. Fidelity (Phone 2) [%]	**38.66** *(11.93)*	8.23 *(1.14)*	0.0001	– 2.5005
Exp. Fidelity (Letter) [%]	44.83 *(9.39)*	38.94 *(14.90)*	0.1985	– 0.4726
Exp. Fidelity (Box) [%]	42.48 *(15.89)*	34.66 *(23.47)*	0.2290	– 0.3904
Head path [px/sec]	**216.11** *(86.19)*	138.84 *(56.46)*	0.0310	– 1.0606
Gaze path [px/sec]	3272.57 *(1028.96)*	**4974.84** *(1844.39)*	0.0279	1.1399

Bold font is used to highlight the significant differences and best results, whereas the italic font indicates the standard deviations

The (–) symbol indicates reverse coded dimensions

Conversely, from the 1-PP, the elements appeared as larger, hence the participants could explore the whole surface while keeping them in their field of view. This behavior could have increased the saccadic movements, leading to low nFix values. Regarding the PercFixInside metric, the 1-PP version showed significantly higher values than the EP version for the ROI labeled as Fabio (60.11 vs 15.17, $p = 0.0001$, $d < 0.05$). This result is probably due to the fact that Fabio is the character seated in front of the main character (Luca) and, hence, in the 1-PP version he directly spoke to the camera. In the EP version, the participants had the opportunity to look at the characters in the same field of view; for this reason, they followed the dialog by just switching the gaze between the two characters (thus reducing the number of fixations registered for Fabio). No statistically significant differences were observed for the head path, probably because there were not enough elements to stimulate the head reorientation. Regarding the Experiential fidelity metric, only the phone was identified as the object valuable for the story. Higher values were obtained for the participants who watched the 1-PP version than for participants who watched the EP version (24.68 vs 12.83, $p = 0.0329$, $d = -0.6079$).

Table 4 reports the objective results obtained with the second script. Focusing on significant results, it can be noticed that larger effects of the POV on the NE were observed, in terms of the PercFixInside metric, for the 1-PP version compared to the EP version on the ROIs labeled as phone (13.29 vs 1.66%, $p = 0.0001$, $d = -6.3989$), agent (15.21 vs 7.89%, $p = 0.0009$, $d = -1.4732$), letter (2.39 vs 1.55%, $p = 0.0344$, $d = -0.9121$), and box (3.57 vs 2.11%, $p = 0.0089$, $d = -1.0641$). These results can be explained by the fact that, with the 1-PP version, details about the story shown on the phone and the letter, as well as through the box content were more naturally visible; probably, this fact also make them more attractive and interesting for the participants who focused their attention onto them. The large difference noticed for the phone ROI was probably due also to the scale of the object as presented to the participants in the two versions. In the 1-PP version, the phone covers a larger portion of the field of view, thus forcing the participant to look at it when presented in front of the camera. This is also confirmed by the analysis of the Experiential fidelity metric, as only this ROI presents statistically significant differences both in the first (49.95 vs 22.53%, $p = 0.0251$, $d = -1.2159$) and the second time window (38.66 vs 8.23%, $p = 0.0001$, $d = -2.5005$).

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, opposite trends are observed, i.e., higher values for the 1-PP and EP version in the head and gaze path, respectively. On the one hand, these results suggest that the 1-PP stimulated the participants to change their field of view more than the EP. On the other hand, the participants covered a smaller distance with their gaze (Gaze path metric). This outcome might be due to the fact that, in the 1-PP version, the participants had to move their head in order to make certain elements of the story/surrounding environment visible (e.g., they had to keep their head down to watch the content of the box). For this reason, with the 1-PP, the participants were more prone to change their field of view (by redirecting their head) and then keep their gaze focused on the element considered as ROI. This assumption seems to be coherent with the result of the PercFixInside metric. The EP, in turn, allowed the participants to see a wider portion of the 360° environment, without the need to change the field of view; therefore, the participants were

more prone to keep their field of view stable and focus on the various elements of the story by redirecting their gaze. Finally, differently than with the first script, in this case no statistically significant differences were observed for what it concerns the nFix metric. This result could be related to the presence of scene elements (relevant for the understanding of the story) that requested the participants to focus on a surface detail (e.g., the countdown and the images shown on the display of the phone, the text in the letter, or the content of the box). In line with the observations above, these details were more evident in the 1-PP version than in the EP one. Thus, an increase of the nFix value was observed only for the 1-PP version, which could have made the differences no more statistically significant.

6 Conclusion and future work

This work investigated the impact of POVs in immersive videos, considering as the main focus the narrative engagement. The subjective results indicate that the viewers who watched the 1-PP versions showed higher values of narrative engagement, overall, and of narrative presence with both the scenes that were produced in this work. The objective data showed that the 1-PP provided better results in terms of gaze fixation on valuable elements of the story, i.e., that it was able to better drive the attention of the viewers on those elements which are regarded as important at a certain time of the story. Moreover, the narrative understanding, attentional focus and emotional engagement seemed to be more affected by the subject of the story and the way the scene was shot than by the POV. Finally, the analysis confirmed the possibility to drive the viewers' attention by making use of lighting and characters' gaze, as suggested in previous works.

The proposed study can be considered as an additional step towards the formation process of the immersive cinema language, as its findings could be useful, in particular, for supporting content creators in deciding which POV to select for their immersive movies, and in understanding how to leverage it to maximize the viewers' immersion in the movie environment as well as their connection with the story and its characters.

Despite the positive results obtained, the proposed study presents some limitations. First, it was decided to focus on a single research direction (i.e., the effects that POV can have on the NE) among the six possible alternatives proposed in (Gödde et al. 2018). This choice was made to avoid possible confounding factors, which could influence the analysis of the viewers' experience. Future works could perform more in-depth analyses which, based on single-factor studies like the one reported in this paper, may derive other findings while considering multiple aspects at a time.

Furthermore, in this study only dialog scenes were considered. Moreover, the scenes were mostly static. Also, the main characters used for the 1-PP versions kept their sitting posture for the whole duration of the scene. Further work should be devoted to extend the study by considering other types of scenes possibly encompassing a wider range of actions, to check whether the subjective and objective results as well as the viewers' preferences remain valid under different conditions or if some script-dependent factors could be more favorable to the EP rather than the 1-PP.

Another possible limitation regards the fact that the actions in both the scripts occurred in a limited region. The remaining of the environment did not have any particular visual and audio elements that could distract the viewers' attention, or maybe convey cues helpful to support a better understanding of the story. Thus, the influence of the POV in scenes containing elements/actions spanning the whole viewers' surrounding and/or with noisy background would deserve further investigation. It could be interesting to also consider scenes in which the main character—and not the viewer—moves in the environment. In this case, the mismatch between the motion of the character and the viewer could lead to cybersickness and impact the level of engagement. Another fascinating way to further extend the study might be to consider higher levels of CVR interactivity (Tong et al. 2021), and investigate the viewers' (users') behavior by sweeping that dimension till the level in which it is possible for them control not only the view but also the body of the characters (as if they were avatars), use their own voice, and maybe interact with elements in the environment, still without resorting to a fully interactive VR experience.

An aspect that may have impacted on results is the fact that the two videos of each script were not perfectly equal and synchronized. Using real actors made it impossible to have two scenes with exact timing, gestures, and moves. Two possible solutions to this problem could be producing short movies with Computer Graphics or shooting the scenes with two 360° cameras at the same time (one per each POV) as done in (Cao et al. 2019). For the latter option, issues should be solved regarding how to occlude the other camera not used for shooting the given version.

Finally, the study group was rather homogeneous. Future investigations should consider participants with a wider variety of backgrounds (including, e.g., VR content creators or VR gamers) and from different age ranges.

Acknowledgements This work was mainly developed at the VR@POLITO, the Virtual Reality lab of Politecnico di Torino.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

Data availability authors confirm that the data supporting the findings of this study are available within the article as supplementary materials.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article.

References

Bellgardt M et al (2017) Utilizing immersive virtual reality in everydaywork. s.l., IEEE, pp 1–4

Bender S (2019) Headset attentional synchrony: tracking the gaze of viewers watching narrative virtual reality. Media Pract Educ 20(3):277–296

Blignaut P (2009) Fixation identification: the optimum threshold for a dispersion algorithm. Atten Percept Psycho 71(4):881–895

Branigan E (1975) Formal permutations of the point-of-view shot. Screen 16(3):54–64

Busselle R, Bilandzic H (2009) Measuring narrative engagement. Media Psychol 12(4):321–347

Cao R et al. (2019) A preliminary exploration of montage transitions in cinematic virtual reality. s.l., IEEE, pp. 65–70

Carey K et al (2017) Toward measuring empathy in virtual reality. s.l., s.n., pp. 551–559

Carroll N (1993) Toward a theory of point-of-view editing: Communication, emotion, and the movies, pp 123–141

Chen M, McGlone MS, Bell RA (2015) Persuasive effects of linguistic agency assignments and point of view in narrative health messages about colon cancer. J Health Commun 20(8):977–988

Christopher B (2020) First-person narratives: examining narrative persuasion in virtual reality. s.l., s.n

Cummins RG (2009) The effects of subjective camera and fanship on viewers' experience of presence and perception of play in sports telecasts. J Appl Commun Res 37(4):374–396

Cummins RG, Keene JR, Nutting BH (2012) The impact of subjective camera in sports on arousal and enjoyment. Mass Commun Soc 15(1):74–97

Dahl TL, Storlykken O, Røssehaug BH (2021) Exploring perspective switching in immersive VR for learning first aid in lower secondary education. Springer, pp 301–316

Dahlstrom MF, Rosenthal S (2018) Third-person perception of science narratives: the case of climate change denial. Sci Commun 40(3):340–365

de Graaf A, Hoeken H, Sanders J, Beentjes JW (2012) Identification as a mechanism of narrative persuasion. Commun Res 39(6):802–823

Dooley K (2020) A question of proximity: exploring a new screen grammar for 360-degree cinematic virtual reality. Media Practice Educat 21(2):81–96

Dooley K (2021) Cinematic virtual reality: a critical study of 21st century approaches and practices. Springer

Emmerich K, Krekhov A, Cmentowski S, Krueger J (2021) Streaming VR games to the broad audience: a comparison of the first-person and third-person perspectives. s.l., s.n., pp. 1–14

Gödde M, Gabler F, Siegmund D, Braun A (2018) Cinematic narration in VR—rethinking film conventions for 360 degrees. Springer, pp 184–201

Gorisse G, Christmann O, Amato EA, Richir S (2017) First- and third-person perspectives in immersive virtual environments: presence and performance analysis of embodied users. Front Robot AI. https://doi.org/10.3389/frobt.2017.00033

Hoeken H, Kolthoff M, Sanders J (2016) Story perspective and character similarity as drivers of identification and narrative persuasion. Hum Commun Res 42(2):292–311

Hoppe M et al. (2022) There is no first-or third-person view in virtual reality: understanding the perspective continuum. s.l., s.n

Lipsey AF, Waterman AD, Wood EH, Balliet W (2020) Evaluation of first-person storytelling on changing health-related attitudes, knowledge, behaviors, and outcomes: a scoping review. Patient Educ Coun 103(10):1922–1934

Marañes C, Gutierrez D, Serrano A (2020) Exploring the impact of 360 movie cuts in users' attention. s.l., s.n., pp. 73–82

Moghadam KR, Ragan ED (2017): Towards understanding scene transition techniques in immersive 360 movies and cinematic experiences. s.l., s.n., pp. 375–376

Nan X, Dahlstrom MF, Richards A, Rangarajan S (2015) Influence of evidence type and narrative type on HPV risk perception and intention to obtain the HPV vaccine. Health Commun 30(3):301–308

Pillai JS, Verma M (2019) Grammar of VR Storytelling: analysis of Perceptual Cues in VR Cinema. Association for Computing Machinery, New York

Punde PA, Jadhav ME, Manza RR (2017). A study of eye tracking technology and its applications. s.l., s.n., pp. 86–90

Rothe S, Hußmann H (2018) Guiding the viewer in cinematic virtual reality by diegetic cues. s.l., s.n., pp. 101–117

Rothe S, Buschek D, Hußmann H (2019) Guidance in cinematic virtual reality-taxonomy, research status and challenges. Multimodal Technol Interact 3(1):19

Rothe S, Zhao L, Fahrenwalde A, Hußmann H (2020) How to reduce the effort: comfortable watching techniques for cinematic virtual reality. s.l., s.n., pp. 3–21

Ruscella JJ, Obeid MF (2021) A Taxonomy for immersive experience design. s.l., s.n., pp. 1–5

Salvucci DD, Goldberg JH (2000) Identifying fixations and saccades in eye-tracking protocols. s.l., s.n., pp. 71–78

Sawilowsky SS (2009) New effect size rules of thumb. J Mod Appl Stat Methods 8(2):26

Schmitz A et al. (2020) Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos. s.l., s.n., pp. 63–72

Serrano A et al (2017) Movie editing and cognitive event segmentation in virtual reality video. ACM Trans Graph 36(4):1–12

Sitzmann V et al (2018) Saliency in VR: how do people explore virtual environments? IEEE Trans Vis Comput Graph 24(4):1633–1642

Stebbins T, Ragan ED (2019) Redirecting view rotation in immersive movies with washout filters. s.l., s.n., 377–385

Tanja A et al. (2021) From FOMO to JOMO: examining the fear and joy of missing out and presence in a 360° video viewing experience. s.l., s.n

Tong L, Lindeman RW, Regenbrecht H (2021) Viewer's role and viewer interaction in cinematic virtual reality. Computers 10(5):66

van den Boom AA, Stupar-Rutenfrans S, Bastiaens OS, van Gisbergen MS (2015) Observe or participate: the effect of point-of-view on presence and enjoyment in 360 degree movies for head mounted displays. s.l., s.n

Wu H et al. (2021) Immersive virtual reality news: a study of user experience and media effects. 147: 102576

Xue T et al. (2021) RCEA-360VR: Real-time, Continuous Emotion Annotation in 360 VR Videos for Collecting Precise Viewport-dependent Ground Truth Labels. s.l., s.n., pp. 1–15

Zhou S (2017) Actor's and observer's perspective in narrative processing, s.l.: s.n

Zielasko D, Riecke BE (2021) To sit or sot to sit in VR: analyzing snfluences and (dis) advantages of posture and embodied interaction. Computers 10(6):73