



Evaluation of user response by using visual cues designed to direct the viewer's attention to the main scene in an immersive environment

Galo Ortega-Alvarez¹ · Carlos Matheus-Chacin¹ · Angel Garcia-Crespo¹ · Adrian Ruiz-Arroyo¹

Received: 12 May 2020 / Revised: 9 September 2020 / Accepted: 23 May 2022 /
Published online: 8 June 2022
© The Author(s) 2022

Abstract

Today the visualization of 360-degree videos has become a means to live immersive experiences.. However, an important challenge to overcome is how to guide the viewer's attention to the video's main scene, without interrupting the immersion experience and the narrative thread. To meet this challenge, we have developed a software prototype to assess three approaches: Arrows, Radar and Auto Focus. These are based on visual guidance cues used in first person shooter games such as: Radar-Sonar, Radar-Compass and Arrows. In the study a questionnaire was made to evaluate the comprehension of the narrative, the user's perspective with respect to the design of the visual cues and the usability of the system. In addition, data was collected on the movement of the user's head, in order to analyze the focus of attention. The study used statistical methods to perform the analysis, the results show that the participants who used some visual cue (any of these) showed significant improvements compared to the control group (without using visual cues) in finding the main scene. With respect to narrative compression, significant improvements were obtained in the user group that used Radar and Auto Focus compared to the control group.

Keywords 360-degree videos · Immersive video · HMD · First person games · CVR

✉ Angel Garcia-Crespo
angel.garcia@uc3m.es

Galo Ortega-Alvarez
galoeffren@hotmail.com

Carlos Matheus-Chacin
carlosalberto.matheus@uc3m.es

Adrian Ruiz-Arroyo
adruiza@inst.uc3m.es

¹ Institute for Technological Development and Promotion of Innovation , University Carlos III of Madrid, Leganes, Spain

1 Introduction

Multimedia content is growing and diversifying exponentially. The rapid evolution of technologies goes hand in hand with the universalization of their use [11], so that increasingly electronic devices are becoming an indispensable tool for all our daily activities. Multimedia content and especially traditional videos now have great potential to bring an enormous amount of information to the viewer. This ability of video to transmit information to the viewer can be extended if the video is provided with more mobility and flexibility [26]. These characteristics offer a closer connection with the spectator's senses, to the point of immersing them completely in a narrative, and thereby generating improved viewer satisfaction while watching a film.

At present there are few ways to improve the narrative in a video, since it is necessary to increase the viewer's presence sense. One of the low-cost and user-accessible alternatives is 360-degree video playback using HMD Head-Mounted Display devices. This way of reproducing videos in 360 degrees allows the spectator to enjoy a high degree of presence and immersion in a film's narrative. On the other hand, while, a 360-degree environment is a technologically feasible step in Cinematic Virtual Reality (CVR), the cost of appealing to new senses to convey more emotions to viewers is relatively high. In addition, there are still many problems that need to be resolved in order to obtain 360-degree cinema. One of the most relevant issues is the direction of the user's attention during the immersive video playback.

Filmmakers have always relied on four cinematographic techniques to guide public attention during the playback of a film [25]. But in highly immersive environments such as 360-degree videos, the rules have changed and a problem has arisen for this type of content, which is that the viewer does not understand or enjoy the 360-degree video narrative. When drastically changing the environment in which the viewer finds himself, his first reaction is to explore the environment in order to gain a sense of security [2]. This activity of exploring implies that the spectator must familiarize himself with the scene, look for the focus of action of the video and connect again to the narrative in each change of scenario. This involves a relatively long period of time that prevents the user from understanding the video narrative and clearly resulting in frustration and even stress. It is increasingly important to resolve the focus of attention and spatial positioning of the user within immersive videos. This is due to the fact that currently many sectors such as cinematography, education, journalism, marketing, tourism, heritage preservation, healthcare (patient rehabilitation), entertainment (video games), data engineering, among others, are interested in the high potential of 360-degree videos to transmit information.

1.1 Software development

We constructed a software that allows visualization of videos in 360 degrees (immersive) with focus orientation and spatial positioning of the spectator. This spatial positioning is done through visual cues designed based on first-person games and cinematographic narrative genre. These visual guidance cues indicate the place where the next action of the narration will be executed and allow the spectator to improve his sense of location within the immersive environment. We developed the software considering that the multimedia content has several main scenes and requires a direct visual cue that is less invasive for the narrative of the immersive content. The development of the software is described in more detail in section 5.2.2.

1.2 Scope

Our research focuses on directing the viewer's attention in highly immersive environments through visual cues. Using visual cues to enhance the narrative experience of 360-degree videos are framed within the study area called Cinematic Virtual Reality (CVR). The objective of this study is to analyze the user's behavior in an immersive environment, while presenting visual cues designed to guide the viewer's attention to the main scene. This analysis is done with data collected through questionnaires and head movement data collected by the developed software. This data is analyzed using statistical methods.

The results of the present study indicate that all visual cues correctly guide the user to the main scene, they also indicate that the Radar and Arrows cues are accepted by the participants to guide them through the immersive environment, however, no concrete result was obtained with respect to the Auto Focus cue. With respect to the two criteria studied (understanding of narrative and focus of attention), the Radar and Auto Focus cues improved the participants' narrative comprehension, as well as correctly guiding the viewer to the main scene. Considering the three criteria evaluated, the visual cue that obtained the best results was the Radar. This is described in more detail in the Discussion section.

The document consists of eight sections and is structured as follows. Section 2 reviews the previous literature related to the study presented. Section 3 details the designed visual cues. Section 4 presents a general description of the actions carried out throughout the project. Section 5 presents in detail the development of the experiment, including information about the immersive environment, the software, the selection of the participants, the procedure of the experiment and the data collection. Section 6 presents the results obtained using statistical methods. Section 7 details the discussions and, finally, section 8 presents the conclusions obtained from this study.

2 Related work

The massification and enormous acceptance of mobile devices has led to the creation of complementary products such as HMDs. This technology allows users to have immersive experiences such as: browsing virtual sites, viewing immersive multimedia content (immersive 360-degree videos), playing with first-person applications in a way different from the traditional ones, and above all offering new experiences to the user. The publication of immersive multimedia content in the computer cloud has grown rapidly. Video capture and playback devices are becoming more common and affordable for the general public [23]. For this reason, today many companies such as BBC29 in London highlight the importance of immersive videos in society to such an extent that they have a research department in this field. 360-degree videos allow the viewer the freedom to watch an entire environment while telling a story [7]. Such freedom also results in a reduced capacity for multimedia content producers since directing the viewer's attention to the focus of the action at a certain angle and time is very complicated. These authors thus carried out countless experiments finding that the combination of auditory and visual cues improve the narrative experience of multimedia content but do not guarantee attention in the focus of the action. For this reason, there are several other studies within this field aimed at improving the viewer's attention within immersive multimedia content.

Visual cues must be well designed in order to guide the attention of the spectator throughout the narration so that users feel the immersion and spatial presence of the film without losing the thread. It is also necessary to use universal visual guidance cues, in case they are not used, these should be explained to the user [29]. At the same time, possible solutions are studied to visualize videos in 360 degrees by means of mobile devices, focusing and reorienting the objectives of the video continuously. To address this issue, two focusing techniques have been developed in the CVR area:

- Autopilot (constantly refocusing the viewing angle). This technique is widely used in 360° sports videos [15].
- Visual guide (indicating the direction of the lens). These can be diegetic (intra- diegetic) or non-diegetic (extra-diegetic) cues. Where diegetic cues refer to cues that are part of the scene, for example, sounds of an instrument being played by a character; while non-diegetic cues are external to the scene, the characters in the film do not know about these cues.

Both approaches are currently used, both diegetic [28] and non-diegetic cues [21]. we used non-diegetic cues in our project, although some previous works indicate that this can break the user's immersion in the narrative, it is also known that some people are afraid of missing something if they are not properly guided [1]. And this feeling of missing out a part of the story can happen even using diegetic cues to guide them [28].

A study carried out in 2017 indicated that the autopilot surpassed the visual guides when it comes to multimedia content in 360 degrees where sports activities are developed in third person. However, visual cues were more accepted by viewers who prefer first-person tourist multimedia content, since one of the characteristics of such content is exploration of the environment [19]. On the other hand, there are immersive environments that are constructed based on spherical images where guiding the viewer's attention is complex. For this reason, the authors present a navigation interface that allows users to control the locomotion and speed of immersive environments based on photographs. This navigation interface shows users visual cues that allow them to indicate the direction in which they can navigate within the immersive environment [35].

Navigating immersive environments can provide an engaging experience for the viewer, but it remains a challenge to overcome the limitations of redirection of focus. Have been proposed three reorientation techniques (SCR: Subtle Continuous Reorientation, SDR: Subtle Discrete Reorientation, OCR: Overt Continuous Reorientation) where it was confirmed that participants were less likely to present presence ruptures when the spectator's focus of attention was reoriented. Especially with SRC continuous rotation techniques, rotating the environment while walking in a way that is imperceptible for the viewer. However, it must be considered that it can only be applied in narratives where the spectator experiences the sensation of walking [34].

To date most research on immersion in 360-degree videos has focused on the issue of improving user experience and in part on improving the spatial position of the viewer, which is key in this type of content. Within this multimedia content several of the studies focus on improving this experience and the known and not so exploited way is through visual aid to the viewer [29]. With this in mind, there is a field of technology with a great deal of experience in the spatial position of the user and narrative in immersive system: video games [36]. These entertainment systems have exploited a diversity of fields and a very important one is that of

video games, specifically “First Person Shooter” (FPS). These focus on the fact that the protagonist of the game is the user and his vision is totally immersive [22]; in view of this situation the producers of games have partially resolved the user orientation and focus of attention through various spatial positioning cues. One of the main challenges for FPS game designers is to compensate for the spectator’s loss of sense of orientation when the environment is highly immersive. One solution that has been used and has a high degree of acceptance is radar superimposed on the viewer’s field of vision, which provides spatial information on the spectator and the objects in the environment [10]. This is the reason why we have decided to use Radar as one of the visual cues to guide the viewer to the main scene, also evaluating whether the use of this cue improves the user’s understanding of the narrative of the story. The color, silhouettes and shapes are the basis for a good design of the data presented in the interfaces of the FPS [37]. In the same way, the color, size and movement are the three characteristics to be considered to improve the design of the interface and prevent the viewer from losing the immersion experience [12]. The colors play a fundamental role as they provide the atmosphere and mood for the viewer. Warm colors such as red, orange, yellow can create an atmosphere of anxiety, stress and emotion, signaling danger or that something is going to happen. On the other hand, cold colors such as blue, purple, green can create the effect of harmony and security [30].

Tools and techniques of visual follow-up have been examined to evaluate the behavior of the look and the movement of the head in 360-degree videos that thereby improve the immersive narrative [13, 20]. Similarly, the user experience can be improved by increasing the horizontal rotational gain (the ratio between the rotation velocity in the virtual (video) space to that of the physical) to view 360-degree videos [14]. We have considered this aspect to improve the user experience, reducing the fatigue or pain that can be caused by making many movements. Regarding spatial positioning in immersive video, visualization and signaling during the immersive experience is important in order to avoid user distraction [27]. On the other hand, the use of a widget (a computer application that provides additional information to improve the usability of computer systems) within the virtual reality for positional information and increasing awareness of events in its surroundings proved to be very useful for the user without causing any discomfort during the experience [32]. For users without experience in immersive environments, 3D arrows allow better positioning results to be obtained than did any other aids, while for experienced users, the Radar is the widget that obtain better acceptance for spatial positioning [8].

Spatial positioning and user orientation within a 360-degree video continues to be an interesting and important research topic for the entertainment industry. Implementing novel and effective solutions will set precedents for maturing this type of technology in favor of improving the user experience.

2.1 Narrative video environments

The cues presented in previous work allow users to be guided in exploratory environments such as tourism, which is not the same for environments where a narrative is told, such as films or shorts. The cues presented in these narrative environments tend to break the immersion of the film, as happens with subtitles in films. One way to address this problem is to introduce visual cues into the film at the editing stage, but the reality shows us that this solution is not adapted in the majority of videos found on the web. For this reason, there are research that propose that future works considering the design of visual cues based on colors and shapes that

prevent breaks in the immersion of 360-degree videos containing a narrative, such as films or shorts. It is also designing cues that help to improve the spatial position of the viewer, also seen as important since some film genres, such as science fiction in 360-degree video, can produce loss of orientation and dizziness [19].

3 Visual guidance cues

We based the independent variables considered for the creation of the software, on the visual guidance cues used in first person shooter games. The visual cues chosen are: The Arrow, the Radar and the camera reframe (Auto Focus), for the location of the main scene. Previous works have shown that visual cues like Arrow and Radar help guide people in virtual environments [8, 9, 24]. The objective of our project is to verify not only if the user's gaze is focused on the main scene, but also if the use of these visual cues help to improve the user's understanding of the narrative.

- Arrow. These are general cues that appear on the periphery of the viewer's field of vision indicating where to look or turn to locate the main scene of the video in 360 degrees.
- Radar. Similar to the operation of a sonar, this shows the referential position of the spectator with respect to the main scene. We also designed a Radar-Compass which shows where the main scene of the video is.
- Auto Focus. This is composed of two cues (Arrow and Radar-Compass) supported by the camera re-frame in the main scene. Visual cues are only there to let the user know about the re-focus movement towards the main scene.

The design of the visual cues has focused on making the viewers feel safe when making movements in the immersive environment, without causing distraction from the main scene. The design of each of the visual cues is described in more detail in section 5.2.2 where the software and hardware used in the experiment is presented.

4 Methodology

For the present study, we considered a mixed approach (qualitative-quantitative) with an exploratory-correlational scope necessary, since it focuses on the analysis of the behavior of the spectators experiencing visual stimulation in an immersive video. This study evaluates the problem of guiding the viewer to the main scene considering not only the positioning of his head (visual focus), but also the understanding of the narrative. All this through the stimulation by visual cues. It also has an experimental design, as it considers the manipulation of three independent variables (Arrow, Radar and Auto Focus) in order to measure the effect on dependent variables (understanding of narrative and focus of attention).

We determined that the degree of manipulation of the independent variables is presence-absence, which implies having a control group and an experimental group for each independent variable, in order to evaluate the effect of visual cues on viewers. In other literature, this type of manipulation of independent variables (presence-absence) is called between group [18]. For the measurement of the dependent variable "focus of attention" the software developed is used as a data collection instrument, since it allows the interaction of the study

subjects to be recorded, while for the measurement of the dependent variable “understanding of narrative” a questionnaire is used. The collected data are processed with statistical methods (ANOVA), for which the SPSS tool has been used.

In addition to the study of these two dependent variables, a study of the usability of the system was carried out. The information is collected through some questions in the questionnaire that the participants must answer, and the data is processed through the SUS scale. The satisfaction of the participants is also analyzed with respect to the visual cues used. The data is collected through some questions of the questionnaire and its analysis is done through a box plot of the 5-scale Likert used. Table 1 shows the features and limitations of the statistical methods used to analyze the data collected throughout the experiment.

Our research focused on using 360-degree videos that contain a plot, such as a film, a short or a video that at least has a very simple script, exempting first-person sports videos that generally do not have a script. At the same time, the devices selected to implement the software were mobile devices in conjunction with HMD devices. We selected these for their features of low-cost immersion and for their high degree of mobility necessary to evaluate the greatest number of participants.

Table 1 Features and limitations of the statistical methods used in the study

| Methods used | Features | Limitations |
|-------------------|---|---|
| ANOVA | <ul style="list-style-type: none"> ANOVA can compare the means across three or more groups, unlike the Independent Samples t Test, which should be used to compare the means for two groups. | <ul style="list-style-type: none"> ANOVA indicates if there is a significant difference between the means of any of the groups, however, it does not tell us which pair of means is different. To know which pair of means is significantly different requires additional analytical processes, such as a Post-Hoc analysis. ANOVA assumes that the data in the groups are normally distributed (or approximate). ANOVA assumes that the groups have the same, or very similar, standard deviations. The samples must be independent. |
| Questionnaire SUS | <ul style="list-style-type: none"> SUS scale is a widely used standard for measuring the usability of a system. SUS is a structured and validated questionnaire for performing usability measurements. Considering that SUS is neutral with respect to the technology used, it can be used to compare the present study with future work in the same area, even if the technology used is changed (it is not necessary to change the structure of SUS questionnaires) [6]. | <ul style="list-style-type: none"> The SUS is a subjective measure of perceived usability. It is a method that analyzes the usability of a system only from the user’s perspective. However, this is no longer a limitation if we consider that usability is divided into efficiency, effectiveness and satisfaction [6]. The first two are verified by checking if users complete the requested tasks and observing how they do it, while satisfaction is analyzed by means of the SUS scale. |
| 5-scale Likert | <ul style="list-style-type: none"> Participants easily understand this type of scale in a questionnaire, as it is widely used in different areas. It allows us to quantify qualitative information, such as participants’ perception of the visual cues used. We can easily analyze the data obtained by calculating the median and frequency on an ordinal scale. | <ul style="list-style-type: none"> There are many scientific controversies about the use of this method The nature of the data is ordinal, and the results obtained from its analysis are also ordinal. This type of questionnaire can force users to answer in a limited number of categories, which are sequential, but not necessarily linear. |

The aspects related to the experiment carried out are described in section 5, however, Fig. 1 shows a diagram of the methodology that has been followed throughout this project.

5 Experiment

5.1 Definition of variables and parameters

The present research uses three types of independent variables and two dependent variables. The independent variables correspond to each visual cue designed. The proposed cues are:

- Arrow.
- Radar.
- Auto Focus.

The dependent variables used to measure the causal effect of manipulating the independent variables are:

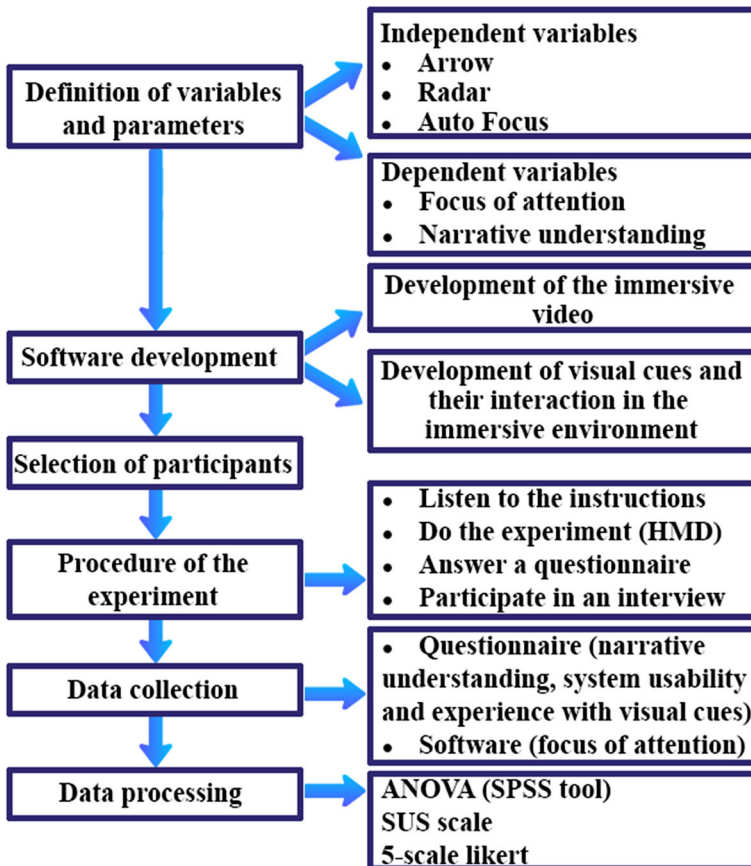


Fig. 1 Project methodology

- Attention of the spectator in scene focus.
- Narrative understanding.

5.2 Experiment development

The experiment consists of two essential components. The first component is the production of a 360-degree video that provides the necessary immersive space to perform the tests. The second component of the experimentation is the development of a software necessary to obtain the values of the dependent variables required for the current study.

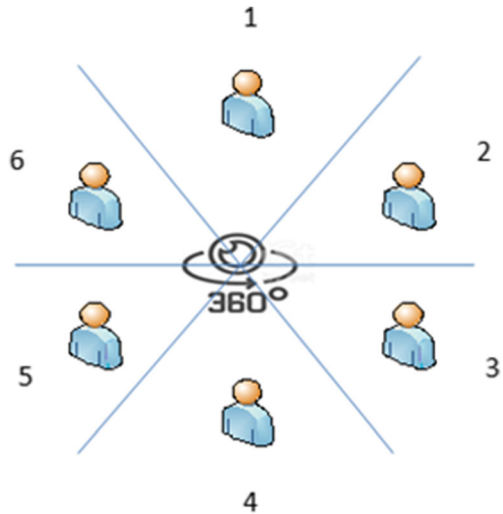
5.2.1 Immersive environment or video in 360 degrees

The immersive multimedia contents using HMD as an interaction device enable the user to look at 360 degrees horizontally and 180 degrees vertically contemplating a rotational displacement on the part of the spectator. The field of vision (FOV) of the human eye is 200 degrees [3]. For a person without peripheral visual training who wears HMD it is around 120 degrees horizontally and 90 degrees vertically, depending on the quality of the device. The main characteristic of an immersive environment is the freedom it gives the user to position the view at any angle, being necessary to build an environment that allows this degree of freedom considering the vision parameters mentioned above. The proposed environment is divided into 6 segments of 60 degrees horizontally, as this is the angle of vision of only one eye to pay effective attention. The radial rotational speed that a person can tolerate in this immersive environment has also been considered. Where the rotational speed is 30 degrees/second and the maximum displacement is 60 degrees without causing dizziness [14]. The device used for filming the 360-degree video was the Samsung Gear 360 VR camera.

The immersive video used in the experimentation must contain a narrative that enables the spectator's understanding to be evaluated, and thus verify the effectiveness of the visual cues. We used a short story, and a questionnaire previously designed and validated by Briceño, Velásquez, & Peinado [4] to evaluate the participant's narrative understanding. The name of the story is "El Lobo Feroz" (The Big Bad Wolf) and it has a duration of 4 min. The story of the video in 360 degrees is narrated through six people located every 60 degrees with respect to the frontal gaze of the spectator, as shown in Fig. 2. Each person narrates a part of the story in a previously determined time and sequence. The narration is supported by graphics to ensure that the viewer looks at the narrator as he or she tells the story. We note that the viewer will be evaluated in terms of understanding the plot and describing the graphics presented. This ensures that the participant has indeed rotated his position to focus attention on the main scene.

We designed the narrative sequence so that the software can evaluate the spectator's ability to turn and fix his vision. This sequence is distributed in the following way 1,6,3,4,5,4,2,3,2,6,1,5 allowing to evaluate which turns can be made during the video to understand the plot. Figure 3 shows the 360-degree spatial distribution of the narrators around a round table seen from a flat perspective. To guarantee the internal validity of the experiment, it is necessary to be isolated from external noise during the execution of the tests. HMDs and noise-canceling wireless headphones can generate a high level of immersion. This ensures that the viewer is isolated from external stimuli that may introduce error during experimentation.

Fig. 2 Spatial arrangement of people for shooting video in 360



5.2.2 Software and hardware

For the realization of the system, the software must consider the programming of a video player in 360 degrees that can support superposition of images and allow recording the visual position of the user based on the central point of the HMD. This allows the values of dependent variables and the data needed to perform statistical calculations to be recorded to then analyze the results. In detail, the software records the movement of the head of the spectator during the experimentation. Data is stored in the device every tenth of a second, saving the viewer's position, elapsed time and position of the main scene.

We designed the video player in Unity 5.4 (Multiplatform videogame engine that allows the creation and representation of immersive environments), while we developed the programming scripts for the superimposition of images in Visual Studio 2015 with the C# language. We designed the signage in Photoshop 5.5 and used Android Studio 2.3.3 to generate the



Fig. 3 360-degree video seen from a flat perspective

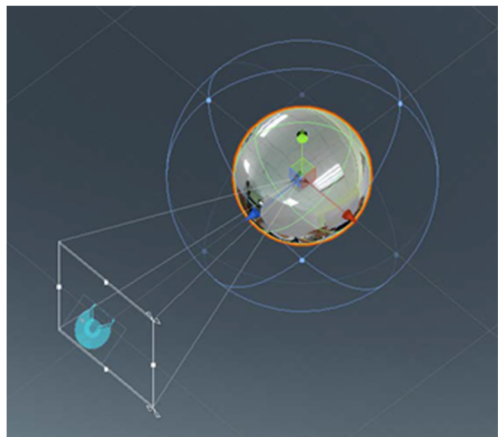
executable file in the Android operating system KitKat version. Additionally, we used the Java Development Kit version 8 update 131. The instruments to run the application are the mobile device Xiaomi MI4 with “1920 × 1080” screen resolution, a Snapdragon 801 Quad Core processor and 2 GB of RAM. In addition, we used a HMD device of VR BOX brand version 2.0.

As shown in Fig. 4, we can see the spatial distribution of the components used by the system software. The visual cue “Radar” can be observed in light blue, perched on a plane superimposed on the viewer’s vision. This superimposed plane has a “Follow” behavior, which follows the user’s vision by showing information about his position, with respect to the position of the main scene of the immersive video. At the same time, the video player can be observed in 360 degrees contained in a sphere that emulates the immersive environment for the viewer. Finally, in the center of the sphere a cube can be seen, which represents the viewer’s vision within the immersive environment.

Regarding the design of the visual cues, aspects such as color, position, among others, were considered. As mentioned earlier the visual cues chosen are: The Arrow (Fig. 5), the Radar (Fig. 6) and the camera reframe (Fig. 7), for the location of the main scene.

We consider several aspects for the design of the Arrow, such as the spatial positioning within the angle of vision of the viewer, the color of the symbol, the size and the appropriate animation to attract the attention of the viewer (the Arrow have an animation that moves to the center of the screen, in order to make it less intrusive to the viewer). For design of symbols for increased information in cases where the user’s attention is critical such as driving, recommend using the periphery of the view to inform the user about an action to be performed. The effective range for placing visual cues is 60 degrees measured from the center [17]. A characteristic of the Arrows is that they are global symbols of direction. They are recognized by the periphery of the user’s eyesight. We configured them with a light blue color and a brightness of 60% that denotes safety as shown in Fig. 8. As for the position of the visual cue, it is recommended placing it at the bottom, which in highly immersive environments denotes the direction of the feet providing greater security in the spatial positioning of the viewer [30]. We have considered necessary to implement a horizontal bar that more clearly denotes the direction of the ground since environments that contain narratives where the spectator does not see the ground can cause dizziness. The animation used is the movement of the Arrow in the

Fig. 4 Software system seen from outside the immersive environment



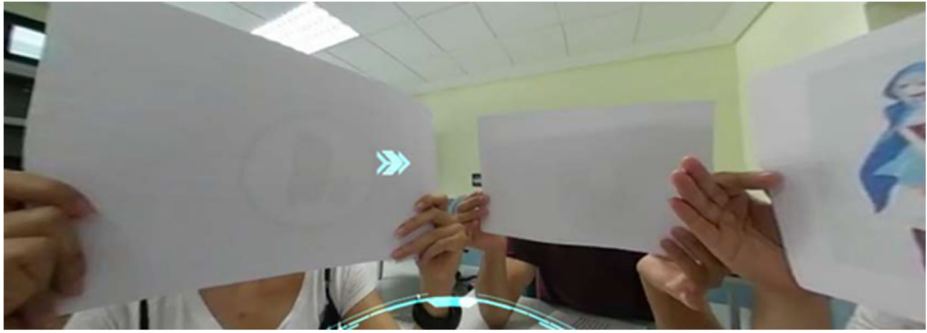


Fig. 5 Arrow cue in immersive video

direction in which the spectator's attention should be fixed. The representation of the Arrow cue on the ready-to-use mobile device is shown in Fig. 8.

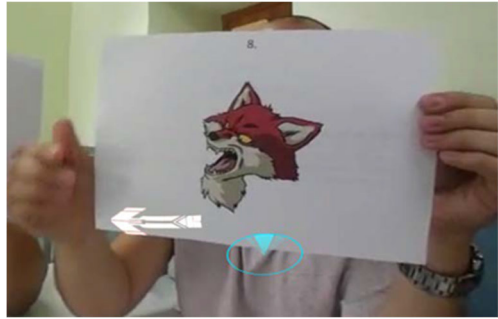
The Radar is one of the most effective spatial positioning cues that does not cause distraction [32]. It should be noted that Radar is not a generally known symbol, therefore, the user needs to be instructed or trained on its use before performing the experiment. The Radar design is widely used in first-person games to determine an object's location relative to the user's location. In this case, we use the Radar to identify the quadrant where the next action will be executed, telling the viewer where to turn to locate the main action. In order for the viewer to be able to locate the position of the next scene with the periphery of the view, we designed a light blue (50% brightness) pincer-shaped Radar, with the location of the scene in bright red. The Radar location is at the bottom with a 60% tilt angle to the front of the user to simulate depth descending parallel to the ground. The representation of the Radar cue on the ready-to-use mobile device is shown in Fig. 9.

For the implementation of the Auto Focus cue, we considered several design aspects as is the conformation of a new immersion space that is limited by the viewer's radius of vision. Since the environment is self-framing, it prevents the viewer from performing a body movement to focus on the main scene. In this case, we have designed a software that allows the automatic re-framing of the scene with smooth turns and scene translation. The operation is such that if the next scene is within the range of the periphery of the eye in a horizontal direction, there is a smooth movement of environment towards the center. If the next scene is behind the viewer, the scene is immediately superimposed as if it were a new scene. We

Fig. 6 Radar cue in immersive video



Fig. 7 Auto-Focus cue with Arrows and Radar-Compass



accompanied this scheme by a new Radar design where it constantly indicates the center of the main scene with respect to the visual position of the viewer. This information allows the user to constantly maintain his sense of location. This symbol has light blue-green colors with a high degree of brightness and a low position with respect to the viewer's viewing radius. Representation of this cue is shown in Fig. 10.

All symbols designed for experimentation and the 360-degree video player are evaluated by the participants using a SUS (System Usability Scale) questionnaire. This allows the software quality to be assessed, providing valuable information for feedback.

5.2.3 Sample selection

The criteria considered to select the participants for the experiments were that they had completed secondary education and that they did not have any type of visual or auditory disability, preferably that they do not get dizzy easily in immersive environments. These characteristics are assessed by means of a questionnaire prior to the experimentation. Figure 11 shows some of the participants in the experiment.

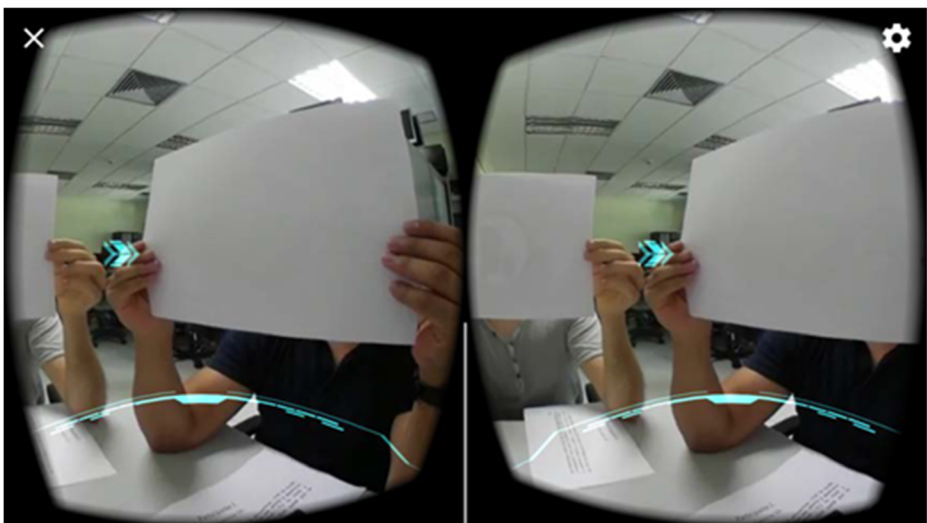


Fig. 8 Visual cue “Arrow” operating on mobile device



Fig. 9 Visual cue “Radar” operating on mobile device

We recruited fifty-two people for the experiment: 29 men and 23 women with an average age of 28.87 and a standard deviation of 6.64. The 52 participants were divided into 4 groups, one that represents the control group and 3 groups that evaluate the symbols to direct attention. Accordingly, each group is made up of 13 people chosen randomly at the beginning of the experiment.

5.2.4 Procedure

Figure 12 shows the stages in which the experiment is divided.

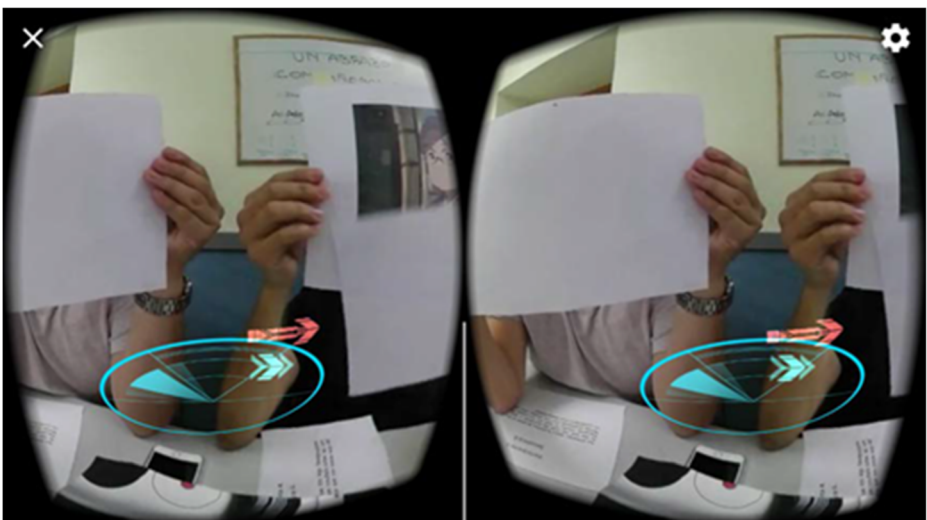


Fig. 10 Visual cue “Auto Focus” running on mobile device



Fig. 11 Participants performing the experiment

5.2.5 Data collection

We used the software developed (described in section 5.2.2) to collect navigation data (Focus of attention) and the questionnaire and the interview as measurement instruments for the dependent variable “Narrative understanding”, these instruments consist of open and closed questions, previously validated to obtain the information. The questionnaire has 29 items. This design consists of 6 general questions, 8 questions to measure the level of understanding of the narrative, 5 questions to evaluate the understanding of the visual guidance cues (likert scale) and 10 questions to evaluate the usability of the software (System Usability Scale).

It should be noted that to evaluate understanding of the narrative, we used a questionnaire previously validated by Briceño, Velásquez, & Peinado [4], who base their study on the reading of a story in order to measure the level of attention of people with a minimum educational level (“Obligatory Secondary Education”). At the same time, to evaluate the usability of the software developed for the present research, we used the survey validated by Brooke [5], which is widely accepted for carrying out global evaluations of computer devices with a high level of interaction with users. Furthermore, the design of the survey responses to understanding the visual guidance cues, as well as the design of the survey responses to assess the usability of the software were based on the Likert scale (as we mentioned previously). Likewise, to assess the narrative’s understanding, we used a single-selection response, which allows a better coding of the data for later analysis. In addition, this questionnaire contains a

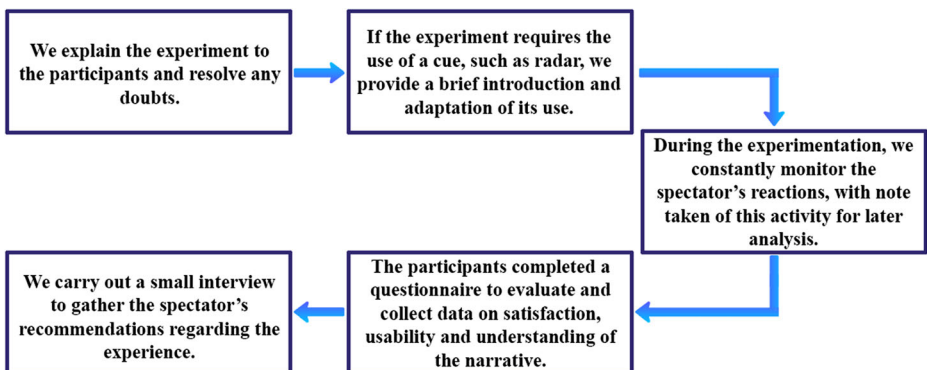


Fig. 12 Procedure of the experiment

section where the participant's consent is requested to carry out the experiment. This is necessary as these types of experiments can cause dizziness.

6 Results

We analyzed the results obtained from the experimentation in four stages. The first stage represents the results obtained about the previous knowledge that the participants have, and the comfort that they had when carrying out the experiment. The second stage corresponds to the evaluation results regarding the degree of narrative comprehension. The third corresponds to the evaluation of the design of the software systems presented. Finally, in the fourth stage, we validated the usability of the developed software as a tool to visualize videos in 360 degrees.

6.1 The first stage: General study of the participants

Prior to the experiment, we checked whether the participants had had experience with HMDs, as well as whether they were accustomed to the directional cues in less immersive environments (first-person games), in order to know whether they required training beforehand. We also asked the group of participants who agreed to learn about first-person games if they were familiar with the operation of the Radar cue. We determined that 66.7% of the participants knew this cue. All this is shown in Fig. 13.

One of the most relevant data regarding the evaluation carried out on the participants was whether they felt any discomfort during the experimentation. The results show that 25.6% did feel some discomfort during the experiment. This value is within an expected range, considering the percentage of participants who had no experience with HMDs and first-person games.

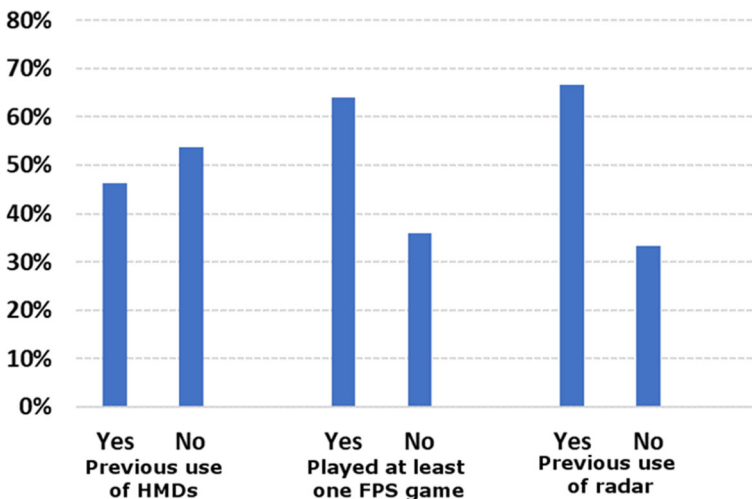


Fig. 13 General study of the participants

Table 2 Normality tests

| Qualification | Group | Kolmogorov-Smimov ^a | | | Shapiro-Wilk | | |
|---------------|---------|--------------------------------|----|--------|--------------|----|-------|
| | | Statistics | gl | Sig. | Statistics | gl | Sig. |
| | Control | 0.207 | 13 | 0.132 | 0.903 | 13 | 0.147 |
| | Arrow | 0.202 | 13 | 0.148 | 0.914 | 13 | 0.207 |
| | Radar | 0.191 | 13 | 0.200* | 0.938 | 13 | 0.430 |
| | Focus | 0.213 | 13 | 0.111 | 0.915 | 13 | 0.216 |

* This is a lower limit of true significance

^a Correction of meaning of Lilliefors

6.2 The second stage: Experimental results

6.2.1 Dependent variable: Understanding of narrative

To analyze the evaluation results for “narrative comprehension” obtained from the questionnaire applied to the participants, it is mentioned that for 13 samples of each group that is available and for the type of manipulation of the Between-Group independent variables the data must be analysed by ANOVA (Analysis of Variance - Statistical technique that shows whether the measures of two groups are similar or different) [18]. For this, the score of the questionnaire has been made based on 8, where each correct answer adds one point and the incorrect answers will not score.

We have decided that the study should be Between-Groups, since one of the dependent variables to study is the understanding of the narrative. Therefore, it would not make sense for users to repeat the observed narrative content. In order to be able to use ANOVA it is necessary to know whether the results comply with the test of normality, independence of observations, homogeneity of variances and equivalence of groups. Using SPSS, we obtained the result shown in Table 2.

The data in Table 2 meet the normality assumption since the “Kolmogorov-Sminov Statistics” data are in the range with a significance value $p > 0.05$.

Similarly, Table 3 shows the confirmation of homogeneity of Levene variances, as the value is 1.828 with $p > 0.05$.

Table 3 Variance homogeneity test

| | | Levene's Statistics | gl1 | gl2 | Sig |
|---------------|----------------------|---------------------|-----|-----|-------|
| Qualification | Based on the average | 1828 | 3 | 48 | 0.155 |

Table 4 ANOVA result

| Qualification | Sum of squares | gl | Quadratic average | F | Sig |
|----------------|----------------|----|-------------------|-------|-------|
| Between groups | 42.462 | 3 | 14.154 | 5.399 | 0.003 |
| Within groups | 125.846 | 48 | 2.622 | | |
| Total | 168.308 | 51 | | | |

Table 5 Post-Hoc Testing comparisons between groups (Rating)

| | Group (I) | Group (J) | measurement diff. (I-J) | Desv. Error | Sig | 95% confidence interval | |
|-----------|-----------|-----------|-------------------------|-------------|-------|-------------------------|-------------|
| | | | | | | Lower limit | Upper limit |
| HSD Tukey | Control | Arrow | -0.769 | 0.635 | 0.623 | -2.46 | 0.92 |
| | | Radar | -1.846* | 0.635 | 0.027 | -3.54 | -0.16 |
| | | Focus | -2.308* | 0.635 | 0.004 | -4.00 | -0.62 |

By carrying out the study considering the Between-Group modality, it is possible to guarantee compliance with the independence of observations in the experimentation and, therefore, in the data obtained. In addition, the equivalence of groups is fulfilled with $N = 13$ participants per group and therefore we can apply ANOVA.

The results obtained after applying unifactorial ANOVA are shown in Table 4, where the significance is 0.003, well below $p < 0.05$. This means that there are significant differences between the control group, the Arrow group, the Radar group and the Auto Focus group.

In order to know in which groups there are significant differences, we programmed the SPSS with the “Post-Hoc Tests” option. The results are shown in Table 5.

There are significant differences between the control group and the Radar and Focus groups. Both groups used Radar as a focus cue. Therefore:

- The results shown in Table 5 indicate that there is no significant difference ($p = 0.623$) between using the visual cue Arrow and not using any visual cue to guide the user in the immersive video, all this when the narrative understanding is being evaluated.
- The results indicated that there is a significant difference ($p = 0.027$ being $p < 0.05$) between the data of the control group and the group exposed to the Radar cue, when the user’s narrative understanding is evaluated. This confirms that the visual cue “Radar” improves the understanding of the narrative in a 360-degree video.
- On the other hand, with respect to the same dependent variable “narrative understanding”, the results indicate a significance of $p = 0.004$ being $p < 0.05$, when the data of the control groups and the Auto Focus group are compared, which shows that there is a significant difference between both groups. We can confirm that placing the main scene of the narrative automatically, as well as using Radar, increases the understanding of the narrative in 360-degree videos.

Table 6 Normality tests

| | Group | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
|---------------|---------|---------------------------------|----|--------|--------------|----|-------|
| | | Statistics | gl | Sig. | Statistics | gl | Sig |
| Qualification | Control | 0.149 | 13 | 0.200* | 0.948 | 13 | 0.568 |
| | Arrow | 0.181 | 13 | 0.200* | 0.925 | 13 | 0.289 |
| | Radar | 0.177 | 13 | 0.200* | 0.941 | 13 | 0.476 |
| | Focus | 0.180 | 13 | 0.200* | 0.928 | 13 | 0.319 |

* This is a lower limit of true significance

^a Correction of meaning of Lilliefors

Table 7 Variance homogeneity test

| | | Levene's Statistics | gl1 | gl2 | Sig. |
|---------------|----------------------|---------------------|-----|-----|-------|
| Qualification | Based on the average | 3.136 | 3 | 48 | 0.034 |

6.2.2 Dependent variable: Focus of attention

We based this validation on the data obtained by the software systems, which stored the interaction of people during the 360-degree video. Using the same data analysis procedure as above, we obtained the following result:

The data shown in Table 6 fulfill the assumption of normality since the data obtained in the section “Statistics of Koimogorov-Sminov” are in the range of 0.149 and 0.180 with a significance of 0.2 being $p > 0.05$.

In the same way, in the data shown in Table 7 we observe the confirmation of Levene variance, since the statistical value is 3.136 being $p > 0.05$.

As mentioned above, we considered the Between-Group modality. This means that each person was subjected to only one experiment, thus ensuring the independence of observations and hence the independence of the data. On the other hand, we note that the equivalence of groups meets $N = 13$ participants per group. Thus, it can be said that it fulfills all the requirements for applying ANOVA.

The results obtained when applying unifactorial ANOVA can be seen in Table 8. These results show that the significance is 0 complying with $p < 0.05$.

In order to determine in which groups there are significant differences, we configured the SPSS with the “Post-Hoc Tests” option. The results obtained are shown in the following Table 9.

As can be seen from the results shown, there are significant differences between the control group and the other groups. It should be noted that in the Radar and Focus groups, we used the Radar symbol to guide the viewer. Therefore:

- With respect to the dependent variable “focus of attention”, the comparative results show a significant difference ($p = 0$ being $p < 0.05$) between the data of the control groups and the group exposed to the Radar cue. With this we can confirm that the overlap of the visual cue Radar positively increases the direction of attention of the viewer in the main scene of a video by 360 degrees.
- On the other hand, the comparative results show a significant difference ($p = 0$ being $p < 0.05$) between the data of the control groups and the group exposed to Auto Focus. Confirming that placing the main scene of the narrative automatically while using Radar positively increases the viewer's attention in 360-degree videos.

Table 8 Result of ANOVA

| Qualification | Sum of squares | gl | Quadratic average | F | Sig. |
|----------------|----------------|----|-------------------|--------|------|
| Between groups | 513,373,781.3 | 3 | 171,124,593.8 | 76.744 | 0.00 |
| Within groups | 107,030,565.4 | 48 | 2,229,803.446 | | |
| Total | 620,404,346.7 | 51 | | | |

Table 9 Post-Hoc comparisons between groups (rating)

| | Group (I) | Group (J) | measurement diff. (I-J) | Desv. Error | Sig. | 95% confidence interval | |
|-----------|-----------|-----------|-------------------------|-------------|-------|-------------------------|-------------|
| | | | | | | Lower limit | Upper limit |
| HSD Tukey | Control | Arrow | -6119.538* | 585.702 | 0,000 | -7678.31 | -4560.77 |
| | | Radar | -6233.308* | 585.702 | 0,000 | -7792.08 | -4674.54 |
| | | Focus | -8446.077* | 585.702 | 0,000 | -10,004.85 | -6887.31 |
| | Arrow | Control | 6119.538* | 585.702 | 0,000 | 4560.77 | 7678.31 |
| | | Radar | -113.769 | 585.702 | 0,997 | -1672.54 | 1445.00 |
| | | Focus | -2326.538* | 585.702 | 0,001 | -3885.31 | -767.77 |

For this analysis of results the independent variable of the Arrow has not been taken into consideration, since both the Radar and the Auto Focus are the cues used in first person shooter games. We used Arrow as a comparison cue, which is presented as a solution in many previous studies. We will analyze the results of the study of the Arrow cue in the discussion of results section.

6.3 The third stage: Design of the system

Part of the research was aimed at gathering the opinions of the spectators regarding design of the systems presented. Participants completed a 5-scale Likert questionnaire (Fig. 14).

Figure 14 shows that the participants prefer the design and position of the Arrow (Median = 5, Min = 3) and Radar (Median = 5, Min = 3) cues over the Auto Focus cue (Median = 4, Min = 2), where both the Arrow and the Radar obtained very good reception from the participants. On the other hand, participants generally indicated that they were able to understand visual cues with peripheral vision, they indicated that both the Arrow cue (Median = 5) and the Radar cue (Median = 5, with an outlier = 2) were easier to understand than the Auto Focus cue (Median = 4, Min = 3). Additionally, most participants who used the Arrow cue indicated that they felt correctly guided in the immersive video (Median = 5, with an outlier = 2), as did participants who used the Radar (Median = 5, with an outlier = 4), and as did participants who used the Auto Focus cue (Median = 4). In addition to everything indicated above, the participants who used the Arrow cue as a visual guide (Median = 5, with an outlier = 2), felt that they needed to see the cue to orient their gaze to the main scene, this did not happen so much with the Radar (Median = 4) and with regard to the Auto Focus, we cannot obtain valid information referred the need to look at the cue, since the responses provided by the participants were highly variable. Finally, the Radar cue was better rated with respect to the distraction that this cue could cause to the viewer, compared to the other two visual cues, Arrow and Auto Focus.

6.4 The fourth stage: System usability

In order to improve the operation of the software, we continued evaluating its usability. We have decided to evaluate the usability of the system using the SUS scale designed by John Brooke in 1996 [5]. The evaluation of the developed visual cues reveals results observed in Fig. 15.

To better understand these results, it is necessary to know that the results obtained using the SUS usability scale has a specific meaning depending on the score obtained and the range

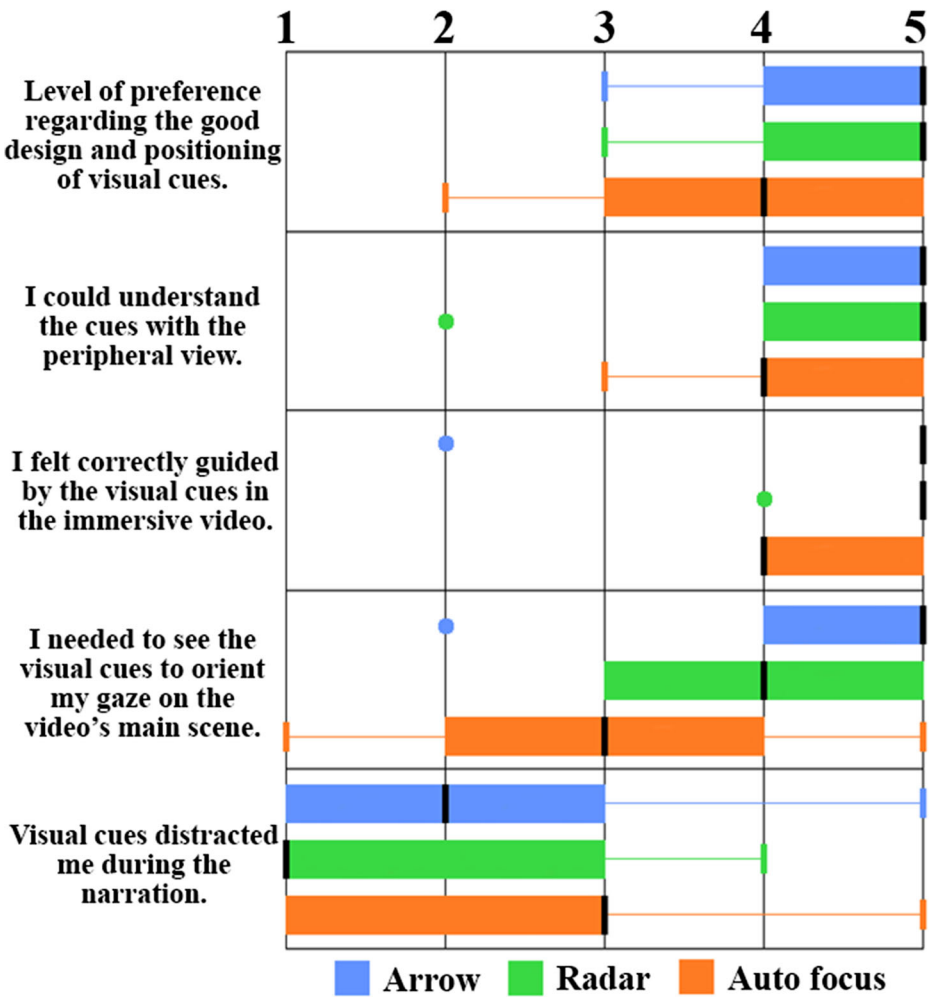


Fig. 14 System Design Assessment, where 1 = totally disagree, 3 = neither agree or disagree, and 5 = totally agree

where it is found [6]. There are several ways to classify it, in our case we use the following ranges:

- Score < 50: Not Acceptable.
- 50 < Score < 70: Marginal.
- Score > 70: Acceptable.

Taking this into account, we can analyze Fig. 15. Regarding the usability of the system, for all users who used the Radar visual cue, the usability measurement indicates that this is an Acceptable system. Regarding the users who used the system with the visual cue Arrow, their responses in general indicated that the usability of the system is acceptable (84.6% of the users), except for the responses provided by two participants, which indicated that the usability of the system is marginal. Finally, with respect to the users who used the system with Auto Focus, their answers do not provide a concrete result, around 69.2% of the results indicate that

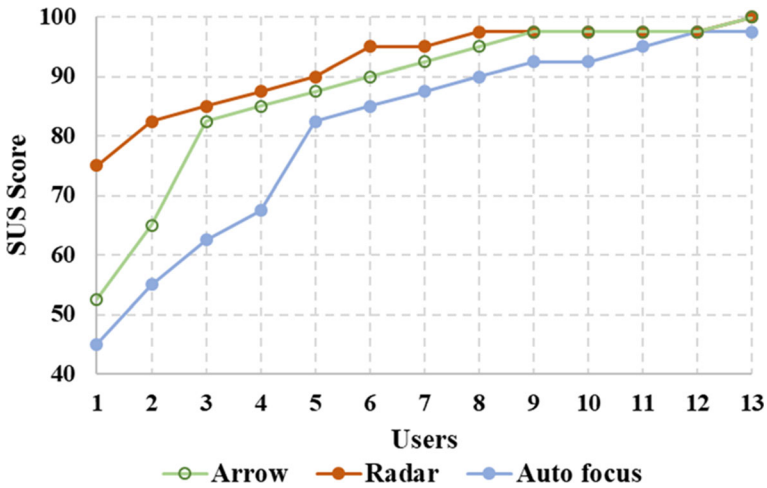


Fig. 15 Usability analysis of the system for each of the visual cues: Arrow, Radar and Auto Focus

it is a system with acceptable usability, however, around 23% indicate that it is marginal and we even obtained a result (7.7%) indicating that the usability of the system is not acceptable.

7 Discussion

7.1 Focus of attention

The three visual cues analyzed in our study allowed us to guide the user to the main scene. When comparing the results obtained by users who used visual cues and those who did not use these cues, a significant difference was obtained ($p = 0$, being $p < 0.05$).

There are previous works where different visual cues have been designed that also allow guiding the user to the point of interest. In a previous study, two visual cues were designed (shadow and bubble) and an experiment was carried out to evaluate the users' focus of attention and compare the results with a control group that did not use any type of visual cue to guide their vision. They found a significant difference when using the bubble cue to guide the users' gaze ($p = 0.021$) [21]. However, unlike us, they used eye tracking as a data collection method, which can generate more accurate information, it would be interesting to extend our study by performing an eye tracking. Another study compares the results of using the central visual cue (an arrow in the center of the field of vision) to guide the user, using a visual cue on the sides (flicker), they indicated that there was no difference between the two cues in the objective of guiding the user, they reached this conclusion by comparing the "Time per Degree to Reach Regions of Interest" using a Wilcoxon Signed-Rank Test, they obtained no significant difference between the cue arrow ($Md = 0.05$) and flicker ($Md = 0.03$), $z = -1.89$, $p = 0.06$, with a medium effect size ($r = 0.27$), however, they indicate that the visual cues on the sides are less invasive [31]. This is the main reason why we have designed the visual cues on the sides of the field of vision. A third study made the comparison between different environments, 3 using diegetic cues, 4 using non-diegetic cues and one without using a cue to guide the user, in this study it was concluded that the best cue to guide the user is the visual cue that they called "Object to Follow", which used an object that the user had to follow

to reach the point of interest (For the experiment with the HMD device), this was an implicit non-diegetic cue that allows interaction [33].

7.2 Understanding of narrative

The visual cues Auto Focus and Radar managed to obtain a significant improvement in the understanding of the narrative of the video compared to the group that did not use guide cues ($p = 0.004$ and $p = 0.027$ respectively, being $p < 0.05$), while the Arrow cue did not achieve such a significant improvement in the understanding of the user's narrative. This verifies that the control over the spectator's gaze while viewing a film does not guarantee control over comprehension and affective response, this has been described in previous works in the field of cinematography [16]. However, this has not been extensively studied in immersive virtual environments.

7.3 User's experience

Auto Focus was the least accepted visual cue by users with respect to usability, even though users had better narrative understanding when using this cue compared to not using a visual cue, and even their results were better than the results of participants who used the arrow as a guide cue, which is consistent with previous work. One study concluded that the Auto Focus (forced rotation) is the least accepted way of guiding with respect to the user experience, even though it was the third best with respect to the identification of objects of interest [33]. On the other hand, considering that this is an immersive environment that allows exploration (it is not a sports video which we must follow each movement), the fact that the visual Auto Focus cue is not accepted by users is in line with conclusions obtained in previous works [19].

One of the higher rated questions in the survey on visual cue design was whether users felt correctly guided by the visual cue in the immersive video (for any cue used). In this question, most participants indicated that they felt correctly guided in the immersive video (only one participant indicated the opposite). On the other hand, discarding the participants who used Autofocus, only one participant indicated that it was not necessary to see the visual cue to look for the main scene, while 5 others indicated a neutral response, the rest of the participants indicated that they needed to see the visual cue to look for the main scene (discarding the participants who used Autofocus, since the screen automatically looked for the main scene). The behavior described above proves that participants can feel the need to be guided when they are in an immersive environment, this has been proven even in environments where the user is not required to perform a specific task [21]. With respect to the level of distraction that users perceived with the use of visual cues, the highest rated cue was Radar (Median = 1, Max = 4) compared to Arrow (Median = 2, Max = 5) and Auto Focus (Median = 3, Max = 5).

8 Conclusions and future work

In our study, we proposed that the correct selection and design of visual cues not only allows us to guide the viewers' gaze to the main scene of the video but also improve the user's narrative understanding. For this reason, we study the use of three visual cues to guide the viewer through the narrative of the video. The three cues studied were; the Arrow, which is widely used in the CVR area; Radar, a cue widely used in FPS Games; and finally, the Auto Focus, which is also widely used in CVR environments.

We have found that even by correctly guiding the user's gaze to the main scene, the viewer may not correctly understand the narrative of the video. In the CVR area it is necessary to consider three criteria, the focus of the users, their correct understanding of the narrative and the experience of the users in the immersive environment, since the visual cues can achieve two of these without guaranteeing the third. In our work, only the Radar visual cue obtained a positive response in these three aspects evaluated.

Based on the results obtained, the design of the Radar cue has a greater acceptance of use than the Auto Focus system. But we should also point out that the Auto Focus enables the user to have a higher level of concentration on the plot at the price of limiting freedom within immersive environments. The software system with Arrows is well accepted by users, since the design is simple and the cue is widely known to users, capturing the user's attention and guiding the viewer towards the main scene of the immersive video. But contrary to this level of attention produced by the cue, the Arrow cue does not improve the user's narrative understanding. We have observed that improperly designing a visual cue can allow the user to be guided to the main scene, while still being comfortable with the experience, but not achieving a correct narrative understanding, probably because the visual cue distracts the user. We could conclude that Arrows can cause a certain degree of distraction that does not allow the user to concentrate on the video narrative. Therefore, for future work, we will consider studying an eye tracking, instead of tracking the movement of the head, in order to observe in more detail, the exact points where the user focuses the gaze.

On the other hand, after analyzing the results, we can conclude that the user tends to look at the visual guidance cues when there is a change of scene of 180 degrees. These scene changes should not be considered when creating or producing an immersive video.

For future work we propose to carry out a study on listening comprehension in immersive videos, orienting the study to the area of education. Since it is a valid alternative to generate new products for use in education. On the other hand, it can be considered important to carry out a study to analyze how much an immersive video can influence the purchase preference of a product within the field of digital marketing. This study can focus on other areas, such as in the development of applications focused on tourism and heritage preservation; similar studies can be conducted to develop support applications for rehabilitation treatments in the area of health; or any other application that requires not only that users observe a specific objective, but also that users understand what is happening and what they are hearing.

On the other hand, as mentioned above, we believe that it is necessary to expand our study by conducting an eye-tracking evaluation. In addition to this, in future work we plan to design other visual or sound cues, diegetic and non-diegetic, to obtain better methods that allow us to improve the user's focus of attention and the narrative understanding.

Authors' contributions Conceptualization, Galo Ortega-Alvarez and Angel Garcia-Crespo; methodology, Galo Ortega-Alvarez and Angel Garcia-Crespo; software, Galo Ortega-Alvarez, Carlos Matheus-Chacin and Adrian Ruiz-Arroyo; validation, Galo Ortega-Alvarez and Carlos Matheus-Chacin; formal analysis, Galo Ortega-Alvarez, Angel Garcia-Crespo and Carlos Matheus-Chacin; investigation, Galo Ortega-Alvarez, Angel Garcia-Crespo, Carlos Matheus-Chacin and Adrian Ruiz-Arroyo; resources, Galo Ortega-Alvarez and Angel Garcia-Crespo; data curation, Galo Ortega-Alvarez, Carlos Matheus-Chacin and Adrian Ruiz-Arroyo; writing—original draft preparation, Carlos Matheus-Chacin; writing—review and editing, Galo Ortega-Alvarez and Angel Garcia-Crespo; visualization, Galo Ortega-Alvarez and Adrian Ruiz-Arroyo; supervision, Angel Garcia-Crespo; project administration, Galo Ortega-Alvarez; funding acquisition, Galo Ortega-Alvarez and Angel Garcia-Crespo. All authors have read and agreed to the published version of the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflicts of interest/competing interests Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Aitamurto T, Zhou S, Sakshuwong S, Saldivar J, Sadeghi Y, Tran A (2018) Sense of presence, attitude change, perspective-taking and usability in first-person split-sphere 360° video. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI18) (pp 1–18). <https://doi.org/10.1145/3173574.3174119>
2. Balakrishnan B, Sundar SS (2011) Where am I? How can I get there? Impact of navigability and narrative transportation on spatial presence. *Human-Comput Int* 26(3):161–204. <https://doi.org/10.1080/07370024.2011.601689>
3. Boonsuk W, Gilbert S, Kelly J (2012) The impact of three interfaces for 360-degree video on spatial cognition. In SIGCHI Conference on Human Factors in Computing Systems - CHI '12 (pp 2579–2588). Austin, Texas. <https://doi.org/10.1145/2207676.2208647>
4. Briceño LA, Velásquez FR, Peinado S (2011) Influencia de los mapas conceptuales y los estilos de aprendizaje en la comprensión de la lectura. *Revista Estilos de Aprendizaje* 4(8):3–22
5. Brooke J (1996) SUS - A quick and dirty usability scale. (Jordan P, Thomas B, McLelland I, Weerdmeester B Eds.). London: Taylor and Francis
6. Brooke J (2013) SUS: a retrospective. *J Usability Stud* 8(2):29–40
7. Brown AJ, Sheikh A, Evans M, Watson Z (2016) Directing attention in 360-degree video. In IBC 2016 Conference. <https://doi.org/10.1049/ibc.2016.0029>
8. Burigat S, Chittaro L (2007) Navigation in 3D virtual environments: effects of user experience and location-pointing navigation aids. *Human Comput Stud* 65(11):945–958. <https://doi.org/10.1016/j.ijhcs.2007.07.003>
9. Chen CJ, Ismail W (2008) Guiding exploration through three-dimensional virtual environments: a cognitive load reduction approach. *J Int Learn Res* 19(4):579–596
10. Fagerholt E, Lorentzon M (2009) Beyond the HUD - User interfaces for increased player immersion in FPS games. Chalmers University of Technology
11. Fombona Cadavieco J, Sevillano Pascual MÁ, Madeira Ferreira Amador MF (2012) Realidad aumentada, una evolución de las aplicaciones de los dispositivos móviles. *Píxel-Bit Rev Medios y Educ* 41:197–210
12. Fox B (2005) *Game interface design*. (THOMSON, Ed.). Stacy L.Hiquet
13. Haffeege A, Barrow R (2009) Eye tracking and gaze based interaction within immersive virtual environments. In ICCS 2009 Proceedings of the 9th International Conference on Computational Science (pp 729–736). Baton Rouge. https://doi.org/10.1007/978-3-642-01973-9_81
14. Hong S, Kim GJ (2016) Accelerated viewpoint panning with rotational gain in 360 degree videos. In 22nd ACM Conference on Virtual Reality Software and Technology (pp 303–304). Munich. <https://doi.org/10.1145/2993369.2996309>

15. Hu H-N, Lin Y-C, Liu M-Y, Cheng H-T, Chang Y-J, Sun M (2017) Deep 360 pilot: learning a deep agent for piloting through 360° sports videos. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE
16. Hutson JP, Smith TJ, Magliano JP, Loschky LC (2017) What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cogn Res Princ Implications* 2(1):46. <https://doi.org/10.1186/s41235-017-0080-5>
17. Ishiguro Y, Rekimoto J (2011) Peripheral vision annotation: noninterference information presentation method for mobile augmented reality. In Proceedings of the 2nd Augmented Human International Conference - AH '11 (p 8:1–8:4). New York. <https://doi.org/10.1145/1959826.1959834>
18. Lazar J, Feng JH, Hochheiser H (2010) Research methods in human-computer interaction. (John Wiley & Sons, Ed.)
19. Lin Y-C, Chang Y-J, Hu H-N, Cheng H-T, Huang C-W, Sun M (2017) Tell me where to look: investigating ways for assisting focus in 360° video. In 2017 CHI Conference on Human Factors in Computing Systems (pp 2535–2545). Denver, Colorado. <https://doi.org/10.1145/3025453.3025757>
20. Löwe T, Stengel M, Förster E-C, Grogoric S, Magnor MA (2015) Visualization and analysis of head movement and gaze data for immersive video in head-mounted displays. In Workshop on Eye Tracking and Visualization (ETVIS) 2015 (pp 1–5). Chicago
21. Mäkelä V, Keskinen T, Mäkelä J, Kallioniemi P, Karhu J, Ronkainen K, Burova A, Hakulinen J, Turunen M (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 (pp 13–24). ACM. <https://doi.org/10.1145/3317697.3323351>
22. McCabe H, Kneafsey J (2006) A virtual cinematography system for first person shooter games. In International Digital Games Conference (pp 25–35). Portalegre
23. Neng LAR, Chambel T (2010) Get around 360° hypervideo. In 14th International Academic MindTrek Conference: Envisioning Future Media Environments (pp 119–122). Tampere. <https://doi.org/10.1145/1930488.1930512>
24. Nguyen TTH, Duval T, Fleury C (2013) Guiding techniques for collaborative exploration in multi-scale shared virtual environments. In GRAPP International Conference on Computer Graphics Theory and Applications (pp 327–336). Barcelona, Spain. <https://doi.org/10.5220/0004290403270336>
25. Nielsen LT, Møller MB, Hartmeyer SD, Ljung TCM, Nilsson NC, Nordahl R, Serafin S (2016) Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In 22nd ACM Conference on Virtual Reality Software and Technology (pp 229–232). Munich. <https://doi.org/10.1145/2993369.2993405>
26. Ramalho J, Chambel T (2013a) Immersive 360° mobile video with an emotional perspective. In 2013 ACM international workshop on Immersive media experiences (pp 35–40). Barcelona. <https://doi.org/10.1145/2512142.2512144>
27. Ramalho J, Chambel T (2013b) Windy sight surfers: sensing and awareness of 360° immersive videos on the move. In Proceedings of the 11th European Conference on Interactive TV and Video - EuroITV '13 (pp 107–116). Como. <https://doi.org/10.1145/2465958.2465969>
28. Rothe S, Hussmann H, Allary M (2017) Diegetic cues for guiding the viewer in cinematic virtual reality. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (p 54). ACM. <https://doi.org/10.1145/3139131.3143421>
29. Sarker B (2016) Show me the sign!: The role of audio-visual cues in user experience of mobile virtual reality narratives. Master's thesis, UPPSALA UNIVERSITET
30. Saunders K, Novak J (2012) Game development essentials: game interface design. (Cengage Learning, Ed.) (2nd ed.)
31. Schmitz A, MacQuarrie A, Julier S, Binetti N, Steed A (2020) Directing versus attracting attention: exploring the effectiveness of central and peripheral cues in panoramic videos. In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp 63–72). <https://doi.org/10.1109/VR46266.2020.00024>
32. Simeone AL (2016) The VR motion tracker: visualising movement of non-participants in desktop virtual reality experiences. In 2016 IEEE 2nd Workshop on Everyday Virtual Reality (WEVR) (pp 1–4). Greenville, SC. <https://doi.org/10.1109/WEVR.2016.7859535>
33. Speicher M, Rosenberg C, Degraen D, Daiber F, Krüger A (2019) Exploring visual guidance in 360-degree videos. In Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video (pp 1–12). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3317697.3323350>
34. Suma EA, Bruder G, Steinicke F, Krum DM, Bolas M (2012) A taxonomy for deploying redirection techniques in immersive virtual environments. In 2012 IEEE Virtual Reality Workshops (VRW) (pp 43–46). Orange County, California. <https://doi.org/10.1109/VR.2012.6180877>

35. Tanaka R, Narumi T, Tanikawa T, Hirose M (2016) Motive compass: Navigation interface for locomotion in virtual environments constructed with spherical images. In 2016 IEEE Symposium on 3D User Interfaces (3DUI) (pp 59–62). Greenville, SC. <https://doi.org/10.1109/3DUI.2016.7460031>
36. Vosmeer M, Schouten B (2014) Interactive cinema: engagement and interaction. In: Mitchell A, Fernández-Vara C, Thue D (eds) International conference on interactive digital storytelling. Springer International Publishing, Cham, pp 140–147. https://doi.org/10.1007/978-3-319-12337-0_14
37. Zammito V (2008) Visualization techniques in video games. In Electronic Visualisation and the Arts (EVA 2008) (pp 267–276). London. <https://doi.org/10.14236/ewic/eva2008.30>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.