FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Visualizing and Interacting with 360° Web-based Videos using Dynamic Annotations

Teresa Carla de Canha e Matos

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Rui Rodrigues

Co-Supervisor: Rui Nóbrega

February 9, 2018

Visualizing and Interacting with 360° Web-based Videos using Dynamic Annotations

Teresa Carla de Canha e Matos

Mestrado Integrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: A. Augusto de Sousa

External Examiner: Paulo Dias

Supervisor: Rui Rodrigues

February 9, 2018

Abstract

The use of 360° videos has been increasing steadily in the 2010s, as content creators and users search for more immersive experiences. This trend was influenced by the development of virtual reality devices, continually more efficient and more accessible to the common user. This type of videos differs from traditional videos as they completely surround the user in its environment, and allow them to look around the scene.

The freedom to choose where to look at, which 360° videos are able to provide, may hinder the overall experience instead of enhancing it, as there is no guarantee that the user will focus on relevant sections of the scene. In the case of videos that contain a message that is transmitted throughout the video relative to different points of interest, that is, videos with narrative, the message may be lost to the users in this context. Video annotations, such as text boxes, arrows or captions, superimposed on the video and associated to points of interest, can help guide the user through the narrative of the video and better understand its content, while maintaining freedom of movement.

This dissertation consists of the conceptualization, implementation and evaluation of a web-based visualizer of 360° videos with superimposed visual annotations. A set of annotations was created, with the purpose of providing information or guiding the user to points of interest or to the narrative's area. Throughout the video, these annotations may be dynamic, adapting to their linked point of interest's movement or to the user's control over the field-of-view. The visualizer allows users to experience these annotated 360° videos in a computer, using keyboard and mouse, or in HTC Vive and mobile devices with Cardboard VR headsets, to experience the video in virtual reality.

As the development of the visualizer was focused on user design, short user tests were performed throughout the implementation phase. The final prototype was then evaluated, by carrying out usability tests, where users experienced several 360° videos with one or more of the implemented annotations applied to each. The participants answered questionnaires where they evaluated the use of these annotations, comparing the experience without annotations and with two different annotations of the same category. The results obtained from these experiments demonstrate that annotations can assist in guiding the user during the video, although their design is imperative for the user's focus on the video. Annotations with a familiar design, that do not occupy much of the visible area and that adapt to the video's scene can be advantageous, while those that cover a larger area of the video can be considered forceful and intrusive, interfering with the immersive experience. Analyzing these results, it was possible to identify what characteristics in an annotation are of benefit for the overall experience, which can be used to devise new types of annotations in the future.

***Keywords* - 360° video, Virtual Reality, Visualization Interfaces**

Resumo

O uso de vídeos 360° tem vindo a aumentar na década de 2010, com a crescente procura por experiências imersivas, tanto por criadores de conteúdos multimédia como por consumidores. Esta tendência foi influenciada pelo desenvolvimento de dispositivos de realidade virtual, que são cada vez mais eficientes e acessíveis ao utilizador comum. Este tipo de vídeos diferencia-se de vídeos tradicionais devido à sua capacidade de envolver por completo os espectadores no seu ambiente, que têm controlo sobre o que observam.

A liberdade de escolher para onde olhar que os vídeos 360° fornecem pode prejudicar a experiência imersiva, em vez de a melhorar, por não haver garantias que o utilizador irá focar-se nas áreas relevantes da cena. No caso de vídeos que contêm uma mensagem que é transmitida ao longo do vídeo, relativa a diferentes pontos de interesse, ou seja, vídeos com narrativa, os espectadores podem perder esta mensagem neste contexto. Anotações de vídeo, como caixas de texto, setas, ou legendas, que são sobrepostas no vídeo e associadas a pontos de interesse, podem guiar o utilizador ao longo da narrativa do vídeo, e facilitar a compreensão do seu conteúdo, mantendo sempre a liberdade de movimento.

Esta dissertação foca-se na conceptualização, implementação, e avaliação de um visualizador web de vídeos 360° sobrepostos por anotações visuais. Um conjunto de anotações foi criado, com o propósito de fornecer informação ou guiar o utilizador, para pontos de interesse ou para a área da narrativa. Ao longo do vídeo, estas anotações podem ser dinâmicas, adaptando-se à posição do ponto de interesse ao qual estão associados ou a mudanças do campo visual. O visualizador disponibiliza aos seus utilizadores estes vídeos 360° anotados, que podem ser experienciados usando o computador, com teclado e rato, ou o dispositivo de realidade virtual HTC Vive conectado ao computador, ou ainda com dispositivos móveis, utilizando um Cardboard, para visualizar o vídeo em realidade virtual.

Como o sistema foi desenvolvido seguiu um *design* centrado no utilizador, curtos testes de utilizador foram realizados ao longo da fase de implementação. O protótipo final foi depois avaliado, com a realização de testes de usabilidade, no qual os participantes experimentaram diversos vídeos 360° com uma ou mais anotações aplicadas em cada. Os participantes responderam a questionários, avaliando a aplicação das anotações, especificamente, comparando a experiência sem anotações e com tipos diferentes de anotação da mesma categoria. Os resultados obtidos demonstraram que anotações podem ser usadas para guiar o utilizador durante o vídeo, no entanto o seu *design* é imperativo no que se refere ao foco do utilizador no vídeo. Anotações que tenham um *design* familiar, que não ocupem muito espaço no campo visual, e que se adaptem à cena do vídeo são benéficas para a experiência, sendo que anotações que cobrem uma maior área da imagem do vídeo podem ser consideradas intrusivas e incomodativas, interferindo com a experiência imersiva. Analisando então os resultados dos testes de utilizador, foi possível identificar as características de anotações que favorecem a experiência em geral, sendo que poderão ser utilizadas para definir novos métodos de anotação no futuro.

***Keywords* - Vídeos 360°, Realidade Virtual, Interfaces de Visualização**

Acknowledgements

To my mother and father, who have always been a constant and consistent support for all my life ventures. To my boyfriend, Mário, for believing in me when at times I wouldn't, and for all the unwavering patience during this entire phase.

To my supervisors, Rui Rodrigues and Rui Nóbrega, who were always supportive and available to guide me whenever I needed throughout the entire process.

To my close friends and also colleagues, Sara, Ivo, Daniel and Valter, among several others, with whom I shared this experience in FEUP, I was able to overcome many hardships thanks to your company and support.

Teresa Matos

"It is necessary to keep one's compass in one's eyes and not in the hand,
for the hands execute, but the eye judges."

Michelangelo

Contents

List of Figures

List of Tables

LIST OF TABLES

Abbreviations

AR	Augmented Reality
AV360	Augmented Video 360
FOR	Field of Regard
FOV	Field of View
HMD	Head-mounted Display
HUD	Head-up Display
POI	Point of Interest
RQ1	Research Question 1
RQ2	Research Question 2
VE	Virtual Environment
VR	Virtual Reality

Chapter 1

Introduction

Immersive videos are 360° video recordings that display the scene centered around the camera in every direction. This type of content became more popular in the 2010s, which was influenced by the higher consumption of mobile video, and also by the increased use of virtual reality devices, mostly in the entertainment industry [AEBD17]. There is a growing search for more realistic and immersive experiences, and 360° videos can be used for this end. Several content hosting services have taken notice of this trend, such as Youtube and Facebook, which released in 2015 a feature for uploading and viewing 360° videos on their platforms, followed by several other similar companies. Likewise, camera manufacturing companies are investing in the development of 360° cameras, progressively more accessible for both amateur and professional content creators, while also increasing and improving the features provided [Kes16].

Currently, this type of videos is being used in different areas, for example sports, as several teams, sports channels and sports associations are creating and uploading 360° videos on social media to bring their fans closer to the playing field[1]. Several marketing and advertising companies have used 360° videos to provide the audience with a more engaging and detailed view of the product, whether it is a tech device, a car or a house. In 2016, Google created a project, collaborating with 360 Labs and Columbia Sportswear, to test the influence of 360° view in the viewer metrics [Hab16]. Two similar advertising campaigns were uploaded to Youtube, one filmed in 360° and the other in traditional format. Analyzing the view-through rate and viewer retention statistics for both videos, the results showed that, although less users watched the entire 360° video, they interacted more with it, and showed more interest in sharing its content and finding other videos of the same type.

Another industry that is presently taking advantage of the 360° videos features, for its ability to provide a broader picture of the scene, as well as to create emotional engagement with the narrative, is the news industry. As journalists search for more modern approaches to create content and to reach out to the public, VR and 360° videos are viewed as an effective tool for this purpose.

[1] "VIDEO: Watch Our 360-Degree VR Special With NBA Superstar Kobe Bryant". 2016. Eurosport. http://www.eurosport.com/basketball/nba/2015-2016/video-watch-our-360-degree-vr-special-with-nba-superstar-kobe-bryant_sto5700527/story.shtml.

Immersive journalism is a recent field directly related to this concept, and it focuses on creating a first-person experience within an interactive storytelling journalistic piece [VPC12]. Many prominent broadcasting networks, such as ABC [2], CNN [3], Euronews [4], as examples, now regularly produce 360° video documentaries, commonly viewable through the computer, smartphone, or VR devices.

1.1 Context and Motivation

Content creators use videos to create and present a story, whether it is fictional or factual. A great deal of time is spent on planning before filming, by developing a storyboard, creating the script, and deciding how to use lighting and sound effects, all in order to compel the audience to focus on the narrative of the video.

360° videos pose a challenge regarding the audience's focus within the video, as the users are given the freedom to look around the scene, and in doing so, they may miss certain pieces of information and, ultimately, the video's message is lost. Steps to overcome this problem can be taken during the planning phase, which means that storyboarding and other processes must be different from what is typically done for traditional videos. Alternatively, during the editing phase, the content creators may try to artificially draw attention to a certain point in the video by placing additional media content over the video's image. Several 360° documentaries use these media objects, in the format of text or graphic icons, in order to provide more information, or to indicate an important point in the scene, as exemplified in figure 1.1. However, these visual objects are edited into the video, essentially altering the video's file, and they do not solve the issue of having important information out of the user's view.

The positioning of media objects in the video's image is another issue that arises when creating 360° videos. An example of this issue is the case of subtitles, which in most video players are commonly placed at the bottom of the screen, using a text font, color and size to ensure its readability. For immersive videos and experiences, however, there is no established standard. In some cases, the subtitles are placed at several fixed positions in the video's image, so that the user can notice them from any direction, while in others, the subtitles are presented as an object of the head-up display (HUD), making them constantly visible for the user.

Considering that 360° videos are used to create a more extensive and immersive experience, similarly to virtual environments, one can search further into this area to find techniques that are applicable to 360° videos, with the objective of solving these issues. In the case of Augmented Reality, which consists of overlaying virtual objects in real scenarios, the methods used to create these environments, their features and issues, can be examined and adapted to be applied in the context of immersive videos. The tools used for the different stages of the creation of traditional video must then be adapted to 360° videos's characteristics requirements. In 2016, Meira et al.

[2] ABC News VR. http://abcnews.go.com/VR
[3] CNN VR. http://edition.cnn.com/vr
[4] Euronews 360-degree videos. http://www.euronews.com/tag/360-video

Figure 1.1: Still image from a news video "A 360° political immersion in Martinique".
Source: Euronews [Eur17]

[MMJ+16] analyzed different methods of visual annotations in the context of 360° videos and immersive journalism, and subsequently created an annotation tool prototype, which allowed the user to identify points of interest (POIs) using a video mask.

In the following year, a project called Augmented Video 360 (AV360)[5] was created, with the goal of researching and creating editing and visualizing tools to be used in the context of immersive journalism. In order to create augmented 360° video experiences, firstly and immersive editing and annotation tool was developed, which allows the use of virtual reality devices to place dynamic annotations in the immersive's video scene. The project's immersive visualizer is focus of the work described in this document, and it consists of a web video player able to read and play the edited 360° videos containing dynamic annotations, and is compatible with different interaction devices, including VR devices such as HTC Vive.

1.2 Research Questions

Visual annotations can be used in 360° videos to support their narrative. A block of text may be placed next to a point of interest to provide more information about it, or a warning symbol can be used to call the user's attention towards an area that is out of their view. Therefore, the purpose of the annotations, as well as their appearance, can vary greatly, and its application may have different implications in the user's ultimate experience. The focus of this dissertation is to research on different annotation methods, in varied contexts, and afterwards define, implement, and test various annotations in a 360° video. To guide the research work, design and development of the proposed system, the following research questions are defined:

[5] Augmented Video 360 Main Page. https://dei.fe.up.pt/gig/projects/av360-augmented-video-360/

- Research Question 1 - How to **enhance** 360° visualization with **dynamic annotations**?

  What are the different annotations that can be placed in 360°, and what is their purpose in this context? How should they be presented to the user, in terms of positioning and design?

- Research Question 2 - What is the impact in **user experience** and **knowledge intake** of different dynamic annotation paradigms?

  Are annotations in fact helpful for the user, or do they remove hinder the immersive aspect of a 360° videos? Are users able to remember the content and information provided on the video due to the annotations, or are these objects a cause for distraction and loss of focus? How do different annotations perform in the same conditions and what causes potential differences in their efficiency?

These research questions will be continuously referred to throughout the document, as they were relevant for the several stages of work. For simplification purposes, the first and second question will henceforth be referred to as **RQ1** and **RQ2**, respectively.

1.3 Objectives

Considering the formerly established research questions, a list of objectives for the dissertation's work are proposed:

- Research on the evolution and current state of virtual environments and interaction devices, the creation of 360° video content, and methods of information overlay.

- Definition of a list of annotations to be implemented, that may contribute to the immersive video's experience in distinct contexts;

- Design of a visualizer's prototype, which is compatible with the editor tool developed in the context of the AV360 project, and is able to present previously defined annotations;

- Implementation of the web visualizer according to the previously defined requirements, supporting the playback of 360° videos and the overlay of specified annotations, using either a computer, a VR headset or a mobile device;

- Evaluate the finalized prototype through user studies, to analyze the impact of the annotations on the experience of an immersive video, as required for the second research question (RQ2).

Related to the first research question (RQ1), the research on methods of information overlay, or visual annotations, should cover their application in different contexts, their design, and potential visualization issues. The last objective, regarding the user experiments, covers the topic of RQ2. After implementing a visualizer prototype, capable of displaying annotated 360° videos, it was important to run user studies on the system, which compared the experience of a user with and without annotations, for different purposes and video themes. These objectives were used to guide the development process that followed, and are covered throughout the present document.

1.4 Implemented Solution

As part of the objectives for this dissertation, a prototype of a 360° video player was created. The video player is web-based, to be accessible in different devices, therefore a user can access the visualizer through a browser on the computer, and watch and interact with the video using the default screen, keyboard and mouse. Alternatively, they can opt for an immersive experience by using the HTC Vive headset to watch the video in virtual reality. It is also possible, albeit limited, to use the video player on a mobile device, in regular or VR mode, by attaching the device to a Cardboard VR headset[6] or another type of compatible headset.

The visualizer provides a list of available 360° videos, which are complemented by annotations placed in its scene. Besides the standard subtitles, as they are used in traditional videos, there are annotations that display a combination of text and images, These are placed at any point in the scene to provide more information about the video's general theme, or a specific point of interest. The annotations may be dynamic, in order to accompany a moving point of interest, for example, and the users are able to interact with them as well, to obtain otherwise hidden information.

When the user is looking away from one of these annotations, an additional annotation is used to indicate the existence and direction of relevant information that is out of the user's view. Furthermore, as certain videos have an underlying narrative line, some other annotations are applied to draw the attention of the user towards the narrative's area of interest.

1.5 Contributions

The first contribution of this dissertation is the implementation of a prototype for the visualizer, as described in the previous section. This was preceded by a research on virtual environments, their history and applications in diverse areas, as well as the different aspects regarding the production and presentation of 360° videos, and naturally, the review of methods of information overlay, that is, the annotations. The developed system provides an immersive experience with this type of videos, while using some additional visual annotations to guide the user throughout the video. This thesis also proposes a list of annotations that can be used to enhance 360° videos, providing a description of their functionalities, possible designs and more technical aspects relative to their implementation and rendering.

The performed user tests add to these contributions, by providing some considerations on the user's experience regarding the use of the described annotations and the overall efficiency of the system. The results from this stage can be used to guide future work in this area, whether it is to create new types of annotation or to revise and improve the existing ones.

As mentioned previously in this chapter, the work developed in this dissertation is part of the Augmented Video 360 project, which was started in 2016 in the Institute for Systems and Computer Engineering, Technology and Science of Porto (INESC TEC).

[6] Google Cardboard for mobile devices. https://vr.google.com/cardboard/

1.6 Document Structure

This initial chapter introduced the context of the project, presenting the opportunities relative to the use of 360° videos, while also expressing its issues in certain contexts such as journalism. After the definition of the research questions in this chapter, the following Chapter 2 presents the results of the literature review, beginning with Virtual Reality, its history, and current interaction methods and devices, followed by 360° video filming, editing and viewing, ending with a revision of visual annotations used in images and videos, their characteristics and issues that might occur regarding its visualization.

The third chapter (Chapter 3) presents the proposed architecture for visualizer, followed by a list of system requirements, and a conceptualization of annotations that could be implemented. This chapter closes with a description of the work methodology adopted, from the beginning of the development stage to the conclusion of the users tests.

The next chapter (Chapter 4) describes in more detail the implementation of the system, starting with a summary of the tools used for the process, followed by an explanation, from a technical perspective, of the operation of the system, detailing how the video is presented in a 360° view for the different compatible devices, and how each visual object is created and rendered throughout the video.

Chapter 5 covers the testing phase of the project, presenting the test protocol and structure, and the 360° content created for the purpose of these tests. Afterwards, an overview of the gathered data is presented, which is then analyzed to draw conclusions regarding the efficiency of the used annotation methods and the system in general.

Lastly, Chapter 6 reviews and presents some conclusions on the entire process of the dissertation. In the same chapter, a proposal of improvements to the prototype are presented, as well as a description of possible future features and implementations for this project.

Chapter 2

State of the Art Review

The purpose of this chapter is to provide an overview on the areas of study related to 360° videos, immersive experiences, and visual annotations, by analyzing their history and also relevant current works that can be used as a basis for the development of the immersive visualizer. Firstly, the definition and history of Virtual Environments (VEs) will be reviewed, followed by a review of present VR technologies and uses. Secondly, 360° videos, their filming, editing, and viewing will be analyzed, and lastly, visual annotations will be introduced, reviewing their features and common uses in images, videos, and virtual environments, so as to identify approaches that may be implemented in the system being developed.

2.1 Virtual Environments

Virtual reality is a concept that became rather popular recently in the entertainment industry, however it has been a prevalent area of study for decades. It is said that the first use of the expression "virtual reality" to represent this research area occurred in the late 80s, by Jaron Lanier [Her89], however, systems that could stimulate several human senses simultaneously and depict a virtual world have existed since as early as the 1950s. In those early days, the definition of virtual reality was closely related to the technologies, as Lanier himself said, referring to it as "technology that uses computerized clothing to synthesize shared reality. It recreates our relationship with the physical world in a new plane, no more, no less." [Her89]. However, this technological definition was regarded as incomplete, and so it shifted towards a more theoretical description. Steuer proposed that virtual reality should be considered as a type of human experience, using **presence** as the key concept for this definition [Ste92]. Presence, in its general term, refers to the sense of being in an environment, which is achieved through providing information from that environment to the human senses. In the physical world, this occurs automatically and directly, while in virtual environments (VE), this transferral of data is mediated by interaction devices. However, while interacting with a virtual world, one is still aware and perceives information from the real world, and so, in order to efficiently create a sense of virtual presence, the virtual environment must be able to block external data or overpower it.

Immersion is another concept that is frequently referenced and analyzed when discussing virtual reality and virtual experiences, with variations on its definition and even mixing with the concept of presence [CB16]. [SW97] characterizes immersion has "the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant", where each dimension of immersion varies in scale, contributing to higher levels of immersion. Presence is, therefore, dependent on the immersive features of the technologies being used, but also on the individual's psyche, their preferences, their susceptibility and willingness to accept other realities.

Virtual environments which create a higher sense of presence lead to more natural emotions and behaviors, which is why they can be used for training exercises and for some psychological treatments, by allowing individuals to interact with complex environments that would be otherwise too stressful or dangerous in the real world, while obtaining the same positive outcomes.

Creating the illusion of presence is not, however, the main intention of all variations of virtual environments. [MK94] presented the term **Mixed Reality**, which refers to an environment that combines virtual and real entities in varying degrees, located at any point in the virtuality continuum, presented in figure 2.1. **Augmented Reality** (AR) is a well known example of mixed realities, and it is defined as an overlay of virtual objects on real environment, and in real-time. AR is mostly used to provide additional information relevant for the task at hand that is not readily available, and it has applications in several areas, such as navigation systems, education and training, entertainment, and several others [ABF+01]. AR also differs from the entirely virtual environments in the interaction devices, as the purpose of AR is not to isolate the users from the physical world, but instead to allow them to perceive it with the least amount of obstruction.

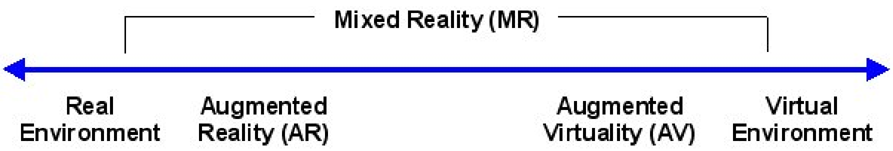


Figure 2.1: Virtuality-Reality Continuum. Milgram and Kishino, *Taxonomy of mixed reality visual displays*, 1994 [MK94]

2.2 VR Interacting Devices

There are several I/O devices available for interacting with a virtual environment through a computer, with varying consequences on the immersion of the experience. With the appearance of more advanced and affordable VR devices, games and other products with a virtual reality component have moved on from simple keyboard and mouse interaction, where the action occurs inside the screen, to a fully engaging and surrounding experience, that places the user in the center of the environment.

Interaction with virtual environments comes in many forms, defined by [Min95] in five fundamental categories, which are movement, selection, manipulation, scaling, and virtual menus and widgets, mirroring the ways one can interact with the real world. Proper interaction with a virtual object must be followed by clear feedback, so that the user can discern if their actions were successful, and also to increase their sense of presence in the VE, which is directly related to the increase of feedback to all human senses [CNdSR16].

In the following subsections, several interaction devices will be analyzed, by describing and comparing their different forms of interaction with virtual environments.

2.2.1 Keyboard and Mouse

The combination of keyboard and mouse is still the most common method of interacting with computer interfaces and, as such, is the one most people are accustomed with. Viewing the virtual environment through the screen, the user can pan around the scene using either the mouse, through clicking and dragging, or keyboard, with specified keys for each camera direction. Selecting and manipulating objects in the scene are typically performed with the mouse, as well as interacting with interface menus and buttons. Even though using these devices is, in most cases, simple and intuitive, there is an added difficulty when three-dimensional actions are required, as a mouse's movement is only two-dimensional. This can be overcome, as an example, by restricting movement to two dimensions, or combining key presses and mouse movement to define the axis of the movement, an example of which can be found in the POV Tech project [1], that uses the scroll button to recreate movement of the pointer in a third dimension. Additionally, it is not the most efficient way to interact with virtual environments, as it will still appear to the user that they are viewing and interacting with the scene from outside.

2.2.2 Touch Screen

The use of mobile devices, that is, smartphones and tablets, is fairly common, with 25% of the population owning smartphones in 2015 [2], not only for communications, but also for entertainment purposes, such as media content viewing. Substituting the physical keyboard for a touch screen facilitates the interaction with the system, and also allows for more complex and rich interfaces. These interactive screens have also been adopted into the computer industry for similar reasons, offering an alternative to the keyboard and mouse.

Touch screens are similar to the mouse, as the user's finger can be used as a cursor on the screen to select and move objects. It can be a more direct and intuitive form of interaction, providing more natural feedback through touch, yet it may be less precise, and provide less forms

[1] Chirls, Brian. 2014. *POV Tech: Introducing Mouse and Keyboard Interactions for Virtual Reality 'Films.'*. PBS. http://www.pbs.org/pov/blog/povdocs/2014/10/pov-tech-introducing-mouse-and-keyboard-interactions-for-virtual-reality-films/.

[2] Statista.com. *Smartphone Users Worldwide 2014-2020*. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

of interaction, as it does not have a hover state, for example. In addition, using a finger or styluses to interact with the screen means that the view of a portion of the screen will always be obstructed. Nevertheless, as many touch screens now support multiple inputs simultaneously, different combination of fingers and their movement on screen can be used to imitate the controls of the mouse. Kim et al. [KGMQ09] used this feature in iPhones to navigate a virtual world on the CAVE immersive space, implementing two interaction methods for translation and rotation of the viewpoint.

Regarding its use to interact with virtual environments, using touch screens may be simpler than keyboard and mouse, however, three-dimensional actions will still be an issue, and feeling of presence will not necessarily increase with its use.

2.2.3 Gamepad

Gamepads and joysticks can be used with computers and mobile devices as an alternative interaction device. For computers, these devices can substitute the keyboard and mouse, and certain interactions may be easier, such as steering. Regarding mobile devices, using gamepads, connected through cable or bluetooth, removes the issue with obstructing the screen with one's hand and overpopulating the interface with virtual buttons. A recent study compared the use of a gamepad against touchscreen with a mobile game, and concluded that touchscreen interfaces had better or similar button selection time, particularly with bigger screens, however, with more complex actions that may require more buttons, with the risk of occupying too much screen area [AI16].

2.2.4 Eye Tracking Devices

The direction of one's gaze can be indicative of different intentions, such as communication with another individual, focus of attention or intent of interacting with a certain event, person or object. Eye tracking devices have been used in order to implement an interaction with virtual interfaces based on gaze direction. These devices can be video-based, where a video stream will be processed in order to identify a person's eyes and their point of focus, video-based with infrared, by creating a corneal reflection that makes tracking the pupil more accurate, or electrooculography-based, where changes in the eye's potential field provoked by eye movements are measured by electrodes placed close to the eye, which makes it a more invasive approach, though more accurate in different lighting conditions and it does not require any image processing software [MB14].

Evaluating eye movement regarding a virtual interface provides several relevant data, which may indicate if the layout is optimal, and certain areas of higher interest. Different eye behaviors may be used to interact with the system, such as fixating on a certain object to manipulate it (as example, to open a menu), or moving the virtual body's position in the direction of the user's gaze. Eye tracking, however, tends to be inaccurate, as it depends fairly on the quality of the equipment, environmental conditions, and even interpretation of movements can be confusing [PB06]. Users with glasses will also have trouble using this interaction method. Nevertheless, if the system does not require highly accurate data from the eye tracker, these devices imply fast interactions, as eye movements are fairly quicker than other body movements, and are simple enough for the user.

2.2.5 Virtual Reality Headsets

Although VR headsets have existed for decades now, only more recently have they become of greater interest to the public, mostly due to the exponential advances in their technology, as well as their use for entertainment purposes. Stereoscopic images are displayed at an optimal distance from the eyes, which combined with tracking sensors that update the scene's viewing direction according to the user's head movement, creates a feeling of presence in the virtual environment. Additional interactions with the environment, such as navigating through the scene or manipulating objects, may require supplementary devices, as not all VR headset models include these functionalities.

Some issues may occur when using these devices that hinder an immersive experience. For one, it is relatively common for users to suffer from motion sickness after using VR headsets for prolonged periods of time, which can be caused by visual overload, or lag between head movement and update in the scene representation, as observed in [HR92]. Furthermore, the weight of the headset and occasional impact of wires and cables against the user are signals belonging to the physical reality that hinder the immersive experience, as described in [SW97].

Nevertheless, VR headsets are the most favorable choice among interaction devices, for both computers and mobile phones, and virtual environment creators would benefit from making such environments compatible with the most common input devices, from keyboard to an HTC Vive, to ensure that nearly all users can experience their work. In table 2.1, a list of current VR headsets is presented, to illustrate the functionalities of these devices.

Table 2.1: Table of currently popular VR Headsets.

VR Headset Model	**Features**			
	Compatibility	**Resolution**	**FoV Angle**	**Additional Controls**
Google Cardboard	Smartphones	Same as device	90°	Side button for screen press
Samsung Gear VR	Smartphones	Same as device	101°	Side control pad or BT controller
Oculus Rift	PC, Xbox One	1080 x 1200	110°	External Controllers
HTC Vive	PC	1080 x 1200	110°	Controllers included
Sony PlayStation VR	PC and Playstation 4	960 x 1080	110°	PS4 Controller, Playstation Move

2.2.6 Body-Tracking

When a VR headset only provides control over the viewing direction, it is necessary to use a supplementary device for any other type of interaction. As seen in the previous section, some headsets include some buttons for this purpose, but when it is not sufficient for more complex interactions, other devices, such as keyboard and mouse or gamepads as mentioned before, are commonly used. An alternative that contributes to the feeling of presence in the environment is a body tracking device. These can be full body trackers, for instance the Microsoft Kinect, or for a specific body part, usually the arms or hands, such as the Wii remote or Leap Motion. To enhance the immersive experience, it is important that the physical movement and the respective virtual movement are synchronous, and a more realistic representation of the virtual body part also contributes to the experience [KGS12] [AHTL16].

There are a few difficulties when attempting to efficiently track a body's movements. One common problem occurs when there is occlusion of body parts, which leads to incorrect reading and representation of the movements with most tracking devices. Another issue is relative to the manipulation of objects in the environment, particularly those out-of-range, and when the available physical area around the user is limited. One existing solution is to use the hand to point towards the object, casting a ray, allowing for interaction with an object intersected by the ray [BH97]. Regarding the limited space issue, in 2016, a VR interaction system was proposed that applies a warped mapping between the real and virtual hand movement, which creates the illusion of a smaller or larger virtual movement [CNdSR16]. However, not all VEs require the highest accuracy and detail in the interaction and representation of the body movements, as they may have simpler interfaces, and the necessary commands and movements are less in number and more limited.

2.3 360-degree Videos in Virtual Environments

The growing interest in VR and immersive systems sparked an interest on the use of 360° videos, as these have certain characteristics that have been proven to heighten the immersion of an experience.

The dimensions over which immersion varies have been studied and presented across many works through the years, and it is essential to understand them to create a virtual experience. Steuer discusses **vividness**, which refers to the richness and quality of the information, whether visual or auditory, transmitted to the user, and **interactivity**, referring to the extent to which the user is able to interact and modify the contents of the virtual environment [Ste92]. Slater and Wilbur later on add to these concepts three others, **inclusiveness**, that is, the ability to isolate the user from sensory information from the physical world, **extensiveness**, referring to the number of senses being involved in the experience, and **surround effect**, which refers to how panoramic the sensory data are, while also emphasizing the need for a **virtual body** in the VE representing the user's physical body, and a correct **match** between the user's body movements and its representation in the VE [SW97].

Observing again the case of 360° videos, they have the ability to visually surround the audience in a scene, which can be complemented by using spatial audio [Rum12]. The loss of viewing boundaries, together with auditory and visual cues, such as sound and lighting effects, may further isolate the audience from the physical reality. Although there is not a virtual body representation, there is still a match between a user's movements and changes in the scene, since head movements change the viewing direction. Regarding vividness, seeing as 360° videos present real images and sounds, the quality of sensorial information is greatly similar to that received from the physical world, even more so when it is stereoscopic. These videos can also be used in the context of interactive storytelling, where the user makes decisions that influence the outcome of the story, such as the project developed by Kwiatek and Woolner in 2010, consisting of a 360° interactive heritage story of a wedding that had several points in the narrative where the user had to choose between two options, which would lead eventually to a different ending [KW10].

It becomes clear that 360° videos are capable of creating a more immersive experience, when comparing to traditional videos, and will be able to induce a more intense emotional response [VTM10] and potentially lead to higher focus on the narrative. The process of creating a film experience with these videos is more challenging, however. Filmmakers no longer have control over what the audience looks at during the film, which means it is not possible to frame a scene, cut between different camera angles or even apply zoom to focus on a conducting point of the narrative. There are various options to tackle this issue: **block** the movement or **redirect** the user's viewpoint, **pause** the scene while the user is not looking in the desired direction, or apply **cues** to guide the user to a certain point. Besides the cues, the other options are more invasive and may hinder the immersion of the system, and even create discomfort and motion sickness, as the image changes unexpectedly unrelated to their head movements. This was noted by a recent study, which compared the use of a visual cue, forced rotation, or no guidance techniques at all, in a panoramic cinematic VR experience, and verified that applying forced rotation had a negative impact on immersion and even made some users feel nauseous [NMH+16].

There is a great variety of cues that can be used for guiding attention, which can be **visual**, by utilizing colour contrast, textures and lighting to distinguish a certain object or entity from its surroundings, although these changes will only have impact if the point of interest is within the user's viewpoint. It is also possible to use the movement of elements in scene to attract attention to certain points, as tested by another study in 2016, which used the actors' actions, such as pointing, waving or even looking in a specific direction, to pull attention to a certain point [BSEW16]. This same study observed that combining visual and spatial **audio** cues has a stronger impact in guiding attention, as these do not depend on the direction of the user's gaze.

All these elements, together with dialogue, the arrangement of objects and actors in the scene, and other parts of filmmaking, are planned before actual filming, using storyboards, pre-visualization models, and other planning tools for it. This phase is also different in the case of 360° videos and cinematic VR, because the traditional tools are restrictive and inconvenient for collaborative work on panoramic scenes. In 2016, a VR planning tool prototype was developed that allowed users to sketch and visualize those sketches with HMD to more accurately grasp the sketched scene [HDC+16].

360° videos are, thus, an effective type of content for the creation of immersive experiences, in which the viewers are potentially more engaged with the events of the video, and its overall narrative. To better understand how these contents are created, the following section describes the technical aspects of filming, editing, and presenting immersive videos.

2.4 Technical aspects of 360° video

To create media content with 360° videos, not only does the creative process need to adapt, as reviewed in the previous section, but also the technical process. In the following sections, the tools and software used during the three main phases of creating a video - filming, editing and viewing - will be reviewed.

2.4.1 Filming

As mentioned before, the rise in popularity of VR and VE has increased the interest in 360° videos, and so, camera manufacturers have been investing in the development of higher quality, and yet affordable, 360° cameras.

There are two main types of 360° cameras, monoscopic and stereoscopic cameras. A monoscopic camera, which is the most commonly found, may use one or more embedded lenses, each pointed at a different angle, with a slight overlap in their fields of view, so that the entire area around the camera is covered. In the case of stereoscopic cameras, a second lens is added to each angle, with a slight distance between both lenses in a field of vision, adding depth to the scene, which is the technique behind 3D movies, as presented in figure 2.2.

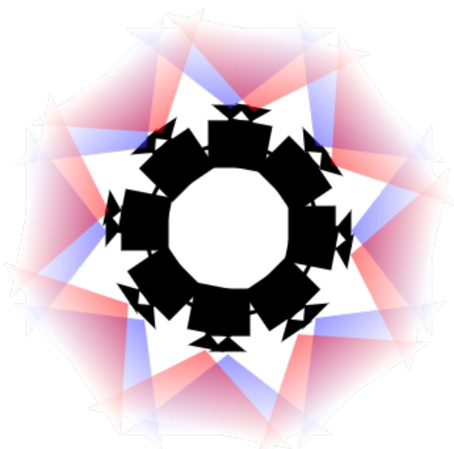


Figure 2.2: Representation of 360° Stereoscopic filming.

After recording, the videos filmed by each lens undergo a process called stitching, which aligns and merges the videos together, producing a seamless large panoramic video. Several algorithms have been developed throughout the years, all of which follow four generic steps [Sze06]: **(1)** define a motion model that maps pixel coordinates between images, **(2)** align the images using pixel-to-pixel matching or feature detection, **(3)** minimize differences between images caused by parallax, motion, light exposure, among others, and **(4)** apply all retouches and project the obtained image in a specified layout, which can be rectilinear, spherical or cylindrical for example.

a) 360° video still image, unstitched. Filmed with Samsung Gear 360 2017.

b) 360° video still image, after stitching with Gear 360 ActionDirector software.

Figure 2.3: Example of a 360° video before and after stitching.

Some cameras perform the stitching automatically, while others require that the user uses stitching software for it. The map projection used is relevant, the most common being the equirectangular format, as the video editor to be used afterwards may not be compatible with it.

While monoscopic cameras can be more accessible to consumers, as well as easier to use, the 3D aspect of videos filmed with stereoscopic cameras enhances the immersive experience [JZDH17]. On the other hand, videos filmed by the latter usually present lower resolution and have more issues during the stitching phase, as most flaws in 360-degree videos caused by parallax, variation in exposure or motion in the scene will be magnified in stereoscopic videos, increasing the level of discomfort for the user. This can be managed with some control over the scene by, for instance, strategically positioning any actors and avoiding movement in the areas of overlap, which is possible for directed films, but not so much for sports events or documentaries [Row15].

Recently, Huang et al. [JZDH17] proposed an approach that allows for the creation of 360° videos with full 6 DOF from monoscopic videos, which only allow for rotational motion, where all new views are synthesized using structure-from-motion techniques to define the scene's 3D geometry and warp the content accordingly.

Currently, there is a variety of 360-degree cameras available in the market, with a wide range of prices, as well as features, to better suit the needs of the content creators. In the following table, a list of 360-degree cameras is presented to better demonstrate the diversity in this market.

2.4.2 Editing

There are quite a few differences in the overall process of creating 360° comparing to normal videos, however for the editing phase, this is not necessarily true, as 360° videos can be edited with mostly any video editor, such as Adobe Premiere, After Effects, and Final Cut Pro, for example. This is possible because these videos are still stored in a 2D format, compatible with common editors, however any operations that are perspective or lens dependent are more challenging with 360° videos. Before editing, the different video angles must be stitched together, as mentioned in the previous section, which some cameras do automatically or at least provide software for it. After the video is stitched in the required format, editing processes such as color correction,

Table 2.2: Table of 360-degree Cameras.

Camera Model	Features			
	Resolution	**FoV Angle**	**3D**	**Stitching**
Samsung Gear 360	4K	360° x 360°	No	Offers Software
Ricoh Theta S	1080p	360° x 360°	No	Offers Software
360Fly 4K	2880x2880	360° x 240°	No	1 Lens, no stitching required
Nokia OZO	2k	360° x 180°	Yes	Offers Software
GoPro Omni	8k	360° x 360°	Yes	Offers Software
Google JUMP Odyssey	4k	360° x 180°	Yes	Performed by Google's VR system

compositing, adding titles and other media on top of the video is the same as with traditional videos. The editor interface allows to pan around the video and apply the desired traditional effects, as well as defining the video resolution, frame rate, and codec, among others, which is important so as to ensure that the video will be compatible with certain devices and platforms. The final edited file can be saved with metadata that identifies the video as being a 360° video (sometimes called VR instead), and monoscopic or stereoscopic, which is vital for most video players to correctly display it.

Although the video editors mentioned before support 360° video editing, it can be cumbersome to interact with the panoramic image to place effects, and to get the complete idea of how the video looks while editing. The video editor subproject of Project AV360, which has been developed as a MSc Dissertation named "Rendering and Editing Tools for Interactive 360° Video", tackles this issue by presenting a solution that allows to create dynamic annotations in a 360° video using VR devices. These annotations are not directly placed on the video file, obtaining a altered file after editing, but instead place the necessary information in a separate metadata file, which has to be interpreted later by a compatible visualizer, such as the one being described in this document.

The AV360 Project and the immersive editor will be detailed further in chapter 3.2.

2.4.3 Viewing

The final phase is the presentation of the created video to the audience, which is the focus of the work presented in this document. Presenting a film requires choosing the appropriate medium according to the needs of the video experience in terms of interaction and visualization. The focus of this section will be on video players compatible with this type of media content and common household displays, such as a computer screen or a mobile phone, and also VR headsets.

User interaction and **experience** are the main concerns when developing and choosing a visualizer, which depends heavily on its **interface**. Common video player interfaces contain controls for playing and pausing the video, as well as volume controls. Some more advanced controls may be added, such as the resolution of the screen, subtitles and viewing size.

In 360° video players, these common controls can also be used, adding controls to them for navigation and guiding. To change the viewing direction, the interface may present directional arrows, or be directly controllable using an input device such as the mouse, keyboard or a VR device. This is generally accompanied by a guiding icon, such as a compass, that informs the user of their current viewing direction (which together with other annotations will be discussed in the following section).

Content creators and their audience expect from the visualizer that the viewing experience suffers as little as possible from video freezing or other disrupting artefacts, as these result in disengagement from the user [KS13] and ultimately hinder the immersive experience [SW97]. Bowman [BMT07] studied and presented the more technical factors that influence the levels of immersion, which are the "FOV and FOR[3], display size, display resolution, stereoscopy, head-based rendering, realism of lighting, frame rate, refresh rate". Although some of these features can be defined in the filming and editing phase, it is relevant to mention at the viewing phase, since some video players do not allow for certain display sizes and resolutions, for example. It is important that the chosen viewing medium does not deteriorate the produced film.

There are several web content hosting platforms currently able to play 360° videos. The most prominent platform, Youtube, has allowed user to upload their 360° videos since 2015, and it is compatible with Google Cardboards and Daydream, Playstation VR, as well as Oculus Rift and HTC Vive, although limited for now. The interface for these videos is similar to the standard Youtube video interface, with added four-directional arrow buttons, to control the gazing direction, which can also be done by dragging the image with the mouse. Besides the traditional video requirements regarding container format and resolution, among others, it also requires certain metadata to identify it as a 360° video. It also allows for the use of spatial audio in the video. Regarding Facebook, the upload of these videos is similar to traditional videos, not requiring the required metadata to be already placed in the video file, as it is possible to add such data during the upload process. The interface contains the controls for regular videos, with the addition of a small interactive compass which indicates the original video direction and the current orientation, and when clicked, moves to the original orientation, as defined by the upload user. Vimeo offers a similar video player interface to Facebook, and it also allows for larger files and resolution up to 8k, however it does not support spatial audio.

It is also possible to instead integrate the video player in a personal website or mobile app. Several companies, such as Wowza Media Systems [4], Bitmovin [5], JW Player [6], or Delight VR [7], provide an HTML5 video player for 360° videos, with an API to create more custom interfaces. These services also include viewer analytics, such as number of views, viewer heatmap, average delay time, and other relevant statistics, and additionally, Wowza and JW Player also provide a

[3] FOV - Field-of-view; FOR - Field of Regard

[4] WOWZA Media Systems - VR & 360° Streaming. https:www.wowza.comsolutionsstreaming-typesvirtual-reality-and-360-degree-streaming

[5] Bitmovin - HTML5 Player for Adaptive Streaming. https:bitmovin.comhtml5-player#VR

[6] JW Player - 360° Video & VR. https:www.jwplayer.comvideo-solutions360-video-vr

[7] Delight VR - The Virtual Reality Player. https:delight-vr.com

video hosting service.

Most of these streaming services use either HTTP Live Streaming (HLS) or Dynamic Adaptive Streaming over HTTP (MPEG-DASH), as the required bandwidth for 360° videos is generally higher than regular videos, meaning that lagging and freezing issues may be more frequent. As such, an adaptive streaming system based on MPEG-DASH was recently proposed, which prioritizes the viewport, providing higher resolution on that section, with the purpose of reducing overall bandwidth [HS].

Table 2.3: Table of 360-degree Web Video Players.

Platform	**Features**					
	Resolution	**FOV**	**Aspect Ratio**	**Frame Rate**	**Stereo**	**3D Audio**
Youtube	8K	Video's FOV	16:9	60 (max)	Yes	No
Facebook	4K	Video's FOV	2:1	60 (max)	No	Yes
Vimeo	4K	custom, 90° max.	2:1	60 (max)	Yes	No
Wowza, Bitmovin, JW Player, DelightVR	4K	Custom	Custom	60 (max)	Yes	No

2.5 Information Overlay

When creating and using content, whether it is an image, a book, or a video, to transmit information, it is sometimes necessary to supplement what is directly observable by the user with additional information. In this section, the methods of conveying additional information in different contents will be reviewed, in order to then identify their general characteristics and purposes, as well as common issues. Observing the applications of visual annotations in different contexts, from maps to virtual environments, will provide a fuller perspective of how these visual objects may be applied and designed for 360° videos.

2.5.1 Labels in Images and Videos

Cartography, the ancient practice of map-making, has presented, since the very beginning, representations of areas of land to which labels have been added, providing additional information and context. Map labeling, refers to the act of placing text to identify map features, such as routes, rivers, buildings, and others, which may be present in some but omitted in other maps, depending on the type and purpose of the map. In figure 2.4, a portion of the map of Peutinger is presented, which is a comprehensive seven meters long map representing the road network of the Roman

Empire, Europe and parts of Africa and Asia, with illustration of rivers, forests, different types of buildings, and even some biblical illustrations [Tal10b].

Figure 2.4: Section of Peutinger Map, online resource [Tal10a].

As the general knowledge of the surrounding area increased throughout the years, maps became more complex with higher density of map features, and as a consequence, the process of labeling became more complicated and the most time-consuming phase. Nowadays, automatic label placement algorithms have been studied and developed so as to reduce or remove the need for manual labeling [KB08].

Textual labels are not the only way to provide more information in a map, as pictograms can be used to transmit data more efficiently than text. Pictograms, also referred to as icons, are visual representations of an object or entity, which are often more compact than their textual descriptions and independent of language, and, if well designed, will be able to convey its meaning to the reader [MJ93]. Some icons are present in almost all types of maps, such as the compass rose or a scale, while others have a more specific use, for instance arrows indicating the street's direction in road maps, or models of prominent buildings in tourist maps. Commonly, icons are accompanied by a key, which provides insight on the meaning of the icons, as can be seen in the map of Paris presented in figure 2.6. Certain icons become so widespread that they eventually become internationally recognized as standard, such as the pictographs present in ISO 7001 [ISO].

Similarly to maps, **illustrations** also often require captions or labels to provide more information and explain what is being represented. This is particularly true for technical, scientific or educational illustrations, which accompany expository text in order to visually explain complex processes or phenomena being described, improving the effectiveness in the learning and memorizing process [LL82], even in the cases of learners unfamiliar with the specific subject [MSBM95]. These informational illustrations come in many forms, presenting an exploded-view of an object and its parts, exemplified in figure 2.7 a) or a cutaway of an object which shows its internal contents and architecture, also exemplified in figure 2.7 b). In the digital era, these illus-

a) ISO graphical symbol for Emergency Exit.

b) ISO graphical symbol for Electricity Warning.

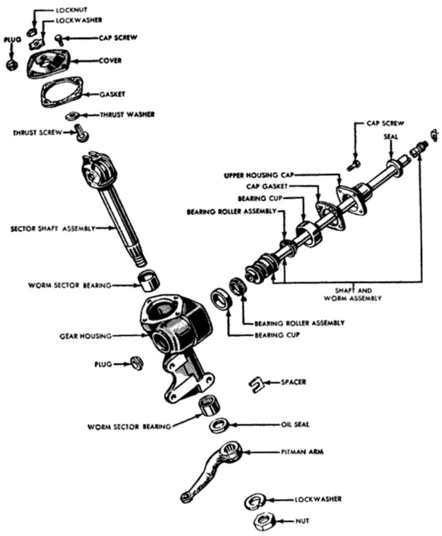
Figure 2.5: Examples of ISO symbols, commonly used and seen in public spaces.

Figure 2.6: Tourist map of Paris, with a description box for the used icons [Par].

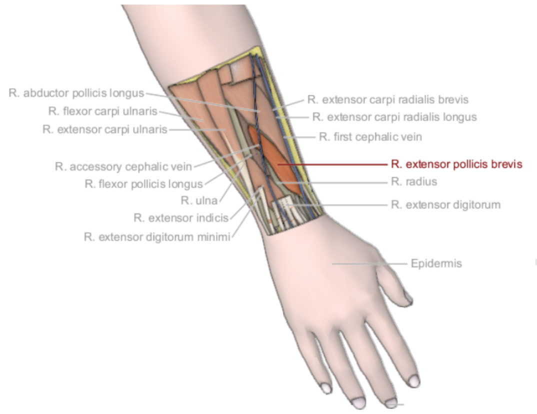
trations are more often created on a computer and can be interactive and animated, allowing the user to rotate and view the object from different angles, preview behaviors through an animation, click, drag, or enlarge/decrease a specific component.

Annotations are still present, however their design must adapt to the interactive and flexible layout of the illustrations, which means they must be dynamic in order to follow the movements of the segment they are associated with, while also taking into consideration any problems of occlusion regarding the illustration itself or other annotations [LAS04]. They may also be interactive, allowing, as an example, for them to be expanded to read more information, or collapsed when they are not necessary, to make the display less crowded, as presented by Sonnet et. al [SCS04].

Moving forward from images to **videos**, this media content has contained textual information since the beginning, as silent movies required text to present the characters' dialogue, and were vital for the audience to understand the plotline. Text was also used, and still is, to present a film's title, producers, cast and film crew, actors, and any other information regarding the producing of the film relevant for the credits. With the appearance of audio in videos, the dialogue texts were replaced by subtitles, for films of foreign origin, or for the deaf or hard-of-hearing audience. Subtitles, unlike the opening and closing credits, appear throughout the entire film, so it is vital that they occupy a minimum amount of screen area, so that they obstruct the scene as little as possible, while at the same time ensuring its readability. Proper subtitling is dependent on the used font,



a) *Exploded view of a manual steering gear for a automobile, from a repair manual of 1956.* Ferguson, "Engineering and the Mind's Eye", 89 [Fer94]



b) *Illustrations generated using our automatic cutaway generation interface.* Li et al., "Interactive cutaway illustrations of complex 3D models", 9 [LRA+07]

Figure 2.7: Two types of technical annotations.

size and color of the text, synchronicity with the dialogues, and even line breaks [Per08].

Videos in a more informational context, such as documentaries, use a larger assortment of media, not only text but also images or graphs, superimposed on the original video, to complement or emphasize the information that the video and narration transmit. Most of these overlaying media are edited directly into the video, except for the subtitles, which are defined in a separate standard file, and then read and placed over the video by the video player. Some video players may allow for other annotations to be placed over the video during its runtime, as it is possible on Youtube, where content creators can edit their videos by placing titles, speech bubbles, notes, links to other videos, and others.

2.5.2 Information in Interactive Displays

Certain human tasks and activities require and involve a great amount of information, such as piloting a plane, where there is a lot of data being displayed to the pilot in different screens, as part of a large and complex cockpit panel. This display of information forces the pilot to alternate from looking forwards to looking down to the panel, and so, **head-up displays** were developed, that present several flight data in a transparent and non-invasive display, placed in front of the pilot, exemplified in figure 2.8 a). Since the 1950s, when they were first developed, HUD technology has evolved greatly, and they have spread from military to commercial aircrafts, and cars as well.

The HUD systems predate the advent of the field of **Augmented Reality**, which refers to systems that combine three-dimensional virtual with real objects, are interactive in real-time [ABF+01].

HUDs are, thus, considered an AR device, in addition to other transparent displays, such as eyeglasses, but the great majority of commercial AR applications are operated with mobile devices, using its camera for scene analysis, and its gyroscopic and GPS sensor for position and orientation, among others [NCJC15]. AR has professional applications, where it is used to provide vital information for performing certain tasks, whether it is repairing a complex system, navigating, or medical training, but currently it is also widely used for entertainment and other casual purposes, such as games or tourism.

Errors may occur while using AR systems, which can be due to the hardware or software conditions. Flawed sensors or displays will create distortion and misalignments, which after noting the different uses of AR, could have serious consequences, so it is vital to check the devices' conditions and perform recalibrations where necessary. The overlay of virtual objects on the scene are also a highly studied problem in AR systems, referred to as view management. The number of objects being added, their spatial positioning, their appearance, and even interaction methods are all vital aspects to take into consideration in order to create an effective augmented reality experience.

This method of information overlay has been widely explored in a particular entertainment context: **games**. A HUD is now a common part of the game's interface, and contains a set of indicators relevant for the gameplay, such the familiar health bar, or a minimap presenting the avatar's position and important surrounding entities, with the purpose of helping the player progress in the game. Although certain elements of a game's HUD have a standard design, so as to be quickly recognized by new players, their appearance is commonly adapted to the overall atmosphere of the game.

a) Photograph taken by a pilot on the VFA-151 of the HUD of a F/A-18C [Com05].

b) Dead Space 2 diegetic HUD, displaying a health (blue light indicator) and statis (an in-game power, represented by the circular yellow light indicator) meter in the character's back. [Wik16].

Figure 2.8: Uses of HUDs in different contexts.

HUD element's design and position on the screen has an influence on the player's experience

[Bab12], which also varies according to the game's genre [CI16]. The influence of these non-diegetic elements on the immersion and sense of presence of the player has also been studied and discussed, some arguing that the HUD reminds players that they are interacting with a virtual environment as outsiders and not as a participant [Wil06], and that they hinder the immersion particularly for more experienced players [ICK+15], while others defend that with careful design and relevant information, HUDs are desirable and enhance the usability of the game [Jør12]. An example of a seamlessly integrated diegetic HUD is presented in figure **??**.

2.5.3 Annotations in Panoramic Media

Panoramic or 360° images and videos place the individual in the center of the scene, which can only observe a segment of the scene at a time. These types of media can also be complemented by overlaying media to convey more information and also interact with the image. Considering the example of Google Maps Street View interface, it contains labels of the street names and a compass, which are common elements in maps, along with navigation arrows, and a minimap that pinpoints the user's position in the 2D map. Panoramic virtual tours present similar elements, such as the Smithsonian Virtual Tour, with additional information and images of the works in display.

360° videos include elements also present in traditional videos, such as the opening credits, subtitles and end credits, as well as elements from annotated illustrations. Most of these elements remain static in a certain section of the scene, which means that they may not be seen by the user in case their viewing direction is not directed to its position. Other elements can be fixed in front of the user's field of view, for instance the subtitles, or moving, usually following the movement of an object in scene [MMJ+16].

These visual objects may also be interactive, which again is the case of Google Maps for 360° images, but it may also be used in 360° videos, as Neng and Chambel described [NC10], creating immersive hypervideo content. This allows users to explore the video's theme even further, by accessing additional information provided by the hyperlinks that are placed in the video.

2.5.4 Definition of Video Annotations

From the diverse examples of overlaying media presented previously, it is possible to define and characterize these elements, which will be henceforth referred to as **annotations**. Visual annotations are media content added to a video, with a textual or graphical representation, subtitles being a common example of the first, and a flag over a point of interest as an for the second type. They can be used to:

- **Guide** the user and support **navigation**
- Attract **attention**,
- Provide additional **information**.

Annotations may be interactive, in this way enhancing its function. To exemplify, a label identifying a building may be selected to show further information of the building, and an arrow, when selected, may move the field of view to focus on the point of interest. For this, it must be defined how to interact with the annotation - through a key press, mouse hover, and so on - as well as its behavior after the event - change in appearance, change of viewing direction, show information, among many others. In 2006, an annotation layout algorithm was presented, regarding the case of interactive and dynamic annotations, where internal and external labels adapt to the movement of their anchor point, minimizing the travel distance as well as anchor distance, while also avoiding occupying used space [GHS06]. Currently, the ability to read and render annotations on top of 360° videos is still a scarce feature in most viewing services, but there are at least two web CMS cases, TransportVR [8] and Viar360 [9], which allow content creators to place interactive elements for video navigation or providing information. However, elements in these services can only have a fixed position throughout the video, even though other types of positioning are of importance to serve other purposes, which is discussed in this following subsection.

2.5.5 Positioning of Annotations

Immersive videos are represented in a three-dimensional environment, projected on the inside of a sphere or a cylinder, for example, and where the user is placed at the center. Objects, such as the visual annotations, that are intended to appear over the video's image will be placed between the video's object and the user's point-of-view. There are three main ways of positioning annotations in a video:

- **Persistent**: The annotation is locked in a point in front of the user's FOV, presented in figure 2.9 a);
- **Fixed**: The annotation is associated and placed in a point in the video's scene, shown in figure 2.9 b);
- **Floating**: The annotation is associated with a point in the video's scene, but its position floats into the user's FOV, as exemplified in figure 2.9 c).

Referring again to subtitles as an example of annotations, they are commonly placed orthogonally facing the user's field of view, appearing static to the viewer. Brown [BTP+17] presented four types of subtitle behavior in terms of positioning in 360° videos, taking into consideration that subtitles are important for the viewing experience, and their position must obstruct and disrupt as little as possible. In 2D displays, their position is **persistent**, which is emulated in 3D displays by defining the position of the annotation according to the user's current viewing direction, maintaining their relative distance and orientation. Other types of annotations may be instead associated to an object in scene, and either **fixed** near the object, or **floating** into the user's field of view, with a line connecting it to the point of origin, this way guiding the user to the object. Maass and Döllner (2008) [MD08] present an annotation technique of three-dimensional objects in a virtual environment that considers the geometry of the annotated object, as well as the viewer's viewing

[8] Transport VR. https://www.transportvr.com/
[9] http://www.viar360.com/media-virtual-reality/

a) Example of a **persistent** annotation (e.g. subtitles).

b) Example of a **fixed** annotation associated to a building.

c) Example of **floating** annotations referring to POIs out of view.

Figure 2.9: Three types of annotation positioning.

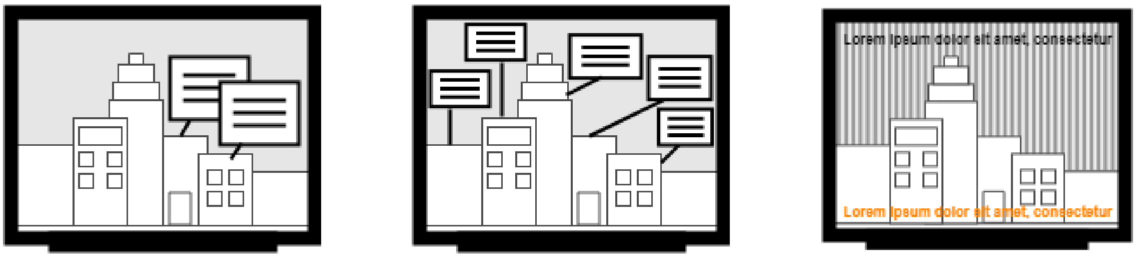
direction. These different methods of positioning annotations are directly related to the purpose of such annotation, according to whether the information should always be visible for the user (static relative to the FOV), the information should be seen only when faced by the user (static independent from the FOV), or when there is a need to indicate a point of interest containing relevant information. There are also cases where the point of interest is moving throughout the video, which requires a dynamic annotation that follows its movement.

2.5.6 Issues with Video Annotations

From the study of the different types of information overlay, such as labels and annotations, it is possible to determine issues that may emerge when placing them in videos:

- **Occlusion** - When an annotation is covered by others, observable in figure 2.10 a);
- **Visual Overload** - When annotations cover too much scene space, as shown in figure 2.10 b);
- **Readability** - When the information on the annotation is difficult to distinguish, exemplified in figure 2.10 c);
- **Out-of-view** - Specifically for panoramic images and videos, refers to the location of annotations outside of the current viewpoint.

Particularly a concern with textual annotations, the **readability** of an annotation is affected by its own attributes as well as the characteristics of the video's scene. The text's font type, color, size, weight influences its readability, and it should adapt to the background's appearance, structures, color and textures. Higher contrast between text color and background color and simple types of font are two steps to take, but are not always enough. Jankowski et. al [JSI+10] tested several annotation designs and observed that text in billboards (text boxes), particularly those with a light foreground over a dark background, had the the best results in terms of readability. Leykin and Tuceryan recognize the relevance of a label's background texture on its readability, and present



a) Example of an annotation covering another.

b) Example of visually cluttered scene.

c) Unreadable annotations, due to color and scene texture.

Figure 2.10: Three types of issues with visualization of annotations.

a system that classifies sections of an image as readable or not [LT04]. Later in 2012, a view-management technique is proposed that would place labels while taking into consideration the video content visual characteristics, such as geometrical structures and background color [GLT+]. These and other similar works present annotations that are flexible in nature, in terms of its positioning as well as its appearance (adaptive foreground/background color, transparency, size, font weight, and so on).

In a scene with many points of interest, annotation **occlusion** may occur due to relative positions of all annotations in the scene, which has been the subject of studies in the past: Azuma and Furmank, for example, tested some automatic view management algorithms regarding the spatial layout of annotations relative to others and their point of interest, in order to avoid occlusion [AF03]. It is also important to avoid obscuring important elements of the scene. Occlusion is not the only concern in highly annotated scenes, as a greater number of superimposed elements will most likely lead to information or cognitive **overload**. Such a condition may lead to frustration, inability to focus and even decrease in task efficiency, no matter the context of the information overload [MdPV+06]. The amount of visual information provided by the annotations must be maintained to a minimum, which is possible by using more compact design, conceal annotations that may be irrelevant at a certain point, or make use of the interactivity of the system, keeping them closed until the user demonstrates interest in them, by clicking or hovering, for example. Clayvision is a project developed with the concern of cognitive overload in mind, which presents an alternative to labels used in outdoors AR, by manipulating the geometry of the city buildings' virtual representation in order to make certain points of interest more evident than its surroundings [TP12].

There is also the case of **out-of-view annotations**, which occurs in 360° images and videos, where only a section of the content is being displayed at a time. The question that arises is of how to inform the user of their existence, without being distracting or disturbing. Neng and Chambel refer to this issue in 360° as link awareness in a study, and presented a solution that placed indicators in the edges of the scene, informing the user of a certain annotation's (called hotspots) proximity and position relative to the current viewing angle. A minimap representing the video and indicating the user's viewing direction, alike the example in Google Street View, could also be used to point out annotations that are currently active. Additionally, considering what was mentioned previously regarding media cues to guide the audience through a video's narrative, one can infer that similar results can be obtained by attaching a spatial sound cue to an annotation, that will attract the user's attention to the annotation's current position. It is, of course, necessary to consider, when applying these or other methods of annotation awareness, their potential impact in the viewing experience.

2.6 Summary

As it was seen in this chapter, virtual environments are appealing for their ability to make users feel like they are present in them. This sense of presence has been seen to heighten one's emotions and assist in developing and training certain skills, as users are engaged in the action, and think and act similar to their normal behavior in the physical world. 360° videos can be used to create more immersive videos, with the goal of engaging the audience more in the narrative. To create this type of content requires different tools and process when compared with traditional videos, and filmmakers must adapt to the transfer of camera control from them to the audience, by taking manipulating conditions in the scene, such as lighting and contrasting colors, and adding visual and audio cues to attract attention to certain points where they would otherwise be able to frame or zoom in.

Films with a more informational purpose, such as documentaries and news videos, may be superimposed by non-diegetic elements to provide more information or to guide the viewer's attention to a certain point of interest. By analyzing cases of information overlay in other areas, it is not only possible to determine different types of annotations that can be used in 360° videos, but also what steps can be taken to ensure an efficient and unobtrusive viewing experience. Defining these annotations in a separate file, instead of directly editing them into the video, will allow for easier additions, changes or even removals after production.

The visualizer chosen to display these annotated 360° videos must support the minimum technical requirements, such as resolution and frame rate, so that it does not hinder the desired experience. Compatibility with different interaction devices is valuable, so as to make it accessible to a larger audience. Content creators, from filmmakers to journalists, are aware that for more extensive audience reach, placing their work online is very beneficial. As such, a web video player compatible with this type of videos is necessary, expectedly supported by most desktop and mobile browsers. Looking at the most popular web video player services, it is evident that although most reach the minimum requirements, being able to render different type of annotations over the video is still a limited feature, especially when considering dynamic annotations.

Therefore, the aim of the work described in this document is to develop a web visualizer for 360° that is able to display different types of annotation, used to provide additional information or identify points of interest in the video, for example. The following chapter presents the requirements and proposed architecture for such system, with the purpose of being compatible with different interaction devices.

Chapter 3

Proposed Architecture for the Immersive Visualizer

Having researched on the state of the art relative to 360° video content creation, as well as the use of information overlay objects, or annotations, in different contexts, this chapter details the planning phase of the development of the visualizer. It covers the requirements and architecture of the visualizer, in order to provide answers to the research questions established in chapter 1.2.

Firstly, the context of the visualizer within the AV360 project is elaborated in the section 3.1, followed by the definition of some key concepts regarding the annotations to be implemented in the visualizer in section 3.3. The section 3.4 afterwards presents a list of system requirements, pertinent to the AV360 Project and the research questions of this dissertation. Lastly, a system structure and its interfaces are proposed in section 3.5, taking into consideration the requirements established in the previous section.

3.1 General Description

The application of visual annotations in 360° videos may potentially enhance the video's content, with the purpose of tackling issues that arise with this type of video, already established in section 2.3. From the review of annotations' characteristics and possible uses in section 2.5, it was possible to define a list of varied annotations, with different purposes. As one of the dissertation's research questions is related to the impact that these annotations have on a user's video experience, it was important to develop a 360° visualizer that would be able to display immersive videos with dynamic annotations. A potential user should be able to choose a video to watch, using their computer, mobile device, or even a VR headset. Throughout the video, the user should be able to control their field-of-view on the video, and the implemented annotations should support their experience, either by providing important information on the video's subject or calling attention to an important area in the scene. It is vital that the visualizer's performance, as well as the annotations' appearance and behavior, does not have a negative effect on the immersion of the experience.

As such, the planning of the annotations and the visualizer, which is described further on in this chapter, should take into consideration these general requirements for a 360° visualizer. For the visualizer developed in the context of this dissertation, it was important to review its context within the process of 360° content creation, which is presented in the following subsection.

3.2 Workflow Context

It was mentioned in chapter 1.1 that the work described in this document is part of a larger project called AV360, which focuses on the research and development of tools for 360° content creation in the context of journalism. Before the implementation of the visualizer described in this document, an immersive editing tool for 360° videos was developed.

In the immersive editor subproject, an initial prototype was developed that utilized masking techniques to mark the position of annotations in a 360° video. With this prototype, a mask layer that identified point of interest would be placed over the video, and in a second phase, text would be associated to those mask points. Later on, a different version of the editor was created, in which a user could place visual objects (annotations) in the 360° scene, using either the traditional keyboard and mouse method, or virtual reality devices - HTC Vive or Oculus Rift [Mei17]. The annotations created with this system are recorded in a separate *.srt* file, with the intention of being used in compatible video players. Figure 3.1 exemplifies this last editor's interface, with its menu visible and an annotation placed on the video. In this figure, the annotation contains the text "Montanha Russa", defined in an HTML file. The white line under the text depicts the annotation's path, using keyframes to define its segments, where a keyframe corresponds to a position in the video's image at a specific runtime.

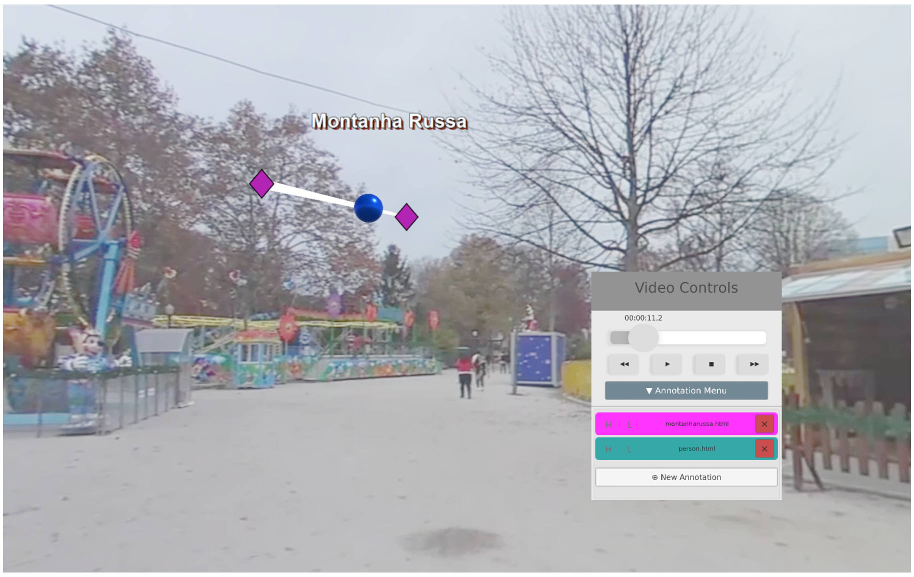


Figure 3.1: Immersive Editor's Interface with mouse and keyboard.

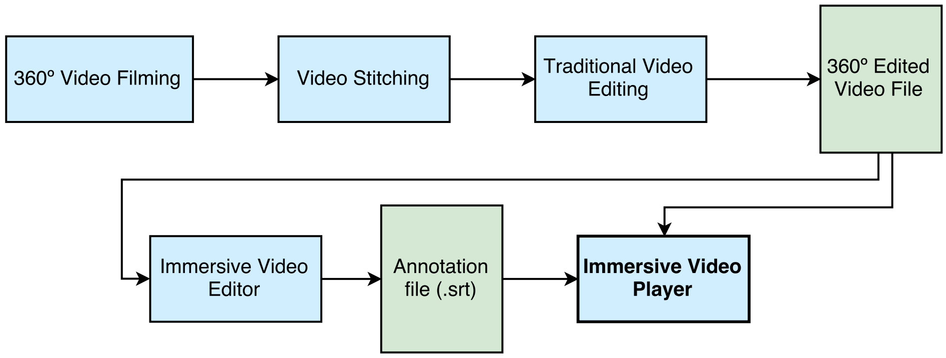


Figure 3.2: Immersive content creation process with project AV360.

As the editor and visualizer are developed in the context of the same project, they are expected to be used collectively, and so, the visualizer described in this document must be compatible with the files that result from the preceding editing process. Once the user finishes annotating a certain video, the files created with the immersive editor are passed to the visualizer, which should be able to read and interpret them, in order to present the video as expected. The Figure 3.2 presents a complete workflow of 360° video content creation, using the two combined systems.

Before defining the visualizer's requirements, it is necessary to delineate the new concepts of annotations that will be introduced in this system, as required from the established research questions, which is presented in the following section.

3.3 Annotations in the Visualizer

In the state of the art review chapter, the section 2.5 analyzes several methods of information overlay objects in different visual contexts, which is directly related to Research Question 1, presented in chapter 1.2, "How to enhance 360 videos with dynamic annotations".

Throughout a 360° video, there may be one or more points of interest at a time in the scene, which can be a building or a moving person, as example. The video may or may not have an underlying narrative that connects the points of interest throughout the video. The content creator may want to attract attention to a certain point of interest, to provide additional information on it, or to guide through the video's narrative. Figure 3.3 presents a possible display of annotations, which are afterwards described, in a 360° video.

For this reason, the annotations implemented in the visualizer must be able to cover all these different circumstances. Four main types of annotation have been identified: **informational annotations**,that have the purpose of providing information to the user, by means of text or images; **directional annotations**, these are visual objects that guide the user to an out-of-view informational annotation; **contextual annotations**, which provide an overview of the relevant annotations' positions in the video's image; and lastly, **narrative annotations** which can be used to identify and guide to the narrative's area of interest, in videos that have a visual path to follow throughout runtime, as planned by the content creators.

Figure 3.3: Proposed display for annotations in the visualizer.

Therefore, adding to the annotation created in the immersive editor exemplified in Figure 3.1, the following annotations are proposed:

Informational Annotations Annotations placed over the video's image, containing text or images pertinent to the video's theme or a point of interest.

- **Subtitles** - Similarly to their use in traditional videos, its purpose is to inform the user of what is being said and heard during the video. This object must be always visible to the user, independently of their field-of-view, so their position must be persistent, exemplified in figure 3.4 a). Subtitles are defined as blocks of text with a start and end time, usually placed at the bottom of the screen so as not to obstruct the view of the scene.

- **Marker** - This is the default annotation defined in the 360° immersive editor, which can be used to provide additional information on the video's topic or a specific point of interest. Its position is fixed, that is, it is associated to an element of the video that can be stationary or moving during the video. For this reason, the marker's position is defined through a list of keyframes, referencing a point in the video at a specific time in the video. A marker's content is defined in an HTML file, which may have a CSS file for further personalization, allowing for a mix of stylized text and images. When a marker has a larger amount of information to present, some may be hidden until the user indicates their intention to read the entirety of marker. Figure **??** represents an example of a marker.

Directional Annotations As informational annotations are only visible when the user is looking in their direction, it is important to have an indicator of their presence, when they are not visible, so that the user does not overlook possibly relevant information contained in the marker.

a) Example of subtitles on a screen.

b) Example of a marker, containing text and an image.

Figure 3.4: Mock-ups for the informational annotations.

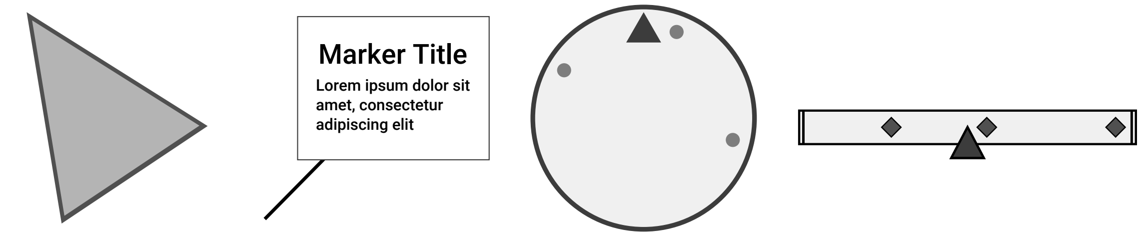
- **Arrows** - A two-dimensional guiding object, as seen in Figure 3.5 a). Its position is floating in the user's FOV, and its position and rotation should make clear how the user must rotate their view to find the marker.
- **Miniature** - An alternative to the arrows, which guide the user to an out-of-view marker, while also providing a hint on the appearance of the marker. Essentially, it is a smaller copy of the marker, using a line connected to the original marker to indicate the path to it, exemplified in Figure 3.5 b).

Contextual Annotations

- **Minimap** - An object that gives general information on the position of all markers currently present in the scene, represented in figure 3.5 c). Similarly to navigation radars, it indicates the user's current viewing direction, and each marker is identified in the minimap's area, giving their relative position to the user. Its position must be persistent, always in view of the user, but placed away from the center of the FOV, so as not to obstruct the view.
- **Horizontal Compass Bar** - This object is an alternative to the minimap, and it consists of a rectangle representative of the 360° video's horizontal perimeter. The center of the bar corresponds to the center of the user's field of view, and markers present in the scene have an indicator placed according to the marker's relative horizontal direction from the user, and rotating the camera moves the indicators to the left or right. The benefit of this object is that it potentially occupies less visual space, as seen in Figure 3.5 d).

Narrative Annotations

- **Vignette** - This effect is based on the concept of vignetting in photography, where a photo's saturation or brightness can be purposely reduced towards its edges, to draw interest to the



a) Example of an arrow guide.

b) Example of a miniature guide.

c) Example of a scene minimap.

d) Example of a compass bar.

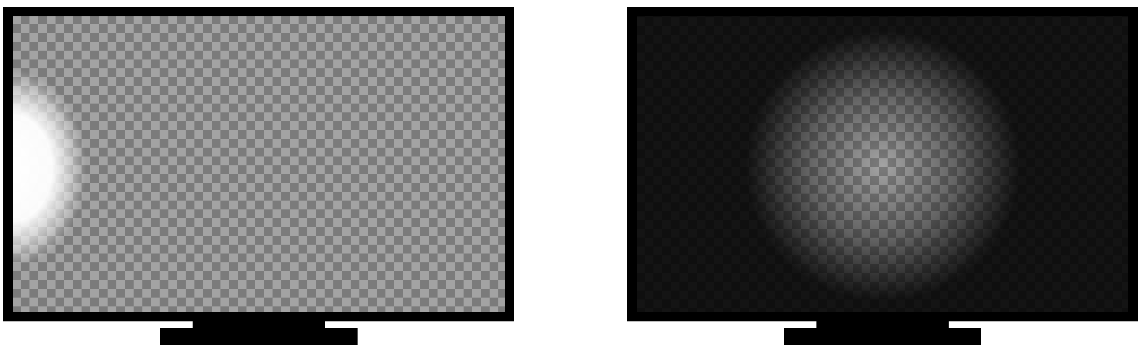
Figure 3.5: Mock-ups for the directional and contextual annotations.

center and to frame the center portion of the photo. Applying this concept on 360° videos, vignetting can be used to draw the user's attention to the current area of interest, while still leaving the remaining area visible, albeit darkened. Figure 3.6 b) demonstrates this effect, centered in the screen.

- **Lateral Lights** - Similarly to the vignette, the lateral lights are used to elicit attention to the current area of interest. While the vignette is permanently visible, the lights only appear when the area of interest is not visible to the user, and they are positioned to the left or right of the user's view to indicate towards which side they should turn, which is depicted in Figure 3.6 a).

- **Viewport** - This object consists of a smaller screen that presents an alternative view of the scene, specifically, the area of interest that is not in view for the user. As its purpose is to display the out-of-view area of interest, it is only used in that context, equally to the lateral lights. Figure 3.7 illustrates the viewport annotation, "A" is the area in the direction of the user, and "B" is the narrative's area of interest, displayed in the viewport object.

Each annotation that is persistent in the camera's view has a specific region which they occupy in



a) Example of a lateral light, applied to the left of the screen.

b) Example of the vignette effect, focusing in the center of the scene.

Figure 3.6: Mock-ups for the narrative annotations.

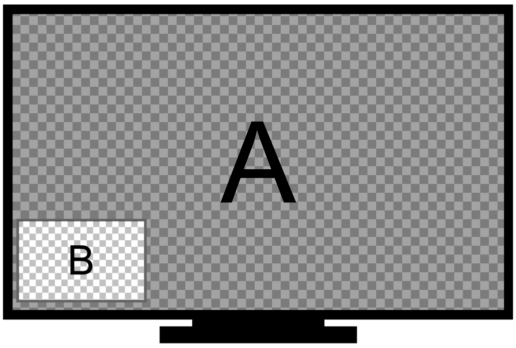


Figure 3.7: Example of the viewport annotation, represented by the rectangle at the lower left corner of the screen.

the camera's frustum, avoiding overlap between different types of annotation. Figure 3.8 demonstrates how the conceived annotations are placed in the visible area: the subtitles are placed, as standard, in the lower region of the screen, which is region B, the floating guides, that is, the arrows and miniatures, as well as the viewport and lateral lights, are placed within region A, closer to the margins, the minimap can be placed at a corner of the screen, such as region C, and the compass is placed on the higher region in region D, as it occupies a large area horizontally.

3.4 System Requirements

To assist in the planning and development of the system, a list of requirements was created, based on the research questions defined in the first chapter, as well as the research on current video players in the second chapter.

The visualizer must first have the basic functionalities of a video player, and be compatible with the most common video formats, such as `.mp4`. The system thus has the following requirements:

- **Basic Video Controls** - The user must be able to perform the basic controls of a 360° video player, which are play or pause the video, and adapt the audio volume.

- **Playing 360° videos** - The user must be able to watch the video in a 360° view and control the camera's rotation using their chosen device.

- **Device Compatibility** - The user must be able to view the video correctly using either the **(1)** computer with keyboard and mouse, **(2)** computer with VR headset (Oculus Rift or HTC Vive), **(3)** mobile device without VR headset, and **(4)** mobile with VR headset (Google Cardboard or Samsung Gear VR).

Because the visualizer is part of a larger project, it must be compatible with the resources obtained from the other tools, namely the immersive editor. As such, the visualizer has the following requirements regarding annotations:

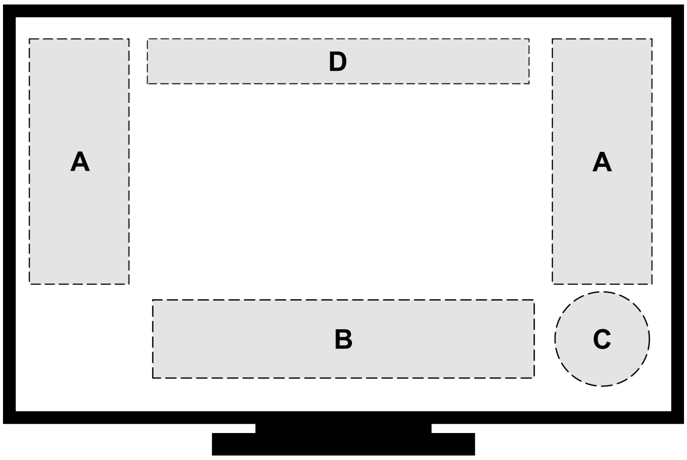


Figure 3.8: Proposed layout for the persistent and floating annotations. Area A is for the marker guides, B is for the subtitles, C is for the minimap, and D for the compass bar.

- **Read Annotation File** - The system must be able to read and render the annotation's data on the *.srt* file.
- **Dynamic Annotations** - The system must be able to correctly render an annotation's movement in the video's scene.
- **Annotation Interaction** - The user must be able to interact with annotations and obtain some feedback.

Some additional requirements can be defined for the system, that were not a focus of the work on this dissertation, but are nevertheless relevant for future work. These are mostly related to the content creator's use of the system, as follows:

- **Video Upload** - The user must be able to upload a 360° video to the system, together with the annotation files.
- **Edit Video Viewing Settings** - The user must be able to change basic settings of the uploaded video's playback, such as its title and description, the initial viewing angle and the field-of-view angle.
- **View List of Video's Annotations** - The user must be able to view the list of the annotations of a specified video uploaded by them, in a user-friendly format.
- **Manage Annotation's List** - The user must be able to add or remove annotations from the existing list, using the system's interface instead of directly modifying the file.
- **Save Statistics** - The system must save relevant viewer statistics on the server side.
- **View Statistics** - The user must be able to view the video's statistics, to know how many people have watched it, how many times the video was watched until the end, and what were the areas in the scene that were more focused on.

3.5 System Architecture

In this section, the general architecture of the system is presented, based on the expected features and requirements of the system defined in the previous sections of this chapter.

The visualizer has two potential types of users: the content creator and the consumer, which require different functionalities. After filming and editing a 360° video using the immersive editor of the AV360 project, the content creator submits the output files from the editor to the visualizer, to make them available for others. The content creator may correct the annotation file for a submitted video, and read its viewing statistics. The consumer, on the other side, is able to choose a video to watch from a list of available videos, using either a computer, a VR headset or a mobile device.

Figure 3.9 illustrates the architecture of the visualizer, which includes the relationship between the users and the system.

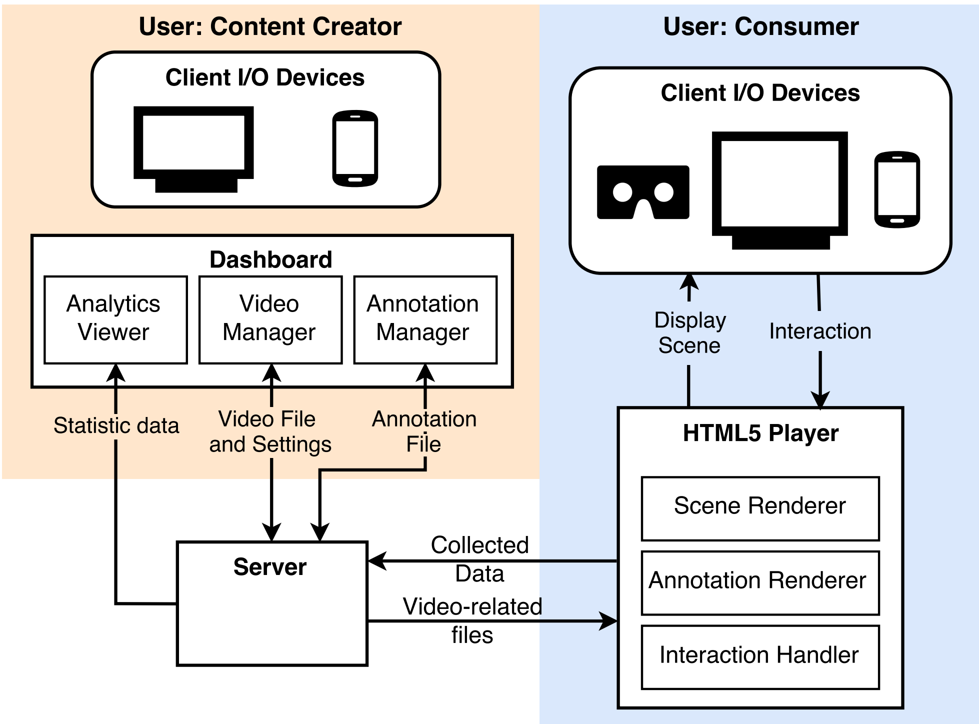
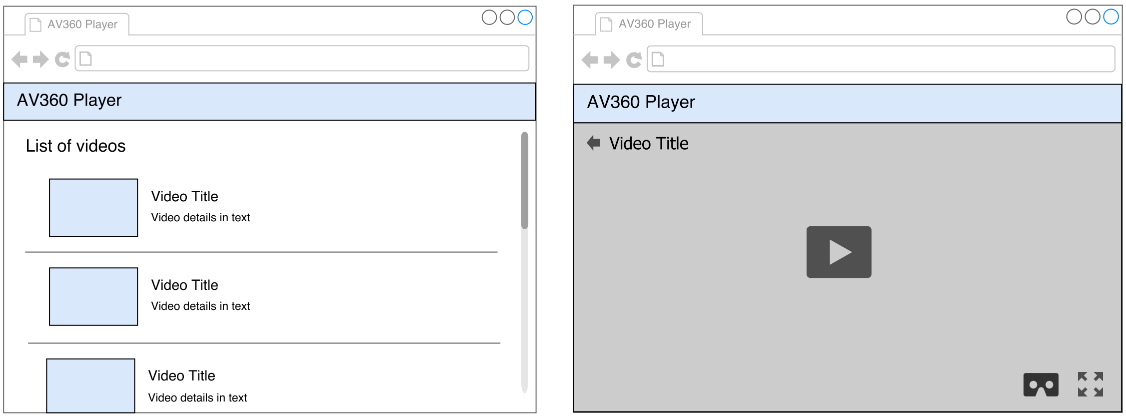


Figure 3.9: Proposed integral architecture for the immersive visualizer.

For this dissertation, the developed prototype focused on only one of the users, the **consumer**, as its associated features and interactions pertain to the established research questions. The requirements relative to the content creator are under implementation, as part of the objectives of the AV360 project.

The consumer requires two main interfaces: the main page presenting a list of videos to choose from, and the video player page. The main page is accessible from any web browser on a desktop



a) Wireframe for the visualizer's main page.

b) Wireframe for the video player's page.

Figure 3.10: Proposed interfaces of the visualizer.

or mobile device, as well as the video player page, which also allows for the use of VR devices such as the HTC Vive. Figures 3.10 a) and 3.10 b) exemplify the described consumer interfaces, to be implemented in the prototype.

3.6 Work Methodology and Preliminary Tests

The development phase of the work progressed iteratively, where each iteration started with a planning step, to decide what would be implemented during the iteration's duration, then the implementation of the planned features, followed by a testing step, where the new functionalities would be tested in different devices, and adjusted if necessary, to ensure the required performance of the system.

At the initial phase of development, the current tools and technologies for web server development and web-based 3D graphics rendering were reviewed, as well as the current state of virtual reality development for web applications. After deciding the tools to be used for implementing the visualizer, the structure of the project was defined, and a basic server and front-end application were created. The video's scene builder and renderer were initialized, starting with a simple display of a video in the 360° view, and basic camera controls. The next step was to implement the annotations of the original version of the annotation file, defined during the immersive editor project, specifically the subtitles and the standard annotation, now labelled "marker".

Once the basic prototype was implemented, compatible with the immersive editor's files, the annotation file's structure and content was redefined, in order to include the new proposed annotations, described in the section 3.3. Each following iteration focused on the implementation of one of the annotations, which would be tested in the compatible devices before moving on to the next feature.

As we reached the end of the implementation phase, the system was tested informally by a small set of users, to obtain external evaluation of the current system, and some enhancements were applied accordingly. From these quick experiments, the following improvements were added to the system:

- **Resized arrows** - The arrows were too large in size, particularly in virtual reality mode, so they were resized to only 5% of the visible area's width;
- **Animated directional annotations** - To become more evident in different video contexts for the user, the arrows and the miniatures are animated with a back and forth movement, towards the user, until the associated marker is seen;
- **Repositioning of peripheral annotations** - In the case of arrows, miniatures, and lateral lights, these were placed too close to the viewing area's periphery in VR mode, rendering them almost invisible. Therefore, these annotations are moved closer to the center;
- **Darkened background for the lateral lights** - Some users mentioned that, in certain areas of the test video, the lateral lights were unnoticeable due to the background image being close to white. As such, a darker color was applied closer to the border of the gradient;
- **Animated vignette when out-of-view** - The preliminary version of the vignette gave no information on the direction of the point of interest when the user was looking at a different direction. Because the darkened effect could be confused with simple video configurations, an animation was applied to the vignette which would help direct the user towards the out-of-view marker;
- **Slower animation for the vignette** - The animated version of the vignette was also tested, and two observations were made: the animation was too fast, and seemed to vary periodically. So, the animation's speed was in general reduced, and reduced further when closer to the effect's starting position.

Lastly, in preparation for the user experiments on the final version of the system, twelve 360° videos were filmed and edited to be used during the user tests, and the system was adapted to save data on the user's interaction with the visualizer.

3.7 Summary

In this chapter, the context of the visualizer was further detailed within the AV360 project, to support the definition of the system requirements. Taking into consideration the established research questions, a list of additional annotations was defined, to be implemented and tested in terms of its efficiency and influence on the user's experience.

A system's architecture was proposed, by illustrating the relationships between the users and the visualizer, and the possible user interfaces. The outlined architecture and required features were used to guide the following work phase, the development of a visualizer prototype. The last section of this chapter describes the overall progress of the development phase, which is covered in the following chapter 4.

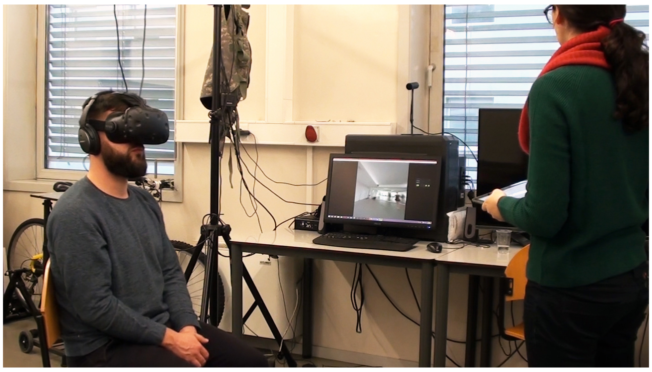
Proposed Architecture for the Immersive Visualizer

Chapter 4

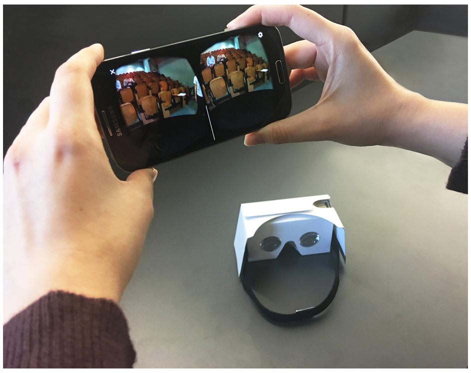
Development of the Immersive Visualizer

The implemented system is a browser-based 360° video player, able to present different types of annotations over the video, with the purpose of guiding the user's focus to the relevant areas of interest at different periods of time in the video. The visualizer's main page contains a playlist of 360° videos, created during the development process, which showcases the implemented annotations in different scenarios. From the main page, a user may choose a video from the list, directing them to the player's page, where the user may watch the video on their computer or mobile device. On the computer, the user can control the scene's rotation using the keyboard or the mouse, and they may also also watch the video in VR using the HTC Vive headset. On a mobile device, the scene rotates with the rotation of the device, using its motion sensors, and it is also possible to watch the video in VR, using any Cardboard headset.

In this chapter, the development process will be reported in more detail, by firstly presenting the tools used in section 4.1, and an overview of the system in section 4.2. Afterwards, section 4.3 presents the changes carried out in the annotation file, followed by an explanation of how



a) User interacting with visualizer, using the VR headset HTC Vive.



b) Visualizer in a mobile device, in VR mode for the Cardboard headset.

Figure 4.1: Use of the visualizer in different devices.

the different 3D scene objects are rendered in section 4.4. The rendering of the scene in VR is described later on in section 4.5, and section 4.6 describes the different interaction methods available to the user. The chapter ends in section 4.7, with the description of the system's features developed for the testing phase of the work, presented and analyzed in the coming chapter 5.

4.1 Development Tools and Technologies

The technologies used to develop the visualizer were researched and decided on an initial phase of development, taking into consideration the system requirements, as delineated in chapter 3.4. The main goal of the visualizer was to be browser-based, to be accessible in different types of devices, and to have the ability to present VR content.

WebVR[1] is a JavaScript API that allows web apps to have access to connected virtual reality devices, such as Google Cardboard or HTC Vive. This fairly recent API, first announced in 2016, creates the connection between the VR device and the 3D scene, sending to the web app information on the device's position and movement, and sending to the headset the rendered scene to be displayed.

Three.js[2] was used for the creation and display of the video and 3D scene. This JavaScript library is based on WebGL, and is able to display 3D graphical scenes across different browsers. At this early stage of development, several examples of VR scenes created with three.js and WebVR were available, which facilitated the beginning of the development process.

Node.js[3] was used for the server side of the system. Considering that Node.js is able to execute JavaScript code for server-side scripting, the system could be developed using a single programming language. It is also cross-platform, which satisfies the compatibility requirement.

4.2 System Overview and Workflow

Using the tools indicated in the previous section, the proposed system in chapter 3 was implemented, in regards to the requirements of the consumer as the user. A Node.js server was created, which controls the navigation through the web application, transmits data regarding the videos to the front-end, and saves viewing statistics of a video in an appropriate file. Creating a server with this tool was also useful to deploy the visualizer application, by setting it up in Heroku[4], a cloud platform for web deployment. This simplified the process of testing the system in different devices, particularly mobile devices.

The first interface to be implemented was the video player's page, which is presented in the figure 4.2, as it is prioritized for the dissertation's objectives. The video to be played is identified

[1] Mozilla's WebVR API documentation. https://developer.mozilla.org/en-US/docs/Web/API/WebVR_API
[2] Three.js Main Page. https://threejs.org/
[3] Node.js Main Page. https://nodejs.org/en/
[4] Heroku Main Page. https://www.heroku.com/home

Figure 4.2: Example of a video containing marker "Mariana PTKS" - "A", a miniature with its blue line connected to a marker invisible to the left of the screen - "B", an arrow at the right - "C", and the minimap placed at the bottom right corner - "D" (the red squares and letters were added to the actual image).

by passing its ID as a variable `video` in the URL. The parsing of this value and the subsequent scene creation functions are performed by the video player's main script, the `sceneBuilder`.

Once the page is loaded, the scene builder script initializes the Three.js scene, its camera and renderer, as well as the appropriate camera controller, according to the user's input device. It then requests from the server the video's data necessary for rendering, and creates the video's scene object, followed by the parsing of the video's annotation file, and consequent creation of the corresponding objects.

The scene builder script controls the render loop, by calling the update functions of each object in scene. If the user starts or stops the video, or if the device requests to enter or leave VR mode, this script performs the required adjustments to the camera and renderer, and calls the adapting function of the annotation objects. Figure 4.3 summarizes this scene building process.

The actions related to the annotations' appearance, position and movement are controlled by

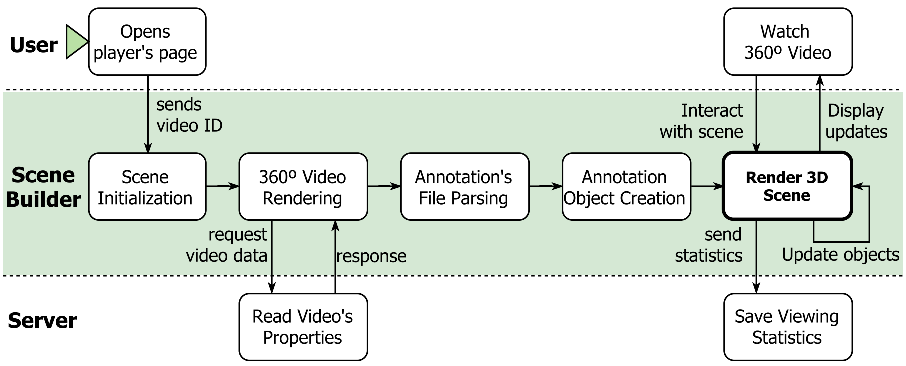


Figure 4.3: Scene building process of the immersive visualizer.

Table 4.1: Table of Implemented Annotations.

Annotation	**Category**	**Position**	**Movement**
Marker	Informational	Fixed	Dynamic
Subtitles	Informational	Persistent	Static
Miniature	Directional	Floating	Dynamic
Arrow	Directional	Floating	Dynamic
Vignette	Narrative	Fixed	Dynamic
Lateral lights	Narrative	Persistent	Dynamic
Minimap	Contextual	Persistent	Static

their specific script. For each implemented annotation, a javaScript prototypical object was created, with a main `Annotation` object, which defines the properties and functions that are common among all types of annotations. The JavaScript prototype objects and their hierarchy, relative to the video annotations, is presented in the following figure 4.4.

From the list of proposed annotations presented in the previous chapter 3.3, the following were implemented: the **subtitles** and **marker**, which are the objects compatible with the immersive editor; two indicator annotations, the **arrows** and **miniatures**; one of the contextual annotations as well, specifically the **minimap**; and two narrative annotations, the **lateral lights** and the **vignette**. To summarize, table 4.1 presents the implemented annotations. Their categories were previously defined in Chapter 3.3, while positioning and movement was defined in Chapter 2.5. The figure 4.2, which shows the implemented video player's interface, contains an example of the subtitles, a marker over the person on the scene, and an arrow indicating an additional point of interest which is out of view.

In the following sections, the new version of the annotation's file, which includes the definition of the implemented annotations, will be presented, and later on, the rendering of each scene object will be detailed further in section 4.4.

The implemented main page interface presented in Figure 4.5, similar to its proposed version

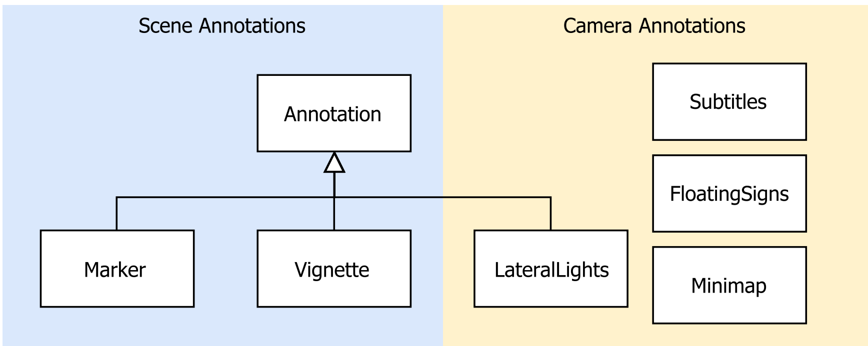


Figure 4.4: JavaScript objects for the annotations. Scene annotations are placed over the video's image, and camera annotations are placed in front of the camera.

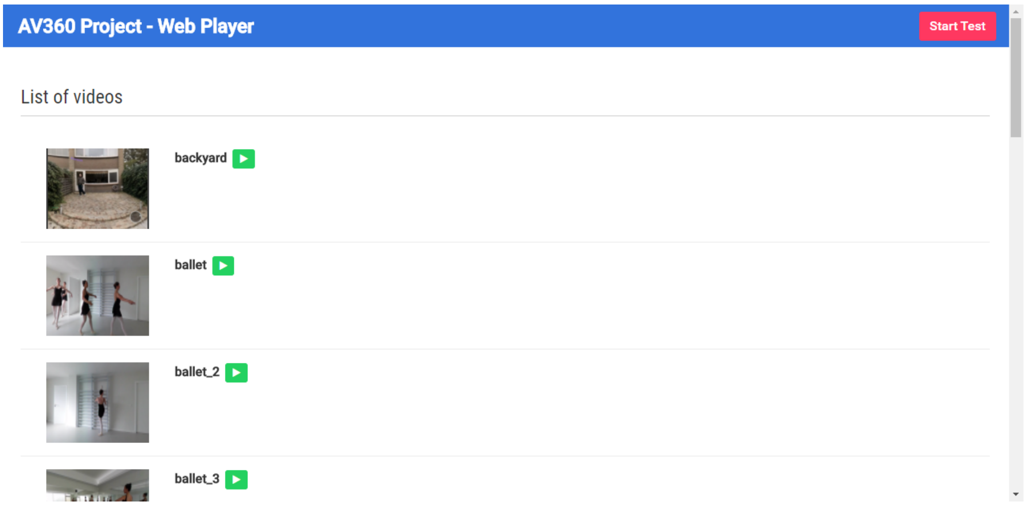


Figure 4.5: Visualizer's main page, presenting a playlist of 360° videos.

in Chapter 3.5, presents a list of available videos, which corresponds to the subfolders of the `videos` folder. Each subfolder contains the video file and the associated annotations' files. The user can choose from the list a video to watch, and the server will redirect the browser to the video player's page, described above.

For the purpose of the user tests, an additional test form page was created, in which relevant information for the system is provided regarding the test session to be performed. Once the form is submitted, the browser is redirected to the video player's page, and the server saves the submitted data in the browser's session, for the purpose of playing the specified list of videos for the test version. The gathered data relative to the test sessions, and other technical details are defined in the section 4.7.

4.3 Extended Annotation File

The file that contains the definition of the annotation objects of a specific video is a SubRip Text (.srt) file, as explained in the description of the immersive editor in chapter 3.1. This type of file, which is commonly used for defining the subtitles of a video, was extended in the context of the immersive editor, by adding a new section, formatted as JSON data, with a list of the annotations to be placed in the video's scene. This method of extending the standard subtitles file allows for it to be readable in traditional players, which are able ignore the JSON block of data and read the remaining file as normal.

In this context, an annotation corresponds to a plane which presents information, using text and/or images, and is placed at a point relative to the video. Each annotation object has an HTML file, which corresponds to its `content`, a `color`, declared in hexadecimal code, `sizeX` and `sizeY`, which are the width and height of the annotation, as power of two, and also an array of `vertices`, or keyframes.

A keyframe is an object with three attributes: `time`, which refers to the timestamp of the video at which the annotation should appear, and a `posX` and `posY`, with values in the range [0-1]. These values correspond to a relative point in the video's area, which is exemplified in Figure 4.6.

The file must always start with an initial subtitle of index 0, with a zeroed timestamp, and the block of JSON data may be placed in any part of the file, between subtitle blocks.

An example of this file, as defined and used in the immersive editor, is as follows:

```
0
00:00:00,000 --> 00:00:00,000
{
"annotations": [
{
    "content": "annotation1.html",
    "color": "#FF0000",
    "sizeX": 512,
    "sizeY": 512,
    "vertices": [{"time": "00:00:00,000",
"posX": 0.45,
"posY": 0.55}]
}]}

1
00:00:03,500 --> 00:00:06,900
Test subtitle 1
```

The annotation file's format was altered to include the new types of annotations defined for the visualizer. The original type of annotation in the editor's file is now labelled `marker`, as mentioned in the previous section. The arrow and miniature are floating objects associated to a marker, so to define them in the file, a simple option is to add the `guide` attribute, which can have the value "none", for a marker without directional annotations associated to it, or "arrow" or "miniature" for the intended type of guide. The subtitles are declared as demonstrated above, and the minimap can be defined as a boolean attribute `minimap`, declaring whether the minimap should be used in the video or not.

The remaining annotations, vignette and lateral lights, are associated to a point of interest, that may change throughout the video. For this reason, the format used for the markers could be adapted to these narrative annotations, which include the definition of `sizeX` and `sizeY`, used to represent the dimensions of the area of interest, and the array of `vertices` for dynamic points of interest, while removing the dispensable `content` and `color` attributes. An additional attribute

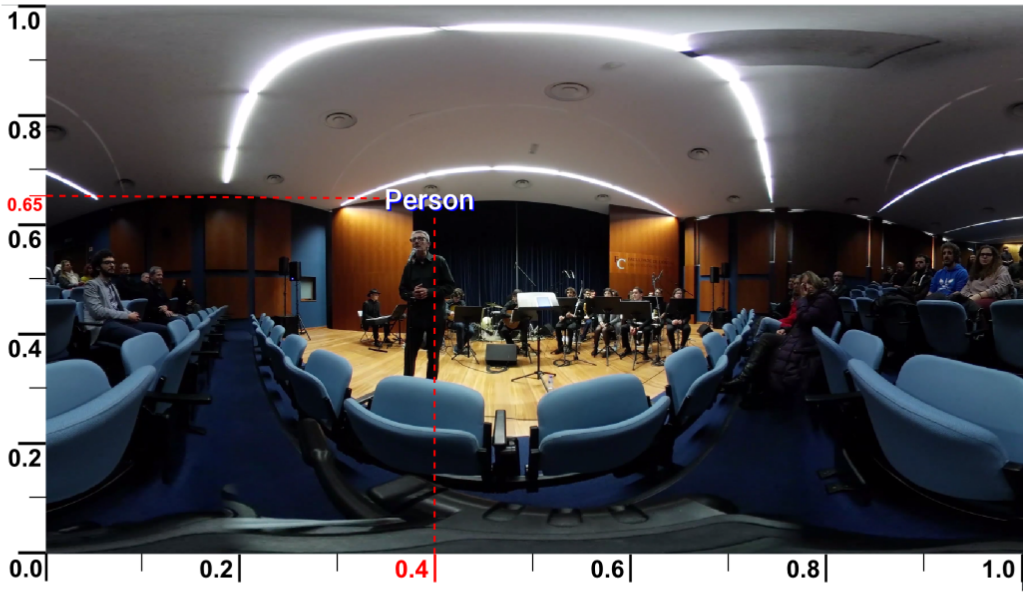


Figure 4.6: Exemplification of a marker being placed at position [0.4, 0.65] in the video's image. A keyframe defines an annotation's position through x and y coordinates that vary between 0.0 and 1.0.

`narrative` is used to declare whether the narrative annotation to be used is a vignette or a lateral light. The following diagram illustrates the structure of the altered version of the annotations' file.

```
Annotation JSON
 ├─minimap: true | false
 └─annotations [0...n]
    └─annotation
       ├─type: marker | narrative
       ├─content
       ├─sizeX
       ├─sizeY
       ├─guide: none | arrow | miniature
       ├─narrative: none | vignette | light
       └─vertices [0...n]
          └─vertex
             ├─time
             ├─posX
             └─posY
```

The system parses through this annotation file to create and render the appropriate scene objects, which is specified in the following sections.

4.4 Scene Building and Rendering

In this section, the process of creating the video's scene is detailed, subdivided for each implemented annotation. Each subsection describes how each annotation is represented in the scene, and how their corresponding scene object is positioned and updated throughout the video playback. Figure 4.2 presents an example of a 360° with different types of annotation in the scene.

4.4.1 360° Video Rendering

The first object to create and place in the scene is the 360° video. As explained previously, when a user advances to the player's page, its URL defines the ID of the video to be played as a parameter. This ID refers to a subfolder on the system's `videos` folder, and it contains a video file named `inputVideo`, with unknown file extension, width and height. The server uses a node module called ffprobe[5] to obtain this information from the video's file, and transmits it to the frontend once requested.

An HTML5 video object is created with the data obtained from the server, and then used as texture for a sphere object created in the scene, which is able to render the video's current state. In order to recreate a sense of depth, and also to allow for enough space for annotations to be placed over the video, the sphere's radius was defined at 10 scene units.

The scene's camera is placed in the center of the sphere, creating the 360° view for the user. To facilitate the use of the video player, particularly in VR mode, the video only starts playing when the device enters VR mode, or alternatively when it enters fullscreen.

The next step is to read the annotations' file and create the annotations as scene objects, which is described in the following section.

4.4.2 Annotation Rendering

As described previously, the annotation file contains the description of the subtitles, in the standard format of SRT files, and the remaining annotations are described as JSON data. The video's main folder has an `annotations` folder, containing the .srt file and all files required for the content of the annotations.

The subtitles' and annotations' sections on the file are divided and parsed separately. In the case of the subtitles, each subtitle block, with attributes `[id, startTime, endTime, text]`, is recorded as an object and saved in an array. In the scene, the subtitles are rendered in a plane object, by using a HTML5 canvas object as the plane's texture and drawing the subtitles' text to it.

This was not the initial method of rendering the subtitles over the video. Originally, the subtitles were simple HTML paragraph objects, placed over the scene, however this was discarded once VR mode was implemented in the system, on account of HTML objects not being visible when the device is in VR mode.

When rendering the subtitles' object, firstly, the video's current time is used to check which subtitle block, if any, should be shown. The subtitle block in question is then drawn on the canvas, using "Helvetica" font, a reference font defined on a W3C Subtitles Specification document[6].

[5] ffprobe information page. https://www.npmjs.com/package/ffprobe

[6] TTML Profiles for Internet Media Subtitles and Captions 1.0. https://www.w3.org/TR/2016/REC-ttml-imsc1-20160421/

Figure 4.7: Example of a video with subtitles.

The font size is also responsive, in order to adapt to desktop and mobile screens. The figure 4.7 demonstrates the final result of the subtitles.

Regarding annotations, once the JSON data is parsed, the type of each annotation is checked to create the appropriate object. Different classes were created for each type of implemented annotation, which are subclasses of an Annotation class, as several attributes are repeated among the annotations. Each type of annotation is detailed in the following subsections.

4.4.3 Marker Rendering

For the marker representation in the scene, the content defined in its HTML file must be placed on top of the video image, moving according to the path, also defined in the JSON data. As explained in the previous section 4.3, the annotation's path is defined using keyframes, defining video coordinates at a specific video's runtime, and the marker moves from keyframe to keyframe, disappearing after reaching the last of the array. In figure 4.2, object A corresponds to this annotation.

The marker's scene object was initially created as a THREE.CSS3DObject, together with its specific renderer, which allows for the mix of DOM elements in a WebGL scene. However, considering WebVR is not compatible with this type of object and renderer, the alternative found was to create an HTML5 canvas object from the original HTML file content, using a JavaScript library called html2canvas [7], and applying said canvas as texture to a plane object.

The position and rotation of the annotation throughout the video is continuously updated to ensure its readability and correct positioning over the video's image. At each update call, the script calculates the current position in video coordinates, using linear interpolation defined as follows:

$$currentPosition(x,y) = startPosition(x,y) + time * translationVector(x,y)$$

[7] html2canvas Website. https://html2canvas.hertzen.com/

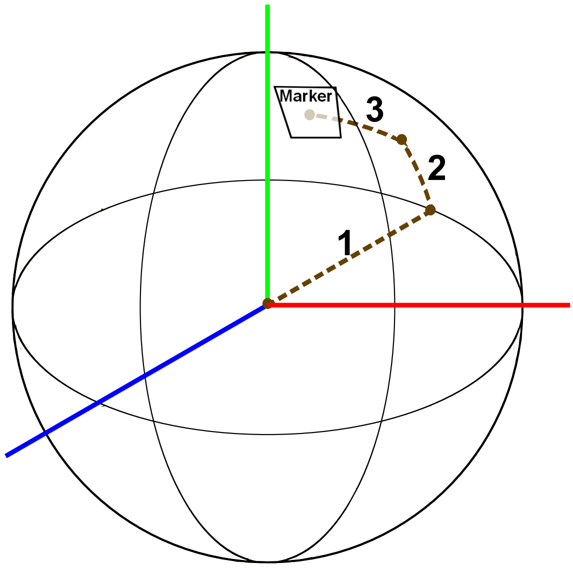


Figure 4.8: Positioning of the marker in the video's scene. **Step 1:** Translation in the z-axis; **Step 2:** Rotation around the x-axis; **Step 3:** Rotation around the y-axis.

where `startPosition` is the position defined in the current keyframe, `time` is the time between the current and subsequent keyframe, and `translationVector` is the vector between the current and subsequent keyframe's position.

The calculated position is used to make the necessary geometric transformations to the plane object. This object, which is initially positioned at the origin, same as the camera, is translated backwards, so that its front side is facing the user, and then it is rotated horizontally and vertically to its position in the sphere, as exemplified in figure 4.8. The distance from the camera is 8 scene units, as it must be smaller than the video's sphere radius, so that there are no intersections between these geometries, but far enough to maintain the user's sense of depth. The angle used for the rotation is calculated using the previously calculated current video coordinates, as follows:

$$\theta(x) = -\frac{\pi}{2} + \pi * currentPosition.y$$

$$\theta(y) = -\frac{\pi}{2} + \pi * currentPosition.x$$

In the next section, the floating guides associated to these markers are described in more detail. These guides, the arrows and miniatures, only appear in the scene as a component of the marker.

4.4.4 Floating Guide Rendering

Arrows and miniatures are also defined as a plane object, however their positioning and dimensions are calculated very differently from a marker. As presented in the proposed layout 3.8 in the previous chapter, these objects occupy a limited peripheral area in the field-of-view, so that the user's view is minimally obstructed. Their dimension is calculated, according to the camera's

FOV and aspect ratio, in order to occupy 5% of the camera's width, and is recalculated when the screen is resized or when VR mode is entered or stopped.

When the respective marker is updated, so is the floating guide. For both the arrow and miniature, the first step is to calculate whether the marker is closer to the left or right side of the camera, and then to place the guide on the left or right region accordingly, vertically closest to the marker. In the case of the arrow, its plane is then rotated so that it is pointing towards the marker.

While the marker is not within the camera's field-of-view, and if it has not been seen by the user before, the guides are visible, with a back and forth motion to attract attention. All four vertices of the marker must be in the viewing frustum for it to be considered visible to the user. The figure 4.2 presents an example of an arrow and miniature floating guide.

The arrow and the miniature objects are exemplified in Figure 4.2, as object C and B, respectively.

4.4.5 Minimap Rendering

A minimap is used to orientate the user in videos with more than one marker in scene, as it gives immediate hints on every marker's position relative to the user. It is represented by a circular plane, as presented in figure 4.2 as object D, with a diameter that is 15% of the camera's width. It contains a small orientation arrow corresponding to the user's orientation, and small dots for each marker in the scene.

If the `minimap` attribute has value of `true` in the JSON data, a minimap object is created, placed on the lower right corner of the field-of-view. For each marker created, a point is drawn in the minimap. When a marker's position or visibility state is updated, so is its corresponding point in the minimap. The orientation arrow is also updated when the camera is rotated by the user.

Similarly to other scene objects, when the browser's window is resized, the minimap's dimensions and position are updated to ensure that it consistently occupies the same amount of viewing area, and that is properly visible. An example of the minimap in the visualizer is visible in the figure 4.2, which contains a scene with a visible marker and two out-of-view markers.

4.4.6 Lateral Light Rendering

The lateral lights are placed in the same peripheral regions as the floating guides, so as not to disrupt the user's view. They are represented in the scene by a plane, with a white radial gradient as texture.

As it is a narrative guide, it has an array of keyframes that defines the narrative area of interest's position throughout the video. In order to calculate its relative position to the user's viewing direction and whether the area of interest is visible, an invisible plane was created to pinpoint the area and, much like the marker's plane, is moved according to the established path.

The light's plane object occupies horizontally 15% of the camera's width, and it is placed to the left or the right depending on which side is closest to its associated area of interest, calling the user's attention towards that direction. Its width and position is adjusted whenever the window

Figure 4.9: Example of a video with lateral lights, directing the user to the left.

is resized. Furthermore, in order to be more evident to the user and not be confused with the video's scenery, the light's gradient is animated, simulating a pulsating light. The figure 4.9 shows the light object on the left side of the video, indicating that the area of interest can be found by rotating towards that direction.

4.4.7 Vignette

The vignetting effect is another method to attract attention towards the narrative's area of interest throughout the video playback. The goal is to maintain the area of interest entirely visible, while the remaining area is radially darkened. This is accomplished by creating a sphere, with a smaller radius than the video's sphere, and centered around the camera as well. A shader is applied to this sphere, which defines its texture's color and transparency. This color varies from complete transparency to black with 80% opacity along the z axis of the sphere's vector normal.

The shader's gradient has two stop points, which divide the gradient in three sections: from 1 to the first stop, the applied color is black with 0% opacity, from the first to the second stop, the color's opacity varies linearly from 0% to 80%, and after the second stop to -1 the color is black with 80% opacity. The figure 4.10 illustrates the sphere's appearance when applying the described shader.

Similarly to the lateral lights, the area of interest is represented by an invisible plane to aid in calculations necessary for when the area is out of view. The sphere is rotated so that its transparent section is located over the area of interest, which can be depicted in the figure 4.11 a).

The visual effect created by the described sphere creates a contrast between the area of interest and its surroundings when the user is looking in its direction, as illustrated by figure 4.11 a). However, if the user is looking in another direction, far enough not to see the transparent section of the vignette sphere, there would be no hint to where to look at, and the darkened effect caused by the sphere may be confused with lighting conditions of the video. It was necessary to create an effect that would call the user's attention towards the right direction in these conditions.

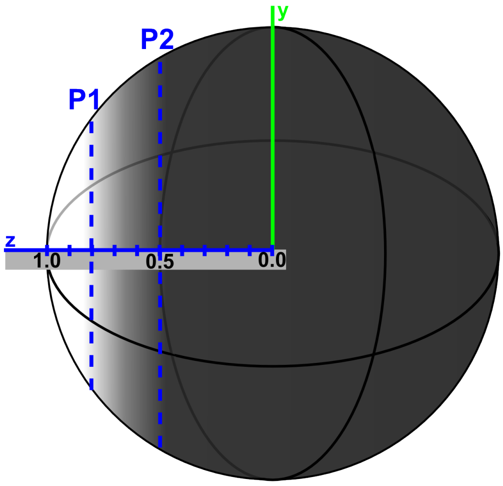


Figure 4.10: Exemplification of the vignette effect applied to a sphere. P1 refers to the first gradient stop, at 0.8, and P2 is the second gradient stop at 0.5.

The solution to the out-of-view issue was to animate the gradient's stop points, varying their position from the initial value to a point visible to the user, and back, until the user turns towards the desired area. Once the area of interest is within the viewing frustum, the darkened area closes around the relevant area.

4.5 VR Mode Rendering

In section 4.1, it was mentioned that the visualizer uses the WebVR API to communicate with VR displays connected to the system. The process of displaying the scene in VR devices starts with an API call to obtain a list of connected devices. If at least one device is identified, the user is given the possibility to enter VR mode. Once the user requests to enter VR mode, the system begins displaying the scene to the user, by drawing the scene twice, one for each eye. The system

a) Example of a video with the vignette effect focusing on a carroussel, which is in the center of the current field-of-view.

b) Example of a video where the area of interest is not centered in the field-of-view. The vignette moves towards the right to guide the user's attention.

Figure 4.11: Vignette's appearance with different viewing directions.

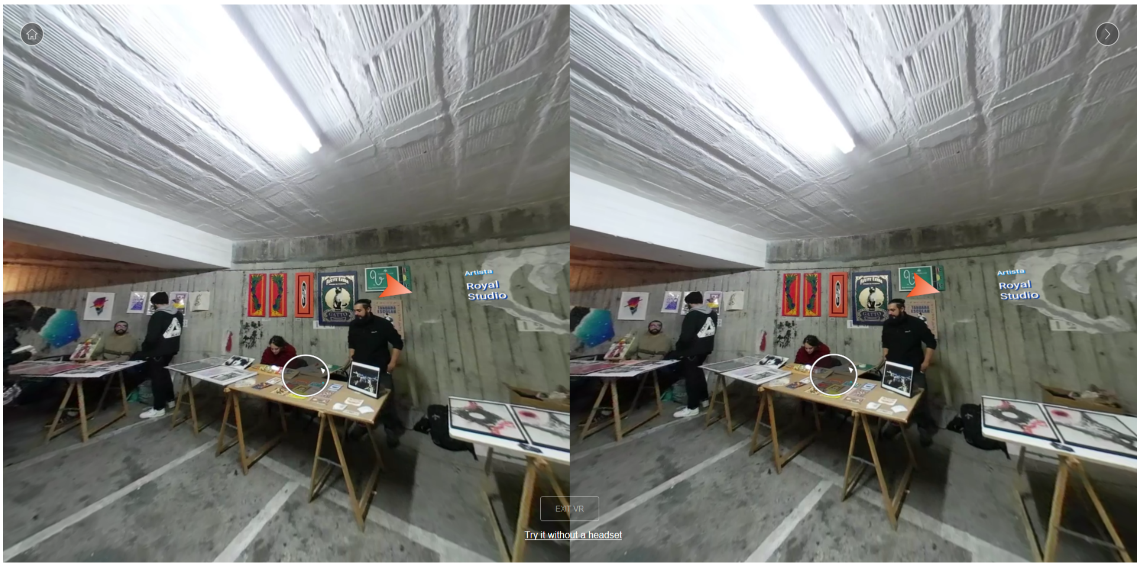


Figure 4.12: Still image of the immersive visualizer in VR mode, using HTC Vive and Firefox browser.

uses the connected device's context data, specifically its rotation and movement data, to control the scene, so that the user's head movements are recreated by the camera.

The technique used to create the three-dimensional view of the scene is relevant for the correct presentation of some scene objects. While the window's width is maintained when entering the VR mode, the viewing frustum occupies approximately half of the actual width, as the scene is drawn twice, side by side. When describing the rendering of the different annotations in section 4.4, it was asserted that those that are persistent or floating annotations depend on the window's width. This implies that, when adapting the scene's object to the change in display, the calculations for those annotations' size and position must consider the available width as being half the actual value.

When testing the initial prototypes of the floating guides in VR mode, it was observed that they were placed too close to the viewing boundaries. To improve the visibility of the guides in this context, the margin to the edges is increased when placing the annotations in VR mode. Additionally, when calculating whether the marker or area of interest is in view, the field-of-view angle used is reduced by 60%, so that it is only considered visible when it is closer to the center of the camera's frustum.

In terms of browser compatibility, currently, there are specific browsers that support WebVR for the different devices. For mobile devices, the compatible browser is Chrome (version 61 was tested and used during implementation), while for HTC Vive, the user must use Firefox's most recent version (version 57.0.4 was this browser's version used for the test sessions). Figure 4.12 demonstrates the user interface of the visualizer in VR mode, on the Firefox browser, using HTC Vive as the virtual reality device.

4.6 Interaction Techniques

In order to freely view the entire image of the video, it is fundamental that the user is able to control the camera's rotation in any supported device. Firstly, the keyboard and mouse controls were implemented, allowing the user to rotate the camera by pressing the arrow keys, or instead by dragging the scene with the mouse. In mobile devices, the user can move the camera around the scene using a one-finger drag in normal view, as demonstrated in Figure 4.13, or by rotating the device around in VR view, as the WebVR API provides the device's gyroscope data to the visualizer to control the camera. This also applies to the HTC Vive headset, whose rotational data is used to also rotate the camera.

Another requirement for the visualizer related to user interaction, specified in chapter 3.4, is the ability to interact with annotations to obtain more information. Considering that the visualizer is to be used in different devices, from a computer, where the user may use keyboard and mouse, to a mobile device inside a Cardboard headset with no means to physically interact with device, the simplest interaction method is to use the user's gaze direction.

To interact with an annotation, in this case a marker, the user must center their gaze on the annotation, without requiring any other input. To indicate to the user their line of sight, a crosshair object is placed in the center of the camera. The first prototype of the crosshair, which was persistent in the user's view throughout the video, proved to be distracting for some individuals, so the final version fades out after a second of stillness of camera. When the crosshair moves over the marker, the marker moves a bit towards the user, as a visual feedback to indicate to the user that the marker is being interacted with. The marker, then, changes its appearance to display the hidden information. Once the user moves away from the marker, its appearance returns to default version. The following Figure 4.14 exemplifies a marker's appearance before and after it is interacted with.

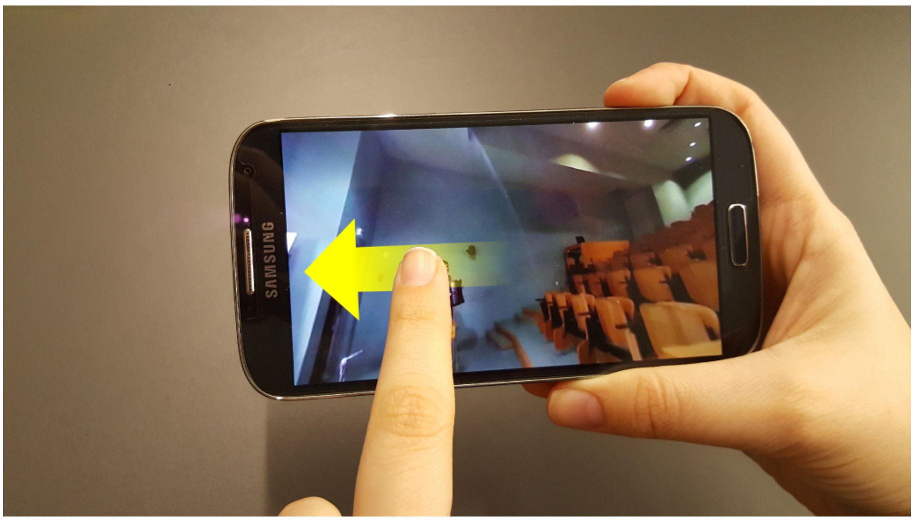
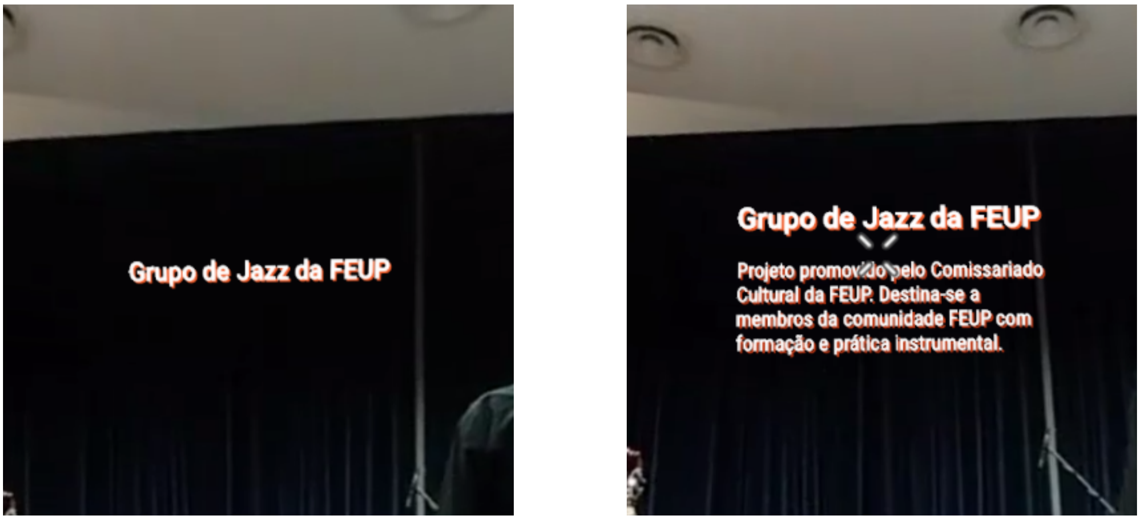


Figure 4.13: Rotating video's scene on a mobile device, using a one-finger swipe, as illustrated by the yellow arrow.



a) The example marker with its default texture, before being interacted with.

b) The example marker, as it is being interacted with, revealing additional information.

Figure 4.14: Example of user interaction with a marker, showing changes in its appearance.

In order to provide the ability to the content creator to choose what information is always visible and what is hidden until the user interacts with the marker, the HTML file which contains the marker's content can be changed for this purpose. The HTML elements to hide can be identified by adding the class `hide` to its definition. The visualizer's system creates a default texture for the marker, with the specified elements hidden, and an alternative texture, which presents all the elements of the HTML file, which are applied alternatively when the user moves over or away from the marker. In case the marker has a miniature as a guide, the miniature's texture will be the same as the default texture.

For keyboard and mouse, it is possible to use other helpful interaction options. Besides camera control and marker interaction, the user can toggle the minimap and floating guides' visibility, by pressing the `M` key for the former, and `F` for the latter.

4.7 Logging User's Interaction Data

One of the goals of the work described in this document is to devise methods of annotating 360° videos, and then understand whether those methods improve the user's experience. After developing the visualizer's prototype and performing the necessary adjustments comes the user test phase, which is more thoroughly described in the following Chapter 5. For this purpose, a testing environment was created in the visualizer.

The tests have four versions, which entails a specific sequence of videos for each version. In the main menu, as visible in the Figure 4.5 above, a button was added to provide access the testing environment, which opens a form page, where the user ID and the test version are indicated. The user ID is used to create the respective file where all collected data is written to, and the test version is used to obtain the sequence of videos that correspond to that test, which is defined in a configuration file. The form is sent to the server, which redirects the browser to the first video. In the video player's page, the Home button ends the test session, and clicking the Next Button makes a request to the server to redirect to the video that is next in the sequence.

When a video starts playing during a test session, the system saves to an array, at every scene update loop, the current video's time, and the coordinates for the point in the video that is in the center of the viewing area. Additionally, for each marker, data regarding interaction from the user is saved as JSON data to the same file, which later facilitates the process of analyzing the test results. The content of the test file is structured as follows:

```
Test JSON
├── userID
├── testID
├── date
└── analytics [0...n]
    └── videoID
        ├── markers [0...n]
        │   ├── name
        │   └── data
        │       ├── startTime
        │       ├── endTime
        │       ├── startTime
        │       ├── startPosition
        │       ├── endPosition
        │       ├── optimalDistance
        │       └── realDistance
        └── dump [0...n]
            ├── time
            └── focusPoint
```

The gathered data during the tests sessions is used to evaluate the implemented system which, together with its purpose and analysis, is described further in the evaluation chapter 5.

4.8 Summary

This chapter described the implementation of the visualizer's system, by detailing the process of creating and rendering the graphical scene and its objects which result in an annotated 360° video. The available interfaces are presented, as well as the supported interaction methods, according to the user's device.

The next chapter reports the testing phase of the work, by specifying the user tests' structure and objectives, as well as the types of data that are collected during a test session, which is afterwards analyzed to assess the efficiency of the implemented system for its established goals.

Chapter 5

Evaluation

The user experiments intends to analyze the efficiency of the developed system, according to the established research questions in Section 1.2. Specifically, they intend to examine whether the annotations assist in the guidance of the user through the narrative of the video, as well as provide additional information, without hindering the immersive experience.

For this purpose, 35 user tests were conducted on the system, from which several data was collected to be analyzed afterwards. This chapter presents in detail the followed protocol and structure of the test sessions in sections 5.1 and 5.2. Section 5.3 describes the 360° videos created for the purpose of the user tests, followed by the definition of the different types of data gathered during those sessions in section 5.4. In the section 5.5, the results obtained from the performed tests are presented, and later on analyzed and discussed in the final section 5.6.

5.1 Evaluation Protocol

The visualizer's annotations described in chapter 4 were created with the purpose of guiding the user, drawing attention to points of interest, and providing more information. The user tests were defined to cover the different applications of the annotations, and a comparison could be made afterwards between videos with and without overlaying information or indicators. Four tasks were defined, each with a different focus, as follows:

- **Task A, Locate an annotated Point of Interest** - For this task, the user must find an annotation in the video's scene, starting from the same position. The task is performed in three videos: one without guides (A.n), another with arrows (A.a), and the other with miniatures (A.m). The purpose is to compare the ease of finding one annotation with and without guides.

- **Task B, Locate several annotated Points of Interest** - In the second task, the user must watch the video and find the different annotations that appear throughout the video. Similar to the previous task, three videos are used, one without guides (B.n) and the remaining with the implemented floating guides (B.a for arrows and B.m for miniatures). Besides examining if the users find the annotations, this task also tests whether the user is aware of their surroundings, and knows where to turn to throughout the video.

- **Task C, Follow the Video's Narrative** - This task focuses in the narrative of the video. With a voice narrating the surrounding area, the user is expected to find the areas of interest referred to by the narration. Three videos are played, each with its own narrative, one without any guides (C.n), another with vignette (C.v), and the other with lateral lights (C.ll). This task serves to examine if the user is able to follow the narrative in a 360° environment.

- **Task D, Obtain Information from the Video** - In the last task, the videos present a scenario with several points of interest that the user may want to observe. In this scenario where the user's attention can be focused anywhere around the scene, a few annotations containing information pertinent to the video's action appear. Comparing the video without guides (D.n) with one with the vignette (D.v) and another with the lateral lights (D.ll), the goal is to analyze the user's intake of knowledge, which refers to the information in the annotations, but also the visual and auditory information obtained from the video.

As mentioned, each task contains three videos with different annotation conditions. For the user tests, the order by which the videos were played was distinct, following a Latin Square permutation, demonstrated in Table 5.1. The specific annotations applied in each video is detailed in the next section.

Table 5.1: Order of the tasks' videos, according to the test version.

	Task A	Task B	Task C	Task D
Test 1	A.n - A.a - A.m	B.a - B.m - B.n	C.v - C.n - C.ll	D.n - D.ll - D.v
Test 2	A.a - A.m - A.n	B.m - B.n - B.a	C.n - C.ll - C.v	D.ll - D.v - D.n
Test 3	A.m - A.n - A.a	B.n - B.a - B.m	C.ll - C.v - C.n	D.v - D.n - D.ll

The interaction device chosen for the tests was the HTC Vive, which allows for an immersive virtual reality experience, contrary to the desktop environment with keyboard and mouse. The HTC Vive headset was also chosen over a smartphone combined with a Cardboard headset, on the grounds that the former has better visual performance, with a wider field of view and higher resolution, besides being more stable in VR mode.

Data was gathered throughout each test session, starting with a short demographics questionnaire, which requested their gender, age, and experience with VR devices, specifically Google Cardboard, HTC Vive and Oculus Rift. After each video experience, some questions formatted in Likert Scale were asked to the users, and in the end of the session, a final System Usability Scale questionnaire was filled regarding the overall experience with the visualizer. More details on the types of gathered data and how they were obtained are presented in the section 5.4.

5.2 Test Session Structure

A user test begins with a short introduction of the system that is being tested, and a explanation of how the test would be conducted. The users were informed that they would be interacting with a 360° video player, which was capable of displaying interactive annotations, which mark points of interest, provide additional information and guide the user in the scene. The structure of the test is afterwards presented, explaining that the user would be performing four tasks, each in three short videos with different annotation conditions. After each video and each task, they would be asked a few questions on the preceding experience.

The device used for the test, the HTC Vive, would then be presented, with a warning for the possible side-effects of using virtual reality headsets, such as motion sickness and general discomfort. The user would be informed that, in case of any symptom, or any other reason, the test session could be interrupted, or even cancelled, at any time, as their participation is voluntary.

Regarding the collected data, the user is then informed that all information gathered with the questionnaires and system would be used confidentially, and exclusively in the context of the system's evaluation. The user is given a consent form, which reiterated these conditions, to be signed if they understood and agreed with the conditions. Furthermore, some test sessions were filmed for the presentation purposes, for which the users were requested to sign an additional consent form regarding their image rights.

After the introduction is finalized and the required documents are signed, the user fills out a demographics questionnaire, as mentioned previously. A short video is presented before the tasks' videos, to ensure the system is working properly, and that the user is comfortable and has adapted to the system. This video contains a standard marker, shown in figure 5.1, to demonstrate to the user how to interact with these objects, and to make any necessary adjustments to the headset, in case the user has trouble reading the information presented in the marker.

Figure 5.1: Preview of the video used before starting the session tasks. The marker visible in the image serves to introduce the users to this type of annotation.

Fill out Demographics Questionnaire			
Task A - Locate an annotated point of interest			
Videos x3	Video A.n	Video A.a	Video A.m
Annotations	Marker, no guides	Marker, arrow guide	Marker, miniature guide
After each video	Questions A.1 and A.2		
After task	Questions A.3 and A.4		
Gathered Data	Optimal distance, Real distance, Elapsed time		
Task B - Locate several annotated points of interest			
Videos x3	Video B.n	Video B.a	Video B.m
Annotations	Markers (3), no guides	Marker (3), arrow guide, minimap	Markers (3), miniature guide, minimap
After each video	Questions B.1 and B.2		
After task	Questions B.3 and B.4		
Gathered Data	Identification of markers found for each video		
Task C - Follow the Video's Narrative			
Videos x3	Video C.n	Video C.ll	Video C.v
Annotations	No guides	Lateral Lights	Vignette
After each video	Questions C.1 and C.2		
After task	Questions C.3 and C.4		
Gathered Data	Identification of points of interest found for each video		
Task D- Obtain Information from the Video			
Videos x3	Video D.n	Video D.ll	Video D.v
Annotations	Markers (2), no guides	Markers (2), Lateral Lights	Markers (2), Vignette
After each video	Questions D.1 and D.2		
After task	Questions D.3 and D.4		
Gathered Data	Recollection of video's points of interest and markers' information		
Fill out System Usability Survey Questionnaire			

Table 5.2: Summary of the test session's structure.

Once the user indicates that they are ready to start the test, a test session is created in the visualizer. Each task is explained before commencing, by describing what the user is expected to do in each video, and what are the annotations that will be tested. The questions that the user answers after each video and task are answered by choosing one of the five possible responses, which relate to how much they agree or disagree with each statement. Any supplemental comment that gives further insight into their experience and opinion on the system is written down in the questionnaire.

For the most part of the session, the user does not remove the headset, to reduce the session's running time and avoid needing to readjust the headset. In case the user indicates discomfort, the headset is removed for a while. At the end of the last task, the user fills out a last questionnaire in regards to the overall experience with the visualizer. Table 5.2 summarizes the overall structure of the performed test sessions.

5.3 360° Content Creation for Testing

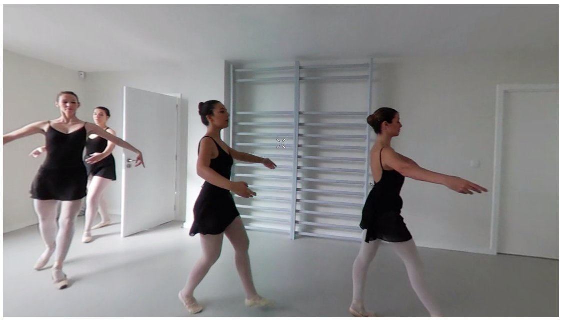
For the user tests described in this chapter, it was necessary to create 360° video content appropriate for each task. After defining each task, different activities occurring in the city were analyzed, in order to find options that would have free access and that would fulfill the requirements for a specific task. To film the 360° videos, a Samsung Gear 360 camera was used, which, as described in chapter 2.4, is capable of filming the surrounding scene entirely, with 4K resolution, besides providing a mobile application that could be used to control the camera through a Bluetooth connection. The resulting chosen scenarios, each illustrated in figure 5.2, were as follows:

- **Task A** - For the first task, the plan was to have a small number of points of interest, possibly dynamic, with a simple background. A local ballet school was contacted, with the request to film a ballet rehearsal, which was accepted. Three videos, of an approximate duration of 30 seconds each, were filmed, depicting the entry and positioning of 4 ballerinas in the rehearsal room, the warm-up exercises, and lastly a short choreography. In each video, a different ballerina was defined as the point of interest, by placing a marker with their names near their location.

- **Task B** - The second task required a space with a large amount of possible points of interest that could be annotated throughout the video. Local events such as exhibits and markets were researched for this task, and the event that was ultimately filmed was an open market of urban art and illustration, where around 60 artists would have an exhibition table to demonstrate and sell their work. Three different points in the market were filmed for each video, and in the editing phase, 3 different artists were annotated in each sequentially through the video's play time.

- **Task C** - For this task, a large open area which had several points of interest surrounding the camera was required. Preferentially, spaces with temporary points of interest were searched, in order to avoid that users have previous knowledge of the scene. The Rotunda of Boavista was decorated for Christmas, with an ice rink and several carrousels and other attractions, and so, three videos were filmed in different points around the rotunda. After filming, a narration was added to each video, which described the scenario by referring to 3 specific points of interest.

- **Task D** - The final task required a scene with many areas of interest that could draw the attention of the user, and that depicted an event that had several related pieces of information that could be presented in annotations. A choir from University of Porto, specifically, Coro Orfeão, had a Christmas Concert scheduled at the Clérigos Church, so a request was made to be able to film the concert. The 360° camera was placed in the center of the stage, around which the choir members, the organist and the maestro were positioned. As it was for the previous tasks, 3 videos were filmed and edited, each presenting a different song being interpreted. Two annotation markers were placed in each video, one containing information



a) Task A - Ballet Rehearsal.

b) Task B - Urban Art Market.

c) Task C - Outdoors Christmas event.

d) Task D - Christmas Classical Concert.

Figure 5.2: Previews of the created content for the four session tasks.

on the performed song, and another identifying either the choir, the maestro, or the organist. These annotations were placed near their corresponding area of interest, close to the normal line of sight so that they are easily spotted.

After filming, the videos were stitched and edited using the camera's editor application, Gear 360 ActionDirector. The resulting stitched videos were then edited using the immersive editor to create and place the annotation markers in the videos, after which the annotation files were adapted to the new annotation terminology and file format defined in section 4.3. The final version of the video and annotation files were subsequently placed in the visualizer, tested and readjusted if deemed necessary.

5.4 Gathered Data

Throughout every test session, different types of data were gathered for subsequent analysis. Part of this data was subjective information obtained directly from the user, by asking some questions regarding their experience in each task, using usability questionnaires. Adding to the results obtained from these questionnaires, there is also the data gathered by the system regarding the user's interaction with it, which can be used to help validate assertions that result from the analysis of the user opinions.

5.4.1 Usability Questionnaires

The different users' impressions on the implemented system may give insight into the efficiency and impact of the video annotations on their experience. A Likert-scale questionnaire was created per task, focusing on its objective for the users, containing 2 questions to be asked after each video of the task, and then two other questions at the end of the task. The answers vary from 1 - Completely Disagree, to 5 - Completely Agree, with an additional option of "Not applicable", in case the user is otherwise unable to answer the question, for not having noticed the annotation in question, for example. Adding to these, for task B, the number of markers found is also written down in its questionnaire, and likewise in task C, regarding the narrated points of interest.

In the case of task D, where the user's ability to intake information from the video and annotations is being examined, a list of nine questions was created for each video, the first three being the same for every video and focusing on the visual information of the video's scene, the next three refer to one of the markers in the video, which was about either the choir, the pianist or the maestro, and the last three were related to the song of the video. The answers for this questions can be "Yes", "No" or "Not sure".

At the end of the session, a last questionnaire was given to the users, which consisted of a System Usability Scale test regarding the overall system experience, with the following questions:

- It was comfortable to use;
- It was entertaining;
- It was easy to use;
- I would repeat the experience;
- I was confident in using the system;
- The system's behaviour was consistent.

The exact questions for each task, and the answers obtained from the tests sessions are detailed in the section 5.5 further on.

5.4.2 User's Interaction Data Logs

During the first task, task A, the system gathered information on the user's viewing direction throughout every video, as well as data on the interaction with the marker. This data was then recorded in a file, unique for each test session, as defined in chapter 4.7. Once the video starts, every 100 milliseconds, the system saves the **time** of the video and the **focus point**, which corresponds to the {x,y} video coordinates of the point in the center of the user's view. Regarding the interaction data, for each marker the following values are recorded:

- **Initial Position** - The x,y coordinates of the video's point at the center of the user's view, recorded when the marker was placed in the video;

- **Final Position** - The x,y coordinates of the video's point at the center of the user's view, recorded when the user interacted with the marker;
- **Elapsed Time** - The time it took for the user to move from the initial to the final position;
- **Optimal Distance** - The length of the shortest path between the initial and the final position;
- **Real Distance** - The length of the path the user took from the initial to the final position.

The optimal distance corresponds to the minimum arc length between the two points in the sphere, while the real distance corresponds to the sum of the arc lengths between each recorded positions, as exemplified in the figure 5.3. As a reference, the arc length of a full horizontal rotation in this context is `2π*radius`. Taking into consideration that the radius of the sphere is 10 units, which was defined in the previous chapter 4.4, then the distance corresponding to a full circumvolution is 62.8318 units.

5.5 User Experiments

In this section, the gathered data will be reviewed more comprehensively, starting with a general description of the user population in subsection 5.5.1, followed by dedicated sections for each task's description, results and analysis.

Figure 5.3: Exemplification of the optimal and real distance, represented by the green and blue line, respectively. The optimal path is a curved line, as it is travelled along the video's sphere.

5.5.1 User Population Overview

In total, 35 user tests were performed as defined in the previous section 5.1. The participants were invited through direct contact or e-mail, with a brief description of the context of the experiment. The ages of the participants ranged from 18 to 43, with a median age of 24. In terms of experience with virtual reality headsets, 65% indicated that they had previous experience with at least one of the referred devices, but only 11% had experienced all of the devices. The general demographics of the 35 participants are presented in the following tables:

Table 5.3: Gender of the Participants.

Gender	Participants
Female	16
Male	19

Table 5.4: Age of the Participants.

Age	Participants
18-22	12
23-27	14
28-32	5
33-37	1
38-42	2
43-47	1

Table 5.5: Participants's experience with different VR headsets.

VR Headsets	No experience	Some experience	A lot of experience
Cardboard	21	9	2
HTC Vive	20	11	1
Oculus Rift	19	12	1

5.5.2 Task A Experiment

Description For the first task, all participants were equally positioned in order to face towards the same direction. The videos used for this task, video A.n, A.a and A.m, presented a ballet scene, with four ballerinas performing different actions in each video. Once the marker appeared in the scene, the users were told they could start exploring the video to look for the marker, which was placed over one of the four ballerinas. Video A.n does not have guides for the marker, while video A.a uses an arrow guide, and A.m uses a miniature. The video's playing order is defined by the version of the test, as explained in section 5.1.

The users were presented with two sentences after every video, and another two at the end of the task, to which they were asked to express how much they agreed or disagreed. The participants were also encouraged to provide additional information regarding their answers and general experience of the task.

- **Sentence 1 (A.1)** - After each video: "I knew where to look to find the marker annotation."
- **Sentence 2 (A.2)** - After each video: "It was easy to find the marker annotation."
- **Sentence 3 (A.3)** - At the end of the task, regarding the guides: "The guide's performance was consistent with my expectations for this task."

- **Sentence 4 (A.4)** - At the end of the task, regarding the guides: "Comparing to video without guides, the guide enhanced the experience of the video in this context."

Additionally, the system starts recording data regarding the user's interaction with the marker, from the beginning until the end of the video, as listed in the previous section 5.4.

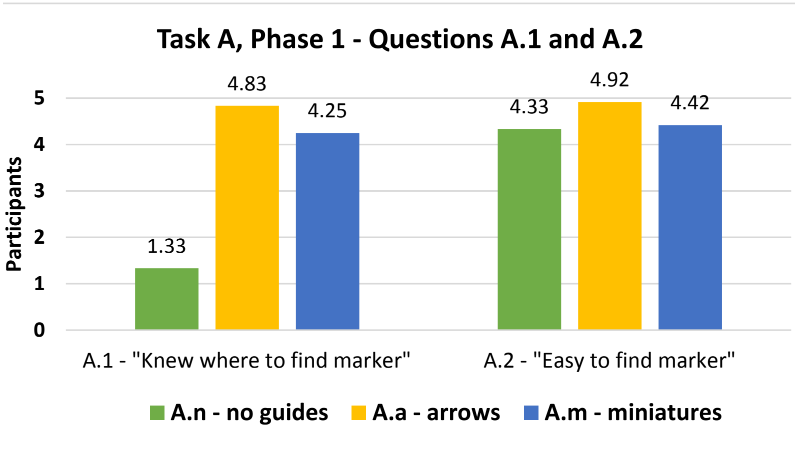
Issues After the first twelve test sessions, some issues with the definition of the task became evident, which were reflected in the values recorded by system, that is, the real and optimal distance travelled from the start point to the marker. In the video with miniatures, due to the dynamic nature of the marker, presented values of the real and optimal distance that varied greatly between the users. The annotation in this video had a longer horizontal translation, comparing to the other videos, which meant that, if the user moved slower to the annotation once they started exploring, then the annotation's position would be a lot further from the beginning, to the point where the optimal path would have been to turn to the opposite side at the beginning.

Furthermore, in the video without guides, these values depended heavily on the initial decision from the user to either turn left, away from the annotation, or right, towards the annotation. Since the marker was positioned fairly close to the user's initial position and at eye level, moving right meant that the users would quickly find the marker. In the 12 tests, only 2 turned left, meaning that the majority of the results had almost no difference between the optimal and real distance. The position of the marker was, as such, poorly chosen, as the results depend on the chance of turning to the right side, and not on how easy it was to find markers in this context.

For this reason, before advancing any more with the test phase, this task A was redesigned. The new markers were placed statically in the scene, to avoid the first specified issue. Additionally, all markers were positioned further away from the normal horizontal line of sight, requiring the users to look around the scene. The data gathered before and after these changes are subsequently presented separately, referred to as Phase 1 and Phase 2 of the tests.

5.5.2.1 Phase 1

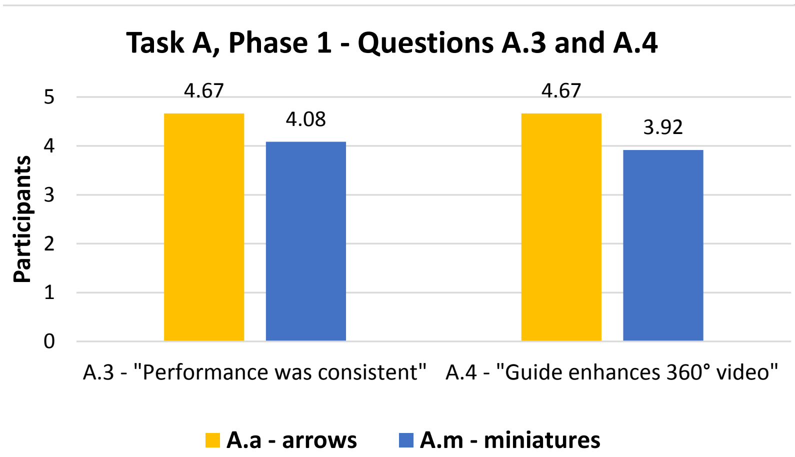
Results The following graphs represent the answers provided for the questions defined above. In the twelve tests that were run during the first phase, for the first question A.1, 11 participants reported that they had no sense of where to look for the annotation, while 10 felt complete confidence with the arrows, confidence that lowered with the miniatures, as seen in the graph 5.4. For the second question A.2, also in graph 5.4, the results were closer for each video, as more than half of the users felt it was easy to find the marker in every video, more so with the arrows. For the questions related to the guides of the task, the arrows were evaluated as having better performance throughout the video, as visible in graph 5.5, while also being able to improve the experience, to a greater extent than the miniatures, shown in the same graph.These graphs present the central tendency value, the arithmetic mean, for each video, for further details the graphs in A.1 presents the number of people that provided each answer option for every question and every video.



Question A.1	1st Quartile	Median	3rd Quartile
No Guides	1	1	1
Arrow	5	5	5
Miniatures	4	4	5

Question A.2	1st Quartile	Median	3rd Quartile
No Guides	4	5	5
Arrow	5	5	5
Miniatures	5	5	5

Figure 5.4: Task A, Phase 1: Question A.1 and A.2.



Question A.3	1st Quartile	Median	3rd Quartile
Arrow	5	5	5
Miniatures	4	4	5

Question A.4	1st Quartile	Median	3rd Quartile
Arrow	5	5	5
Miniatures	4	4	5

Figure 5.5: Task A, Phase 1: Question A.3 and A.4.

Regarding the interaction data, described in the previous section 5.4, the following table 5.6 presents the average time, and table 5.7 presents the difference between the optimal and real distance for each video. Firstly, regarding the time users took to find the marker, the video without guides (A.n) and the video with miniatures (A.m) had very similar results, the former averaging at 7.85 seconds and the latter at 7.94 seconds. Meanwhile, the arrows (A.a) had better performance, diverging in 2 seconds with an average of 5.16 seconds.

Additionally, for the video without guides, the average difference between the optimal and real distance was 12.7 units. For the video with arrows, the value decreases to 5.7 units, and lastly, the video with miniatures showed the worst results, averaging in 16.7 units.

Table 5.6: Task A, Phase 1 - Elapsed time for finding the marker, in milliseconds, for each video.

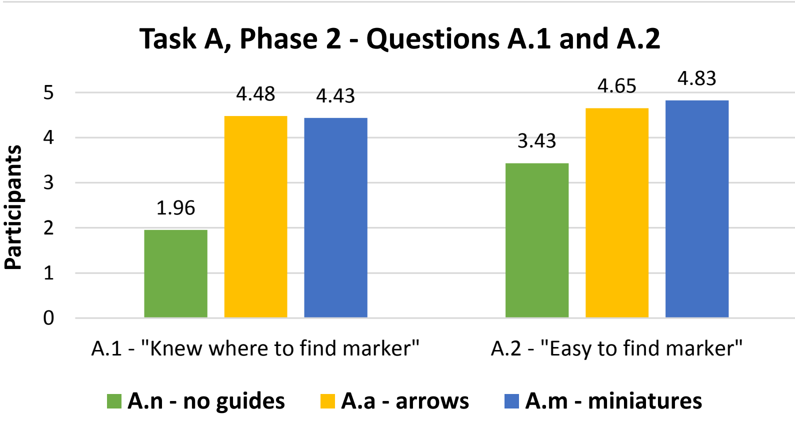
Elapsed Time (ms)	Mean	Range (Max - Min)	Standard Deviation
A.n - No guides	7848	10767	3391
A.a - Arrows	5160	5386	1747
A.m - Miniatures	7938	8335	2691

Table 5.7: Task A, Phase 1 - The difference between the optimal and real distance, in scene units, for each video.

Optimal - Real Distance	Mean	Range (Max - Min)	Standard Deviation
A.n - No guides	12.69761	38	11.8136
A.a - Arrows	5.6836	23	8.9248
A.m - Miniatures	16.6609	42	14.1564

Analysis From the initial version of Task A, the users' answers indicated that, in the video without guides (A.n), most did not know where to turn to, and spontaneously chose a direction. The arrows in video A.a had an almost opposite reaction, as the users in general knew with confidence where to turn to, while the miniatures from video A.m were not as effective as the other guide. Regarding the two guides, the participants commented that they felt more difficulty with the miniatures, as they were not as intuitive as the arrows, taking some time to understand their overall appearance and behavior earlier on in the video. Whereas the arrows, being a commonly used object for navigation and direction, required no initial familiarization. Even though the arrow was the favored type of annotation, some issues with its design were brought up by the participants, specifically that it was too large and potentially distracting.

The data gathered by the system regarding the path of the user's focus supports the answers of the users, and reflects the issues presented above. Using the arrows as guide took less time for the users to find the marker, without significant variation between participants. This was not the case for the other videos, which had similar higher results, as the participants had to explore the video without guides to find the marker, and some found the miniatures to be confusing at first, which could explain the higher average time, in comparison with the arrows. Looking also at the difference between the optimal and real distance for each video, the arrows again had the best



Question A.1	1st Quartile	Median	3rd Quartile
No Guides	1	2	3
Arrow	4	5	5
Miniatures	4	5	5

Question A.2	1st Quartile	Median	3rd Quartile
No Guides	3	4	5
Arrow	4	5	5
Miniatures	5	5	5

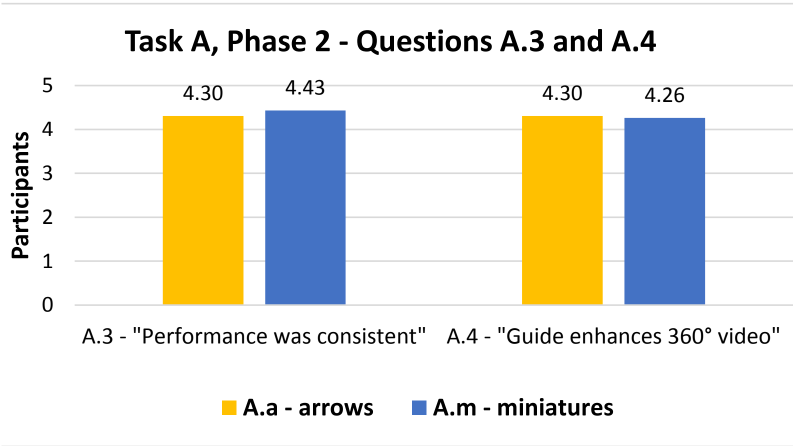
Figure 5.6: Task A, Phase 2: Question A.1 and A.2.

results, which goes with the user comments on the experience. The miniatures effectively had the worst results, which can be explained by the issue identified previously, and was tackled in the second phase of tests.

5.5.2.2 Phase 2

Results After altering the videos' annotations, 23 additional tests were performed. In question A.1, represented in the graph 5.6, 43% of the participants answered that they did not know where to look for the marker, in the video without guides (A.n). With the guides, 60% were completely confident of the marker's direction with the arrow (A.a), and 52% for the miniatures (A.m). For the second question A.2, the difference in results between the video with and without guides is noticeable, as only 30% considered it easy to find the marker without guides, increasing to 65% for the arrows, and 82% for miniatures. In this second phase of the tests, the evaluation on the guides was closer between the two options for the questions A.3 and A.4, which evaluate the perceived performance and efficiency of each guide, as seen in graph 5.7. These graphs present the central tendency value, the arithmetic mean, for each video, for further details the graphs in A.2 presents the number of people that provided each answer option for every question and every video.

The data relative to users' interaction with the video's marker, summarized in the following table 5.8 and 5.9, saw larger differences from the results of the first phase of tests. In terms of elapsed time, the results were fairly similar to the 12 previous tests, as the arrows had again the fastest average time, at 6.01 seconds, and the video without guides averaged at 8.68 seconds, leaving the miniatures with the slowest time, at 9.46 seconds. The followed path to the marker was, in general, longer in this second phase, however this time, the video without guides had the



Question A.3	1st Quartile	Median	3rd Quartile
Arrow	4	4	5
Miniatures	4	4	5

Question A.4	1st Quartile	Median	3rd Quartile
Arrow	4	5	5
Miniatures	4	4	5

Figure 5.7: Task A, Phase 2: Question A.3 and A.4.

worst results, with 24.7 units as average, and the guides resulting in proximate values, with 14.7 and 15.31 units for the arrows and miniatures.

Table 5.8: Task A, Phase 2 - Elapsed time for finding the marker, in milliseconds, for each video.

Elapsed Time (ms)	Mean	Range (Max - Min)	Standard Deviation
A.n - No guides	8674	18252	4384
A.a - Arrows	6013	21222	4418
A.m - Miniatures	9458	15940	4543

Table 5.9: Task A, Phase 2 - The difference between the optimal and real distance for each video.

Optimal - Real Distance	Mean	Range (Max - Min)	Standard Deviation
A.n - No guides	24.7028	118.9323	20.3191
A.a - Arrows	14.7050	86.9492	8.9248
A.m - Miniatures	15.3103	89.7491	20.5879

Analysis The users' responses regarding the video without guides (A.n) maintained similar results to the ones provided in the previous phase, as most participants were likewise unsure of where to turn to find the marker. The responses regarding the videos with guides do not have great distinction between the arrows and miniatures (A.a and A.m), however, through spontaneous comments, some users reported that they preferred the arrows, as they are more familiarized with this type of guide, and because the miniatures are confusing at the beginning, and so, less intuitive. Some users furthered this claim by adding that once the miniature became visible, they were unsure if it was a guide or the actual marker, which added to their confusion.

This confusing quality of the miniatures might help explain the greater amount of time it took for the users to find the marker, as some of that initial time was used to figure out the guide.

Overall, both the elapsed time and path length to the objective marker for each video worsened, when compared to their phase 1 counterparts, as the markers' positions in this phase required larger head movements to be found. Although the guides had reasonably better results than the video without guides, it is possible that their results might have been worse due to the fact that they are only positioned to the left or right of the screen, which may not be the an effective option to indicate objects that are above or below the current field of view.

Regarding the issues that called for the adjustment of the conditions of this task A, these results suggest that the alterations removed the issues, as the ability to find the markers quickly in the video without guides seemed to be less reliant on the initial direction decided by the user, and the miniatures in fact improved from the first phase.

5.5.3 Task B Experiment

Description The second task contained three videos displaying an art market where, in every video, three distinct artists were annotated at subsequent times. The users were once again placed at an initial position, however they could start exploring the video from the beginning. Once they found a marker, the participants were requested to read its text out loud, for the purpose of writing down which markers were found. The videos for this task, B.n, B.a and B.m, are annotated similarly to the first task, that is, B.n only has the markers, B.a uses the arrow guides for the markers and a minimap, and B.m uses the miniatures combined with the minimap.

Similarly to the first task, the users had to answer two questions after each video, and another two at the end of the task, while any spontaneous comment during the answers was written down.

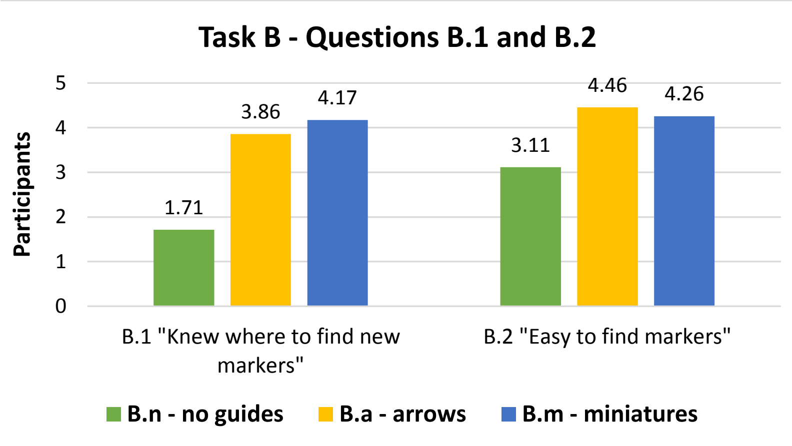
- **Sentence 1 (B.1)** - After each video: "I knew where to look to find new marker annotations." The purpose of this question was to understand if the user was aware that, in a certain direction, there would be a new marker, or otherwise an already seen one.

- **Sentence 2 (B.2)** - After each video: "It was easy to find the marker annotation."

- **Sentence 3 (B.3)** - At the end of the task, regarding the guides: "The guide's performance was consistent with my expectations for this task."

- **Sentence 4 (B.4)** - At the end of the task, regarding the guides: "Comparing to video without guides, the guide enhanced the experience of the video in this context."

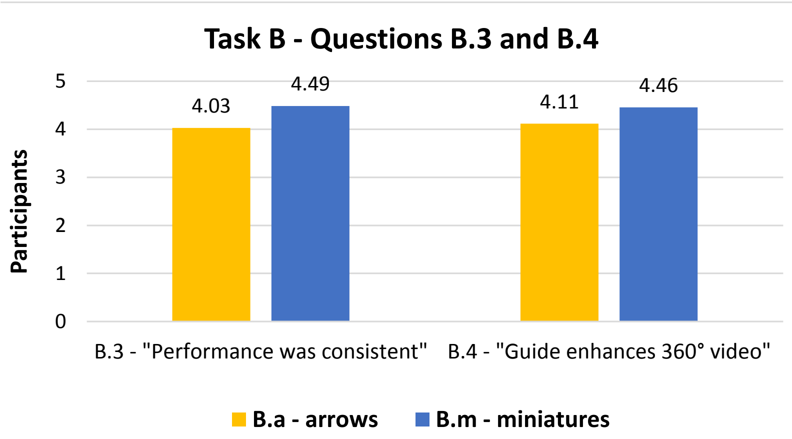
Results In the graph 5.8, the mean value for sentence B.1 and B.2 is presented for each video. For question B.1, 57% of the participants said they did not know where to find new markers in video B.n (no guides), whereas for the other videos, they had better awareness of where to turn to. In regards to B.2, which shows that 62% of users felt it was easy to find the markers in video B.m (miniatures), lowering to 54% in video B.a (arrows), and finally only 14% reported the same for video B.n (no guides). Graph 5.9 present the users' opinions on the performance and efficiency of the used guides, which show an improvement for the miniatures, comparing to the previous task, and the graph 5.10 presents the average of markers found for each video. These graphs present the arithmetic mean of the answers for each video, for further details the graphs in A.3 presents the number of people that provided each answer option for every question and every video.



Question B.1	1st Quartile	Median	3rd Quartile
No Guides	1	1	2
Arrow	3	4	5
Miniatures	4	4	5

Question B.2	1st Quartile	Median	3rd Quartile
No Guides	2	3	4
Arrow	3	4	5
Miniatures	4	5	5

Figure 5.8: Task B: Question B.1 and B.2.

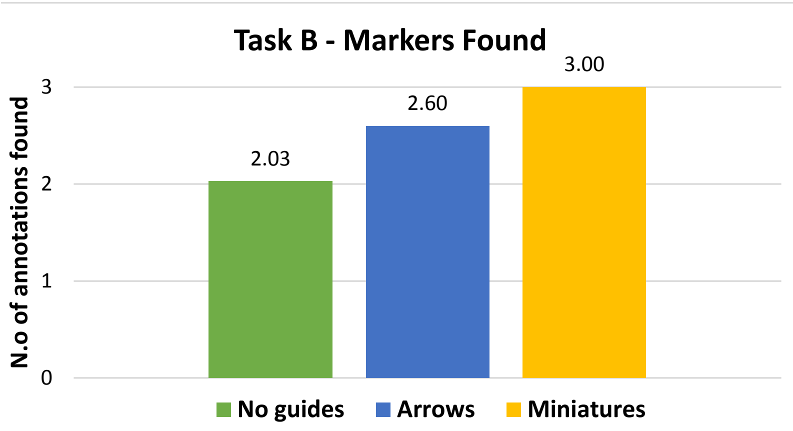


Question B.3	1st Quartile	Median	3rd Quartile
Arrow	4	4	5
Miniatures	4	5	5

Question B.4	1st Quartile	Median	3rd Quartile
Arrow	4	4	5
Miniatures	4	5	5

Figure 5.9: Task B: Question B.3 and B.4.

Analysis Following the first task, some participants reported through spontaneous comments that they found it easier to understand the floating guides, particularly the miniatures, because they had



Markers Found	1st Quartile	Median	3rd Quartile
No Guides	2	2	3
Arrow	2	3	3
Miniatures	3	3	3

Figure 5.10: Task B: Markers Found.

better understanding of how the guides worked and looked like. Although the arrows seemed to be the preferred guiding method in Task A, the preference slightly shifted towards the miniatures in this task, which some users mentioned that due to the textual information in the miniature, it was immediately clear that a new marker appeared, and the blue line helped locate the marker.

In the video without guides (B.n), only 37% of the users were able to find all the markers in the scene. Once the video started, similarly to what happened with the first task, if they chose to turn to the opposite side of the first marker, they almost always missed it. Even some of those that turned correctly did not see the marker, as they moved away before it became visible. It is also relevant to point out that the users reported more difficulty in finding the markers in this scenario, giving the justification that the videos from this task had richer visual information, when compared with the ballet videos.

Using the arrows as a guide in video B.a had comparatively better results in performance of the task, where 65% of the participants were able to find all markers. However, in this scenario, the arrows proved to be confusing when more than one of them was available, particularly when they appeared in opposite sides, as the users felt confused for having to choose to follow one of the guides without additional information.

The video with miniatures (B.m) had the highest success rate, with 100% of all participants finding the three markers in the video. Whereas in the first task, the miniatures were considered less effective and slightly confusing, in this task, the users were able to understand and follow these guides with relative ease.

5.5.4 Task C Experiment

Description The third task of the test sessions was the first experience for the users with the narrative guides: the vignette and the lateral lights. Each video had three points of interest to be found: for the video without any guides, C.n, the users had to find the ice rink, the carrousels, and the ticket office, which are depicted in the figures 5.11 a), 5.11 b) and 5.11 c), respectively. In the video with lateral lights, C.ll, the points of interest were an ice ramp, another carrousel and a Santa Claus' house, shown in figures 5.11 g), 5.11 h) and 5.11 i). Lastly, the video with vignette, C.v, had the same ice rink as the first described video, although further away, bumper cars, and the bus or metro station near Casa da Música, which are presented in figures 5.11 g), 5.11 h), and 5.11 i).

The users were placed at the same position before the start of every video. They were informed that the video contained a narration which mentions a few points of interest that are present in the scene, and they were instructed to look for the referred point of interest within the scene, after which they should indicate that they had successfully found the narrated point. Two questions were asked after each video, and afterwards two other questions regarding the guides used for the task.

- **Sentence 1 (C.1)** - After each video: "It was easy to find the narrated points of interest".
- **Sentence 2 (C.2)** - After each video: "I was confident that I had found the narrated point of interest."
- **Sentence 3 (C.3)** - At the end of the task, regarding the guides: "The guide's performance was consistent with my expectations for this task."
- **Sentence 4 (C.4)** - At the end of the task, regarding the guides: "Comparing to video without guides, the guide enhanced the experience of the video in this context."

Results The mean value of the answers to sentence C.1 and C.2, which vary from 0 to 5, are depicted in graph 5.12, and the participants' evaluation on the narrative guides' performance and efficiency is depicted in graph 5.13. More detailed graphs are provided in appendix A.4, displaying the number of people that chose each possible answer option for the questions of this task.

Analysis The video with lateral lights (C.ll) had highest rate of success, with 34 out of the 35 participants being able to identify all three points of interest, which decreased to 21 users in the video with the vignette (C.v), and lastly, only 12 people, 34%, identified all narrated points of interest without any guides in video C.n. Although 97% of users identified all mentioned objects in the video with the lateral lights, the reported levels of confidence were actually inferior, when comparing to the vignette effect. Some users justified this by saying that they expected feedback once they turned towards the correct area, instead of the lights simply disappearing.

Analyzing each video's experiences, and starting with the video without guides, the users explored this video's scene more, examining different areas possibly related to the narration, to

a) Video C.n, Ice Skate Ring.

b) Video C.n, Carrousels.

c) Video C.n, Ticket Office.

d) Video C.ll, Ice Ramp.

e) Video C.ll, Carroussels.

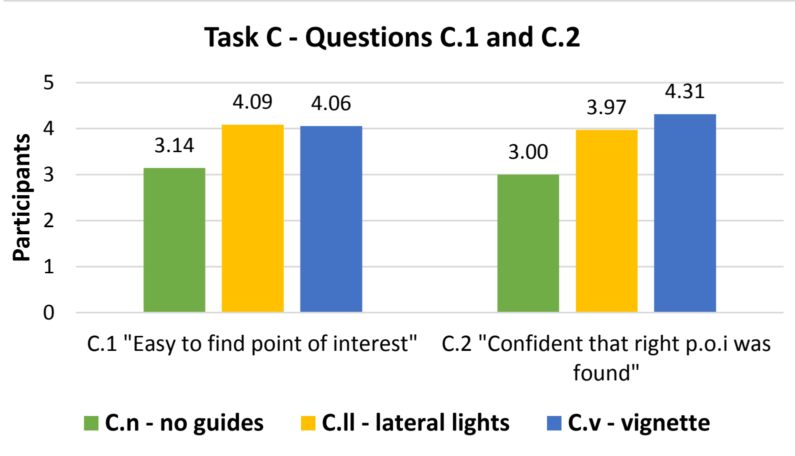
f) Video C.ll, Santa's House.

g) Video C.v, Ice Skate Ring.

h) Video C.v, Bumper Cars.

i) Video C.v, Bus/Metro Station.

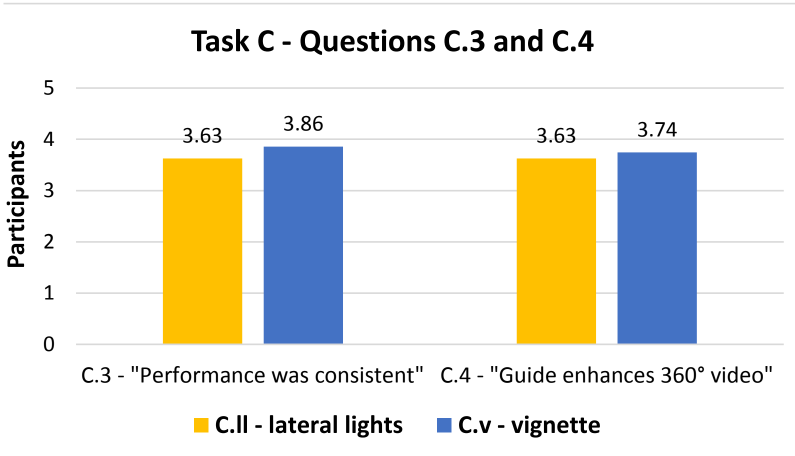
Figure 5.11: Task C - Screenshots of all three points of interest for every video.



Question C.1	No Guides	Lights	Vignette
1st Quartile	2	4	4
Median	3	5	4
3rd Quartile	5	5	5

Question C.2	No Guides	Lights	Vignette
1st Quartile	2	4	4
Median	3	4	5
3rd Quartile	5	5	5

Figure 5.12: Task C: Question C.1 and C.2.



Question C.3	Lights	Vignette
1st Quartile	3	3
Median	4	5
3rd Quartile	5	5

Question C.4	Lights	Vignette
1st Quartile	3	4
Median	4	4
3rd Quartile	5	5

Figure 5.13: Task C: Question C.3 and C.4.

figure out which was correct. Regarding the video with vignette, several users reported that they were unable to identify the first point of interest (42% did not identify the first PoI) because the clear area of the vignette was too wide-ranging, and so they were not able to understand which building they should be looking at. This was also worsened by the fact that the ice rink is further from the camera when comparing to other points of view. Others also mentioned that they were unaware of the presence of the vignette effect during the first 20 seconds, for the same reason, and only when this effect moved towards the second point of interest did they realize its existence.

Finally, the lateral lights were the preferred guiding method for this task, as their intent was clear while not being very invasive. This characteristic, defined by some users as positive, had a negative impact for others, which either did not even notice the lights, or confused them with a possible defect in the video's recording. In two specific sessions, the distance between the HTC Vive's lenses was incorrectly adjusted for the user, which meant that the lights were out of view. This was rectified before advancing towards the last task, and in the questionnaires, the questions regarding this video were answered with "0 - Not applicable".

5.5.5 Task D Experiment

Description For the last task of the test, the users could freely explore the video, as the purpose of this task was to examine their ability to retain information. As the previous task, the narrative guides are used in this task, D.ll using the lateral lights, and D.v using the vignette, while D.n has no guides.

The videos used, which portrayed a concert, had rich visual and auditory information, besides the text presented in the two markers. Specifically, the user could focus on the maestro, the choir, the pianist, the soloist, all dispersed in the surrounding area, or even observe the details of the church's architecture, more visible upwards. Once a user found a marker, they were requested to read the text to the end, to ensure that they could read it entirely. At the end of the video, the participants were asked a few questions regarding what could be seen in the video:

- Question 1 (D.1) - When the video started, was the choir singing?
- Question 2 (D.2) - Was the maestro giving instructions?
- Question 3 (D.3) - Was someone playing the piano?

Additionally, there were questions specific for each video regarding the information on the markers, one about the song being interpreted and another either about the name of the choir, the maestro, or the pianist. For each of these details, the user would be asked if they remembered the name, and then given 2 possible answers, one correct and one wrong. As an example, in the case of the video with the vignette effect:

- Question 4 (D.4) - What was the name of the maestro?
- Question 5 (D.5) - Was the maestro called António Ferreira? (Correct)

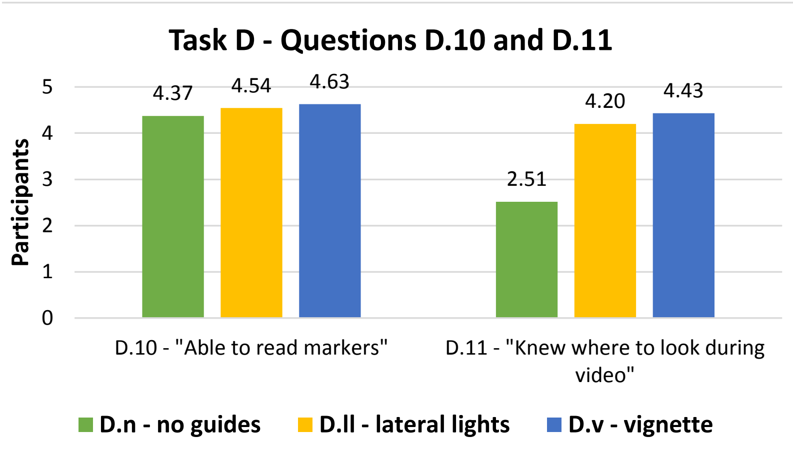
- Question 6 (D.6) - Was the maestro called António Pereira? (Incorrect)
- Question 7 (D.7) - What was the name of the song?
- Question 8 (D.8) - Was it "Et Magnificat"? (Incorrect)
- Question 9 (D.9) - Was it "Et Misericordia"? (Correct)

The two markers that presented these details in each video appeared apart from one another and consecutively, so that the users would have the need, but also the time, to explore the video's scene. Besides these questions, as it was done in the other tasks, the users had to answer two additional questions after each video, and later on, at the end of the task, two final questions about the guides.

- **Sentence 10 (D.10)** - After each video: "I was able to read the annotations clearly until the end".
- **Sentence 11 (D.11)** - After each video: "I knew where I was supposed to look at throughout the video."
- **Sentence 12 (D.12)** - At the end of the task, regarding the guides: "Comparing to video without guides, the guide enhanced the experience of the video in this context."
- **Sentence 13 (D.13)** - At the end of the task, regarding the guides: "The guide was not intrusive for the video's experience"

Results The answers to the questions about the user's experience, D.10 to D.13, are depicted in graph 5.14, and 5.15, while the questions regarding the visual and auditory memory of the video's content are in graph 5.16. The remaining questions are related to the information on the markers, where D.4 and D.7 are questions without hints, and D.5, D.6, D.8 and D.9 provide a false option and a correct option. As such, the following tables present the results of these questions by separating into **correct answers without hints**, referring to users that answered correctly to D.4/D.7, **correct answers with hints**, relative to users that did not answer correctly to D.4/D.7, but were able to get the follow-up questions right, and lastly, **not correct answers**, which encompasses the remaining combination of answers. Table 5.17 relates to the questions regarding the marker that provides the name for the entity or person in the video (D.4 to D.6), and table 5.18 presents the results regarding the marker with the identification of the song being performed. Similarly to the other tasks, appendix A.5 contains more detailed graphs on the answers for each question of this task.

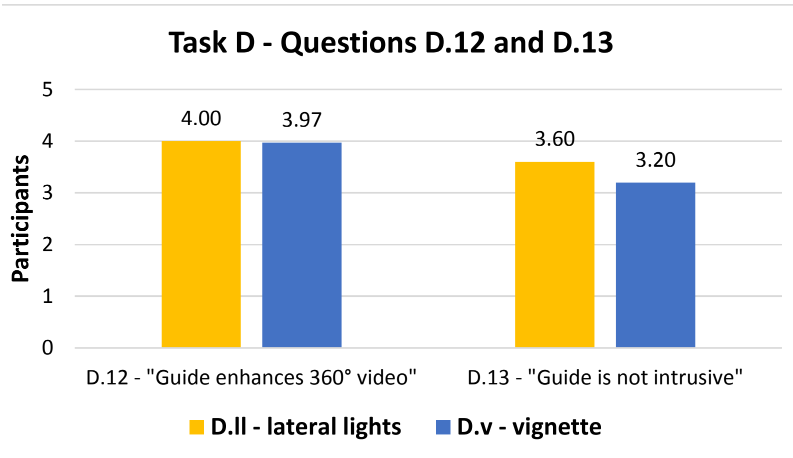
Analysis In the video without guides (D.n), the users explored the scene more freely and, although there were no hints to where to look and so they were not sure where to focus, as demonstrated in graph 5.14, almost all, 29 out of 35, were able to read the two markers to the end. In regards to the visual memory questions, D.1 to D.3, represented in the first table in 5.16, had the worst results for D.3: "Was someone playing the piano?". This could be due to the fact that



Question D.1	No Guides	Lights	Vignette
1st Quartile	5	5	5
Median	5	5	5
3rd Quartile	5	5	5

Question D.2	No Guides	Lights	Vignette
1st Quartile	1	4	4
Median	2	4	5
3rd Quartile	4	5	5

Figure 5.14: Task D: Question D.10 and D.11.



Question D.3	Lights	Vignette
1st Quartile	4	4
Median	5	4
3rd Quartile	5	5

Question D.4	Lights	Vignette
1st Quartile	3	2
Median	4	3
3rd Quartile	5	4

Figure 5.15: Task D: Question D.12 and D.13.

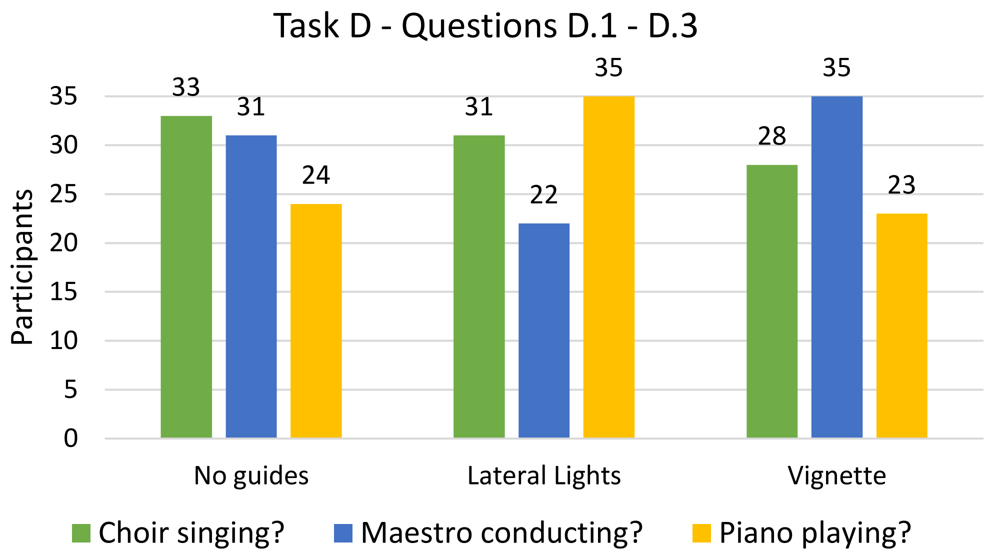


Figure 5.16: Task D - Questions D.1-D.3 (Visual Memory), correct answers with hints, without hints, and wrong answers.

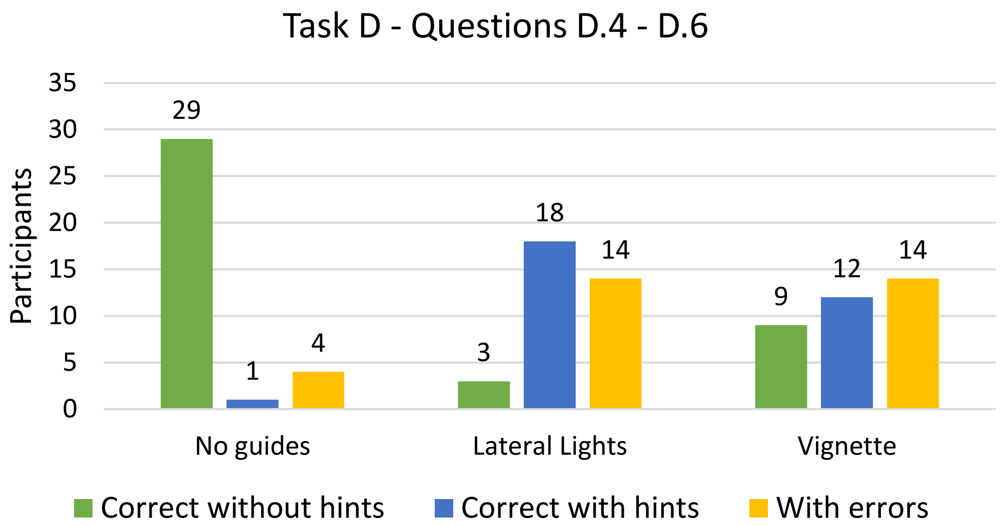


Figure 5.17: Task D - Questions D.4-D.6 (Name Memory), correct answers with hints, without hints, and wrong answers.

the pianist did not have a marker placed near it, unlike the maestro and choir, and so, the users' attention on the pianist would not have lasted as long as on the others. Most users were able to remember the choir's name, but not so much with the song's name, which can be explained by the fact that, without guides, some were not aware of the marker's existence, and only found it closer to its disappearance, not giving enough time to read the marker's text.

The video with lateral lights (D.ll) had less movement from the users, as they typically followed the direction of the lights and remained in the respective area for a while. Even so, some

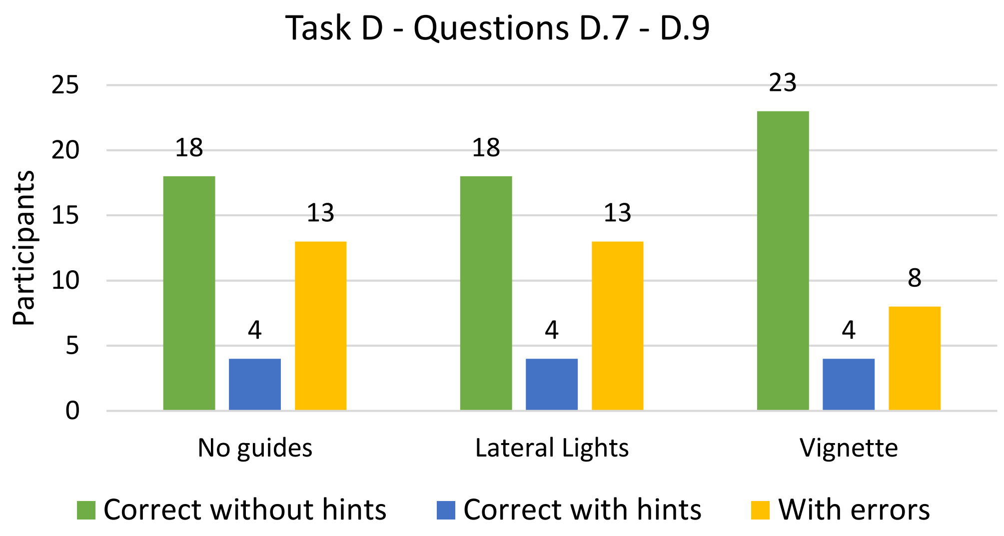


Figure 5.18: Task D - Questions D.7-D.9 (Song-related Memory), correct answers with hints, without hints, and wrong answers.

participants commented that they felt the liberty to look around even when the light was active, as they saw it as a recommendation, not so much as an obligation. The visual memory questions, shown in graph 5.16 had slightly worse results comparing to the video without guides, particularly in D.2: "Was the maestro giving instructions?", as some never turned around to see the maestro, the subject of the question, seeing that it was not the focus of the lateral lights at any time during the video. This video's question about the name of the pianist had the worst results, comparing to the other videos, shown in graph 5.17. These results may not necessarily be due to the guiding method used, but instead to an inherent difficulty in recalling proper names. In fact, while struggling to remember his name, some were able to recall the remaining information of the pianist's marker, such as his job titles and entities he had worked with. This recalling issue has been a focus of attention for cognitive psychologists, as Brennen [Bre93] mentions, who hypothesizes that, because people's names have a wider range of possible phonologies, comparing to common names, they will be more difficult to recall than, for example, profession names. The last memory question, regarding the song of the video, had very similar results to the video without guides, as seen in the tables 5.18.

Focusing now on the video with the vignette (D.v), as it was pointed out in the previous task, the participants felt deterred from exploring the video due to the vignette effect, and this condition persisted in this task, which lead to them being forcefully guided towards the maestro in the beginning, and the choir later on, according to the vignette's movements. Despite the fact that the vignette had the best results in graph 5.14 regarding the user's confidence in where to look at, its movement between areas of interest and its behavior to direct a user towards an out-of-view area were distracting and, at some points, invasive for those who wanted to explore the video. This was apparent from the answers represented in graph 5.15. This condition had an effect in the user's perception of the video's content, has it presents slightly worse results in terms of visual memory, as seen in table 5.16. Nevertheless, with this type of guide, the users had better results in remembering the individual's proper name, the maestro in this case, than the video with lateral lights, shown in 5.17. These results were still not better than the video without guides, which can again be related to the fact that the question was about a proper person's name, and not an entity's name. The final question, regarding the name of the song of video, had the best results for this marker, in graph 5.18, as most users were focusing on the area where the marker appeared in and felt, in general, a resistance to look away, due to the visual nature of the vignette.

5.6 Results' Analysis and Discussion

The performed user tests and their results provided some insight into the influence of the annotations in the user's experience. In videos where there is a point of interest out of the user's view, as was the case in task A, the arrows facilitate the process of finding it, while the miniatures were less effective. Arrows are an extremely common directional icon, meaning that the users do not require a familiarization phase. The positioning of these guides being restricted to the left or right of the view hinders their ability to guide towards marker's located above or below the user, as they create confusion by seemingly guiding towards one or the other side.

For the second task, as it was the second encounter with miniatures, the participants were now more familiar with this annotation, and even preferred it for the task's context. For videos with higher levels of visual and auditory information being passed to the user, annotation guides are even more relevant to pinpoint different points of interest, and at the same time, be aware of one's surroundings. Because the arrows provide no contextual information of the marker they are connected to, indecisiveness occurs when there are more than one arrow in the scene, as the users were troubled to decide which to follow. The miniatures, in contrast, are able to indicate not only the existence of a new marker, but also some indication of the marker's content, which differentiates miniatures from one another. The minimap, in this context, proved to be useful when users were more confused in regards to where to turn, overlooking it when not needed.

Videos with narration, such as the ones used for the task C, benefit from the use of narrative guides, to keep the user aware of the important areas of interest. The lateral lights proved to be an innocuous but effective guide, able to guide the users without removing much of the user's attention on the video and audio. Its discreet nature proved to be harmful at times, as some users did not notice its existence. The vignette, on the other hand, suffers from being too noticeable, ending up being a distraction unless the user kept focused on the area of interest. Although it is more apparent than the lights, it requires some improvements in order to adapt to dimensions of the area of interest, being that it may cover part of larger regions of interest or instead it may cover too much of the surrounding area of smaller regions of interest.

The lights continued to be the favored in the last task, having less issues with it not being noticed by the users, which could be related to the video for this task being darker, and therefore, more contrasting with the lights. The vignette was in this case, for some, even more intrusive, due to the nature of the video which gave the users the urge to explore its scene. Even though, in the most part, the users were able to follow the vignette's indications, it happened that they wanted to ignore it at times to look around, triggering the vignette's animation, a bothersome and distracting effect. From this task, it was clear that videos that depicted actions involving areas of interest spread around the scene, the users would prefer to have the minimal interference from virtual objects, so as to enjoy the full experience of the event as they wish. Otherwise, by pulling the attention of an unwilling user, it creates a distraction, which may lead to the loss of attention and ultimately, the loss of important information. A possible solution for this last issue with the vignette would be to establish a visited state, where the vignette stops its motion or becomes less apparent once the user looks towards the current area of interest.

By analyzing the results of the experiences, one can recognize the impact that virtual overlaying objects can have on the immersion of 360° videos, for either the positive or the negative side. Smaller and more sporadic annotations are preferred, and the need for visual guides increases with the visual complexity of the scene. Annotations that encompass broader areas of the video while, in some way, making those areas less visible can have a restrictive effect and remove some of the user's perceived control over their actions, which is vital for the sense of immersion. From this point on, several improvements may be performed on the tested annotations, basing on the analysis of the user's reports, and additional methods of annotation may be implemented, avoiding certain design characteristics that were evidently harmful for the overall experience. The following chapter 6 further details the future work that would be pertinent for the case of augmented 360° videos.

Chapter 6

Conclusions

Omnidirectional videos, better known as 360° videos, are a type of content that has greater potential of user immersion and engagement than traditional videos, due to its added ability of surrounding a viewer in its scene. The investment in the production of virtual reality content and manufacture of appropriate devices keeps increasing, closely associated with the growth in the average consumer's interest in immersive experiences.

This type of media content has application in contexts which benefit from a more realistic experience and broader view, while not requiring movement in the scene. This is the case of the advertisement industry, which uses 360° videos to present thoroughly the positive features of certain object or space, bringing the user closer to a realistic experience with the object that is focus of the ad. Several sports-related companies are providing to their audience and fans 360° video experiences, bringing them closer to the action without requiring their physical presence. Virtual tours have a similar purpose, as they allow the users to experience a scene which may be otherwise inaccessible.

The news industry has delved into the use of 360° videos for the transmission of news-related events. A branch of journalism focuses on the creation of first-view experiences, to heighten the visual and emotional engagement of the user to the depicted events [dlPWL+10]. As the journalist or content creator plans the story's narrative to be represented by an immersive video, there is additional concern of ensuring that the user is compelled to focus on the scene's areas of interest. The freedom a user has to observe the surrounding scene might interfere with the video's purpose, so proper guidance must be applied, without removing the immersive characteristics of this type of content.

This dissertation focuses on these issues, by tackling the research questions established in chapter 1.2. After researching on previous work relative to virtual experiences and environments, the different stages of 360° content creation, and methods of overlaying and augmenting media objects in various types of content, an omnidirectional visualizer was planned, taking into consideration the different features required to fulfill the objectives of the dissertation and the encompassing project, AV360, as described in chapter 3.

A prototype for the visualizer was developed, whose implementation details were presented in

4. This solution allows users to view 360° videos that have been augmented by different types of annotations, with the purpose of providing more information on the video's topic, or to guide the user through the video. To ensure that this system would be accessible for the majority of users, it was developed as a web application in order to be compatible with computer web browsers, mobile devices, and even virtual reality devices, specifically the HTC Vive.

Also as part of the dissertation objectives, user tests were performed in the final version of the visualizer, to review the system's general performance, as well as the impact of the different implemented annotations on the user's experience. As part of the experiment, the users performed four tasks, each with a different purpose, from finding one and then several points of interest in the video, to following the video's plot line, and reading information regarding the video's topic. The results of the 35 performed tests are discussed in the previous chapter 5, which showed that, in general, the users were able to complete the tasks more easily with the annotation guides placed in the videos. The tested annotations had better impact depending on the context of the video, which could be improved with a upgraded version of these objects.

The objectives of the project, in the context of this dissertation as described in chapter 1.3, were covered by the carried out work, which is detailed throughout this document. From the evaluation phase of the project, it was possible to identify components in the system that could be improved in the future, which are explained in the following sections.

6.1 Future Developments

The user tests served to evaluate the current version of the annotations, and through the provided comments of the participants, a few performance improvements were determined for each annotation. The **arrows** were the preferred guide for Task A, due to its familiar design, still, some users reported some issues with its look and behavior. From the recorded comments, a new version of the arrow should be created of smaller size, with a design that adapts with the environment of the video, so that it is better incorporated into the video, and not so visually intrusive. The **miniatures** had better results in the task B, as participants had the previous experience with this annotation from task A, and it is more effective in videos with more than one point of interest. This guide created some initial confusion with the users, due to it being too similar to the marker it is pointing to. Its design should be altered to differentiate it from regular markers, and the guiding line should a more prominent be more prominent, changing it thickness or color.

For both floating guides, their appearance should change more noticeably after the user finds the associated marker, as it was clear from the tests that the users did not notice the implemented changes for this situation. Some users mentioned that, for this case, it would be adequate if the guides disappeared altogether, as they would no longer be needed, and the minimap would be sufficient to indicate the marker's location. Additionally, their positioning in the lateral areas of the screen or view also created some confusion, so a new positioning method could be implemented and tested where these guides move around the entire frame, including the top and bottom areas. This would naturally require alterations in the subtitles' and minimap's position, to avoid overlapping.

Regarding the narrative guides, a few alteration recommendations arose from the test sessions. Starting with the **lateral lights**, this specific annotation worked well for both task C and D, however, there were some cases of users not noticing them in the beginning or even until the end of the video, particularly in task C. This was possible due to the bright scenario in this task, as the lights did not have enough contrast to be easily seen. For this reason, the lights should adapt their color to the area they are superimposing. Also, alike the floating guides, this annotation could be placed around the frame of view, to indicate areas of interest in any direction.

The **vignette** effect had very mixed reactions, but it was evident that, in its current state, it can be intrusive for the immersive video's experience, particularly in videos where the users want to freely explore the video's scene, as was the case in task D. This narrative guide should be used only on some specific contexts, such as virtual tours with a narration to follow, as it would assist in following the indications of the narration. The way it creates contrast between the area of interest and its surrounding, by darkening around it, could be enhanced so that it does not conceal potentially interesting details from the other objects in the scene. A possible method for this is changing the blending mode[1] of the vignette's layer, using the color data of the video's image to calculate the overlaying color, again with the purpose of focusing on the area of interest. Alternative methods of drawing attention, when the user is looking away from the area of interest, should be researched and tested, taking into consideration its distracting effect on the users.

6.2 Future Work

The prototype developed in the context of this dissertation hopefully established the basics that can be used for future implementations. Coupled with the various improvements suggested in the previous section, there are some features that would be relevant to implement, outside the scope of this dissertation. As described in chapter 3.5, the visualizer may be expanded to include the requirements for the content creators as users of the system, so that they are able to submit their videos, edited with the immersive editing tool.

The research and implementation of new types of annotations, as well as interaction techniques, should be continued, to offer a greater variety of annotations that may be useful for different contexts. This process should always consider the importance of compatibility with different devices, by keeping it accessible to the more common devices, the computer and mobile devices, and it should follow a user-centered design, testing newly implemented annotations with several users to ensure it does in fact complement the video's experience, without negative implications.

Currently, the visualizer does not support spatial audio, also known as 3D audio, which could contribute for the user's sense of presence and immersion in the video's scene. This requires

[1]Example of layer blending modes in Photoshop. https://helpx.adobe.com/photoshop/using/blending-modes.html

research on the requirements and characteristics of this type of audio, and also on available tools to implement this feature on the system.

Efforts could be made so that the visualizer could be embedded into external sites. This would work towards making the system more accessible for both content creators and consumers, by integrating already existing websites.

References

[ABF+01] Ronald Azuma, Yohan Baillot, Steven Feiner, Simon Julier, Reinhold Behringer, and Blair Macintyre. Recent Advances in Augmented Reality. *Ieee Computer Graphics And Applications*, 2001(December), 2001.

[AEBD17] Lemonia Argyriou, Daphne Economou, Vassiliki Bouki, and Ioannis Doumanis. Engaging Immersive Video Consumers: Challenges Regarding 360-Degree Gamified Video Applications. In *Proceedings - 2016 15th International Conference on Ubiquitous Computing and Communications and 2016 8th International Symposium on Cyberspace and Security, IUCC-CSS 2016*, pages 145–152. IEEE, dec 2017.

[AF03] Ronald Azuma and Chris Furmanski. Evaluating label placement for augmented reality view management. In *Proceedings - 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2003*, pages 66–75. IEEE Computer Society, 2003.

[AHTL16] Ferran Argelaguet, Ludovic Hoyet, Michaël Trico, and Anatole Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *Virtual Reality (VR), 2016 IEEE*, pages 3–10. IEEE, 2016.

[AI16] Deepak Akkil and Poika Isokoski. Gaze Augmentation in Egocentric Video Improves Awareness of Intention. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, (May):1573–1584, 2016.

[Bab12] James Babu. Video Game HUDs: Information Presentation and Spatial Immersion James Babu Committee: Dr. Evelyn P. Rozanski. 2012.

[BH97] Doug A. Bowman and Larry F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *Proceedings of the 1997 symposium on Interactive 3D graphics - SI3D '97*, pages 35–ff., 1997.

[BMT07] Doug A. Bowman, Ryan P Mcmahan, and Virginia Tech. Virtual Reality: How Much Immersion Is Enough? (Cover story). *Computer*, 40(7):36–43, 2007.

[Bre93] Tim Brennen. The difficulty with recalling people's names: The plausible phonology hypothesis. *Memory*, 1(4):409–431, 1993.

[BSEW16] A. Brown, A. Sheikh, M. Evans, and Z. Watson. Directing attention in 360-degree video. *IBC 2016 Conference*, (March 2017):29 (9 .)–29 (9 .), 2016.

[BTP+17] Andy Brown, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong, and Maxine Glancy. Subtitles in 360-degree video. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '17 Adjunct, pages 3–8, New York, NY, USA, 2017. ACM.

[CB16] James J. Cummings and Jeremy N. Bailenson. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, 19(2):272–309, 2016.

[CI16] Loïc Caroux and Katherine Isbister. Influence of head-up displays' characteristics on user experience in video games. *International Journal of Human Computer Studies*, 87:65–79, mar 2016.

[CNdSR16] Cristiano Carvalheiro, Rui Nóbrega, Hugo da Silva, and Rui Rodrigues. User Redirection and Direct Haptics in Virtual Environments. *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*, pages 1146–1155, 2016.

[Com05] Wikimedia Commons. Photograph taken by a pilot on the vfa-151 of the HUD of a f/a-18c. https://commons.wikimedia.org/wiki/File:Hud_on_the_cat.jpg, 2005.

[dlPWL+10] Nonny de la Peña, Peggy Weil, Joan Llobera, Bernhard Spanlang, Doron Friedman, Maria V. Sanchez-Vives, and Mel Slater. Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News. *Presence: Teleoperators and Virtual Environments*, 19(4):291–301, 2010.

[Eur17] Euronews. A 360° political immersion in martinique. http://www.euronews.com/2017/04/18/360-video-french-election-on-the-caribbean-island-of-martinique, 2017.

[Fer94] Eugene S. Ferguson. *Engineering and the Mind's Eye*. MIT Press, 1994.

[GHS06] Timo Götzelmann, Knut Hartmann, and Thomas Strothotte. Agent-based annotation of interactive 3D visualizations. *Smart Graphics*, pages 24–35, 2006.

[GLT+] Raphael Grasset, Tobias Langlotz, Markus Tatzgern, Dieter Schmalstieg, and Denis Kalkofen. Image-Driven View Management for Augmented Reality Browsers.

[Hab16] Jim Habig. Is 360 video worth it? https://www.thinkwithgoogle.com/advertising-channels/video/360-video-advertising/, 2016.

[HDC+16] Rorik Henrikson, Bruno De Araujo, Fanny Chevalier, Karan Singh, and Ravin Balakrishnan. Multi-Device Storyboards for Cinematic Narratives in VR. *Proceedings of the 29th Annual ACM Symposium on User Interface Software & Technology*, pages 787–796, 2016.

[Her89] Anne Herbert. *Whole Earth Review*, 1989.

[HR92] Lawrence J. Hettinger and Gary E. Riccio. Visually Induced Motion Sickness in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 1(3):306–310, jan 1992.

[HS] Mohammad Hosseini and Viswanathan Swaminathan. Adaptive 360 VR Video Streaming: Divide and Conquer!

[ICK+15] Ioanna Iacovides, Anna Cox, Richard Kennedy, Paul Cairns, and Charlene Jennett. Removing the HUD: the impact of non-diegetic game elements and expertise on player involvement. *CHI PLAY '15: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pages 13–22, 2015.

[ISO] ISO ISO. Standard graphical symbols: Public information symbols (iso 7001: 2007). 2007. *International Standards Organisation (ISO): Geneva, Switzerland.*

[Jør12] Kristine Jørgensen. Between the Game System and the Fictional World. *Games and Culture*, 7(2):142–163, 2012.

[JSI+10] Jacek Jankowski, Krystian Samp, Izabela Irzynska, Marek Jozwowicz, and Stefan Decker. Integrating Text with Video and 3D Graphics: The Effects of Text Drawing Styles on Text Readability. *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1321–1330, 2010.

[JZDH17] Jingwei Huang, Zhili Chen, Duygu Ceylan, and Hailin Jin. 6-DOF VR Videos with a Single 360-Camera. In *Proceedings of the IEEE Virtual Reality*, pages 37–44. IEEE, 2017.

[KB08] Jill Phelps Kern and Cynthia A Brewer. Automation and the Map Label Placement Problem: A Comparison of Two GIS Implementations of Label Placement. 2008.

[Kes16] Martin Kesselman. 2016 Consumer Electronics Show, Las Vegas: Virtual 3D Cameras and More. *Library Hi Tech News*, 33(3):16–19, may 2016.

[KGMQ09] Ji Sun Kim, Denis Gracanin, Kresimir Matkovic, and Francis Quek. iPhone/iPod Touch as input devices for navigation in immersive virtual environments. In *Proceedings - IEEE Virtual Reality*, pages 261–262. IEEE, mar 2009.

[KGS12] Konstantina Kilteni, Raphaela Groten, and Mel Slater. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[KS13] S. Shunmuga Krishnan and Ramesh K. Sitaraman. Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs. *IEEE/ACM Transactions on Networking*, 21(6):2001–2014, dec 2013.

[KW10] Karol Kwiatek and Martin Woolner. Transporting the viewer into a 360^o heritage story: Panoramic interactive narrative presented on a wrap-around screen. *2010 16th International Conference on Virtual Systems and Multimedia, VSMM 2010*, pages 234–241, 2010.

[LAS04] Wilmot Li, Maneesh Agrawala, and David Salesin. Interactive image-based exploded view diagrams. *GI '04 Proceedings of the 2004 Graphics Interface Conference*, pages 203–212, 2004.

[LL82] W. Howard Levie and Richard Lentz. Effects of text illustrations: A review of research. *Educational Communication & Technology*, 30(4):195–232, 1982.

[LRA+07] Wilmot Li, Lincoln Ritter, Maneesh Agrawala, Brian Curless, and David Salesin. Interactive cutaway illustrations of complex 3D models. *ACM Transactions on Graphics*, 26(3):9, 2007.

[LT04] Alex Leykin and Mihran Tuceryan. Automatic determination of text readability over textured backgrounds for augmented reality systems. In *ISMAR 2004: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 224–230. IEEE, 2004.

[MB14] Päivi Majaranta and Andreas Bulling. Eye Tracking and Eye-Based Human-Computer Interaction. In *Advances in Physiological Computing*, chapter 3, pages 39–65. Springer, 2014.

[MD08] Stefan Maass and Jürgen Döllner. Dynamic Annotation of Interactive Environments using Object-Integrated Billboards. In *WSCG*, pages 327–334, 2008.

[MdPV+06] Ingrid Mulder, Henk de Poot, Carla Verwij, Ruud Janssen, and Marcel Bijlsma. An information overload study: using design methods for understanding. *Proceedings of OZCHI'06, the CHISIG Annual Conference on Human-Computer Interaction*, (October 2015):245–252, 2006.

[Mei17] João Meira. Rendering and Editing Tools for Interactive 360° Video, 2017.

[Min95] Mark R Mine. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept*, 1995.

[MJ93] Jackie Moyes and Patrick W Jordan. Icon design and its effect on guessability, learnability, and experienced user performance. *People and computers*, (8):49–60, 1993.

[MK94] Paul Milgram and Fumio Kishino. Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, E77-D(12):1321–1329, 1994.

[MMJ+16] João Meira, João Marques, João Jacob, Rui Nóbrega, Rui Rodrigues, António Coelho, and A. Augusto Sousa. Video Annotation for Immersive Journalism using Masking Techniques. *2016 23° Encontro Português de Computação Gráfica e Interação (EPCGI)*, 2016.

[MSBM95] Richard E. Mayer, Kathryn Steinhoff, Gregory Bower, and Rebecca Mars. A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, 43(1):31–41, mar 1995.

[NC10] Luís A R Neng and Teresa Chambel. Get Around 360º Hypervideo. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 119–122, 2010.

[NCJC15] Rui Nóbrega, Diogo Cabral, Giulio Jacucci, and António Coelho. NARI : Natural Augmented Reality Interface Interaction Challenges for AR Applications. (May), 2015.

[NMH+16] Lasse T. Nielsen, Matias B. Møller, Sune D. Hartmeyer, Troels C. M. Ljung, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology - VRST '16*, pages 229–232, New York, New York, USA, 2016. ACM Press.

[Par] On the worlf map, detailed tourist map of paris with metro. http://ontheworldmap.com/france/city/paris/large-detailed-tourist-map-of-paris-with-metro.jpg.

[PB06] Alex Poole and Linden J. Ball. Eye Tracking in HCI and Usability Research, 2006.

[Per08] Elisa Perego. Subtitles and line-breaks: Towards improved readability. In *Between Text and Image: Updating Research in Screen Translation*, pages 211–223. 2008.

[Row15] Matt Rowell. Stereo vs mono 360° video for VR. http://360labs.net/blog/stereo-vs-mono-360-video-vr, 2015.

[Rum12] Francis Rumsey. *Spatial audio*. CRC Press, 2012.

[SCS04] Henry Sonnet, Sheelagh Carpendale, and Thomas Strothotte. Integrating expanding annotations with a 3D explosion probe. In *Proceedings of the working conference on Advanced visual interfaces*, pages 63–70, New York, New York, USA, 2004. ACM, ACM Press.

[Ste92] Jonathan Steuer. Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4):73–93, 1992.

[SW97] Mel Slater and Sylvia Wilbur. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 1997.

[Sze06] Richard Szeliski. Image Alignment and Stitching: A Tutorial. *Computer Graphics and Vision*, 2(1):1–104, 2006.

[Tal10a] Richard JA Talbert. Peutinger map: seamless whole, in color, with overlaid layers, 2010.

[Tal10b] Richard JA Talbert. *Rome's world: the Peutinger map reconsidered*. Cambridge University Press, 2010.

[TP12] Yuichiro Takeuchi and Ken Perlin. ClayVision: the (elastic) image of the city. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, page 2411, New York, New York, USA, 2012. ACM Press.

[VPC12] Bregtje Van Der Haak, Michael Parks, and Manuel Castells. The Future of Journalism : Networked Journalism. *International Journal of Communication*, 6:2923–2938, 2012.

[VTM10] Valentijn T. Visch, Ed S. Tan, and Dylan Molenaar. The emotional and cognitive effect of immersion in film viewing. *Cognition & Emotion*, 24(8):1439–1445, dec 2010.

[Wik16] Dead Space Wiki. Dead space 2 screenshot05. http://deadspace.wikia.com/wiki/File:Dead_Space_2_Screenshot05.jpg, 2016.

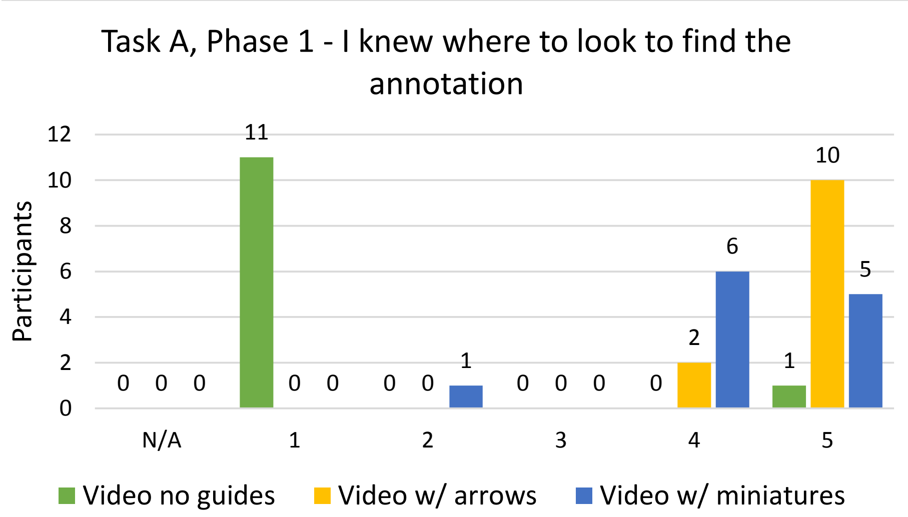
[Wil06] Greg Wilson. Off With Their HUDs!: Rethinking the Heads-Up Display in Console Game Design, 2006.

REFERENCES

Appendix A

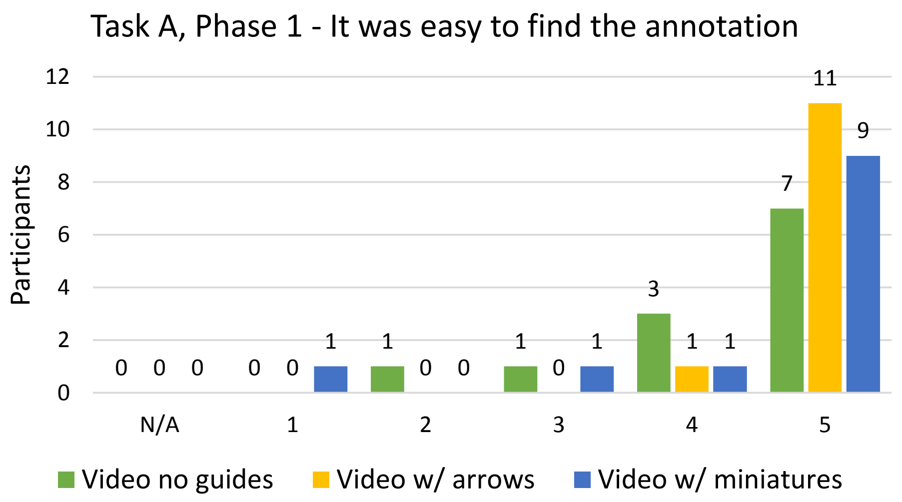
User Experiments' Results

A.1 Task A, Phase 1 Graphs



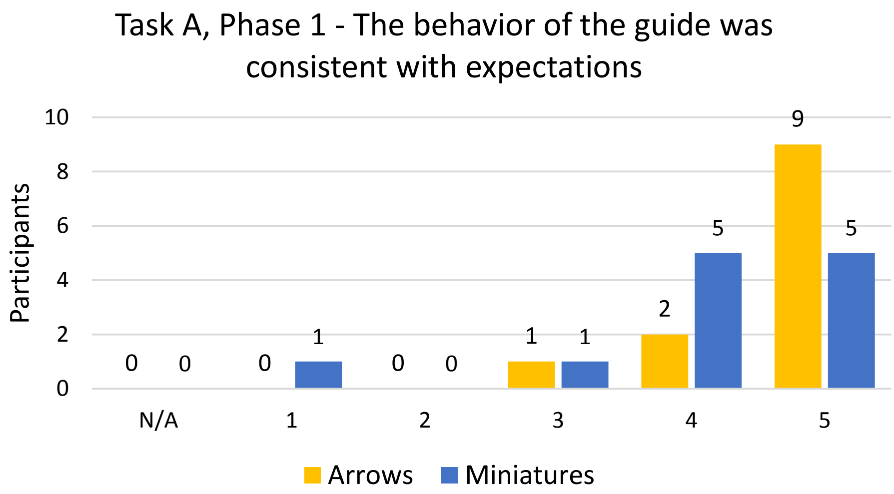
Question A.1	1st Quartile	Median	3rd Quartile
No Guides	1	1	1
Arrow	5	5	5
Miniatures	4	4	5

Figure A.1: Task A, Phase 1: Question A.1



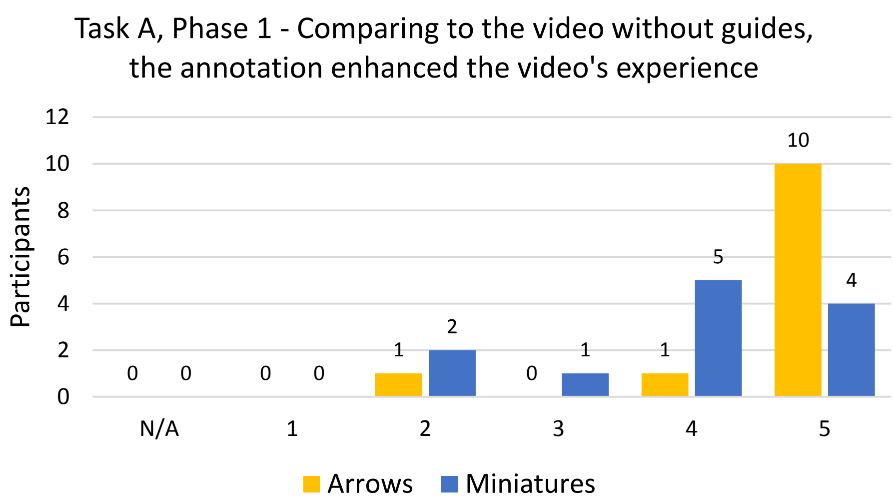
Question A.2	1st Quartile	Median	3rd Quartile
No Guides	4	5	5
Arrow	5	5	5
Miniatures	5	5	5

Figure A.2: Task A, Phase 1: Question A.2



Question A.3	1st Quartile	Median	3rd Quartile
Arrow	5	5	5
Miniatures	4	4	5

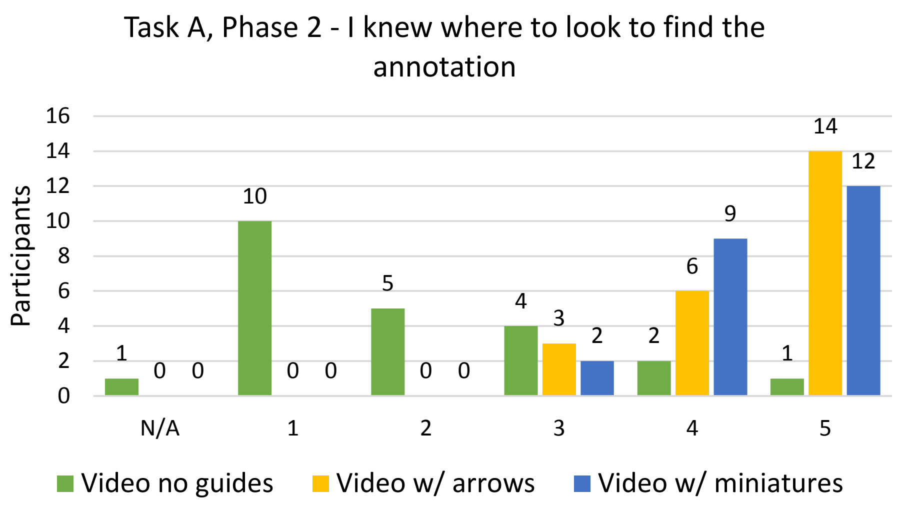
Figure A.3: Task A, Phase 1: Question A.3



Question A.4	1st Quartile	Median	3rd Quartile
Arrow	5	5	5
Miniatures	4	4	5

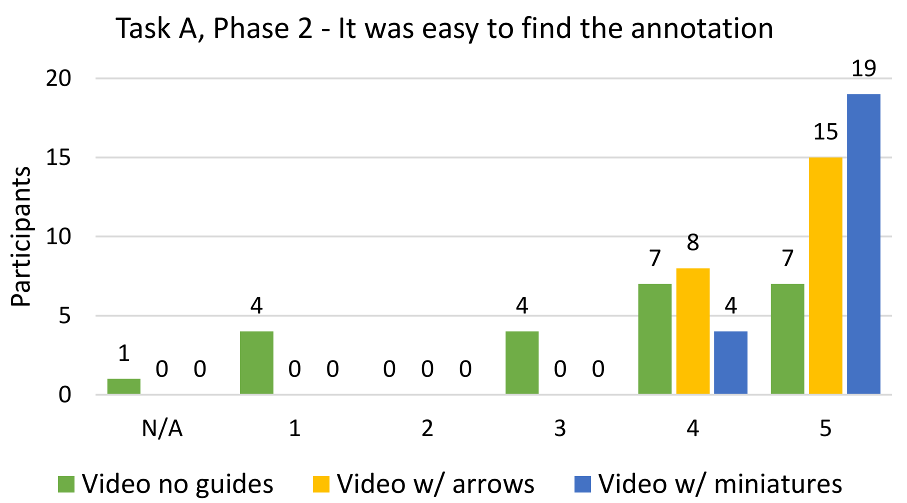
Figure A.4: Task A, Phase 1: Question A.4

A.2 Task A, Phase 2 Graphs



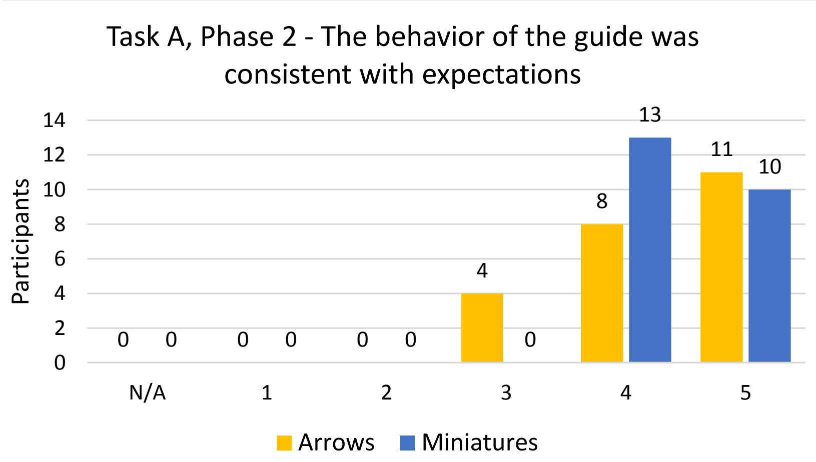
Question A.1	1st Quartile	Median	3rd Quartile
No Guides	1	2	3
Arrow	4	5	5
Miniatures	4	5	5

Figure A.5: Task A, Phase 2: Question A.1



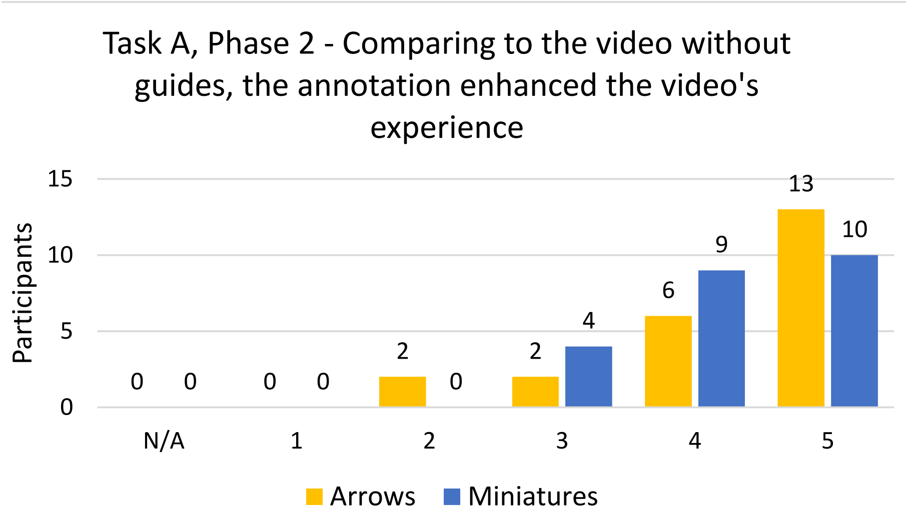
Question A.2	1st Quartile	Median	3rd Quartile
No Guides	3	4	5
Arrow	4	5	5
Miniatures	5	5	5

Figure A.6: Task A, Phase 2: Question A.2



Question A.3	1st Quartile	Median	3rd Quartile
Arrow	4	4	5
Miniatures	4	4	5

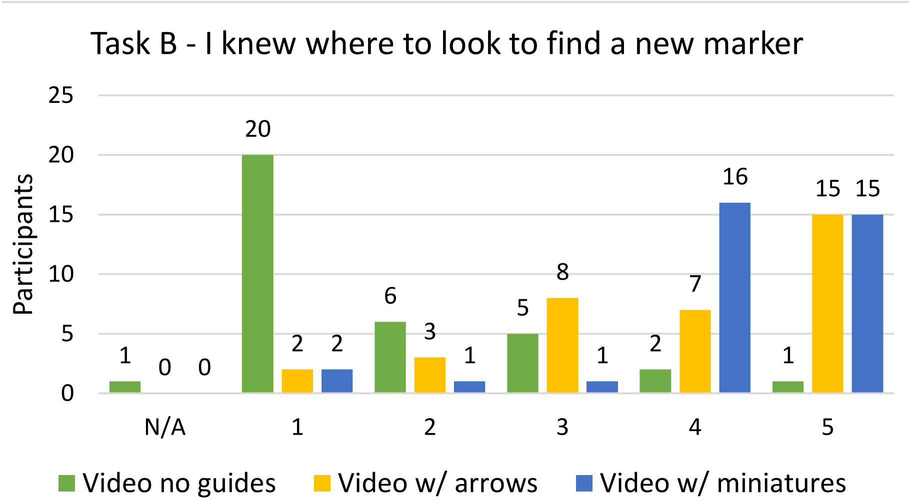
Figure A.7: Task A, Phase 2: Question A.3



Question A.4	1st Quartile	Median	3rd Quartile
Arrow	4	5	5
Miniatures	4	4	5

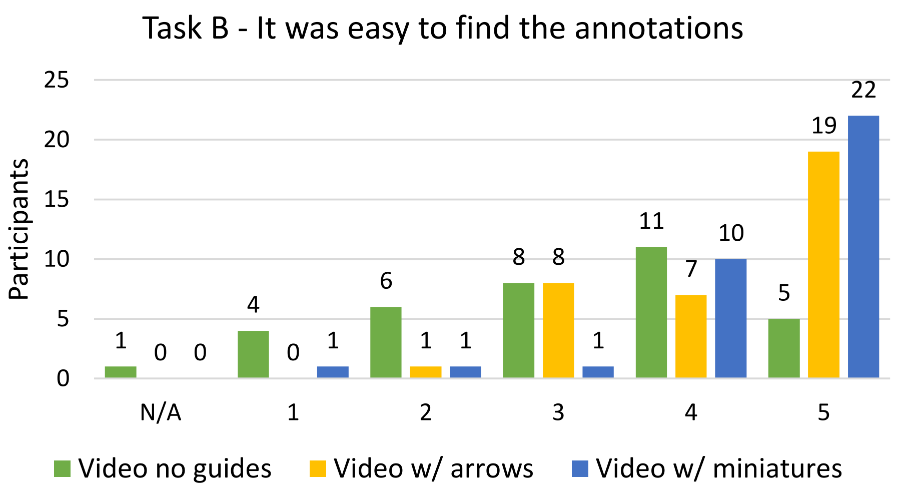
Figure A.8: Task A, Phase 2: Question A.4

A.3 Task B Graphs



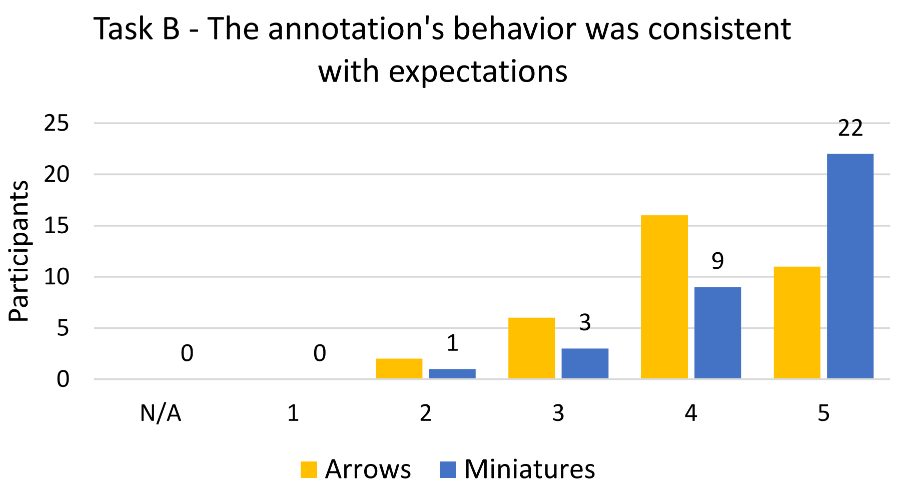
Question B.1	1st Quartile	Median	3rd Quartile
No Guides	1	1	2
Arrow	3	4	5
Miniatures	4	4	5

Figure A.9: Task B: Question B.1



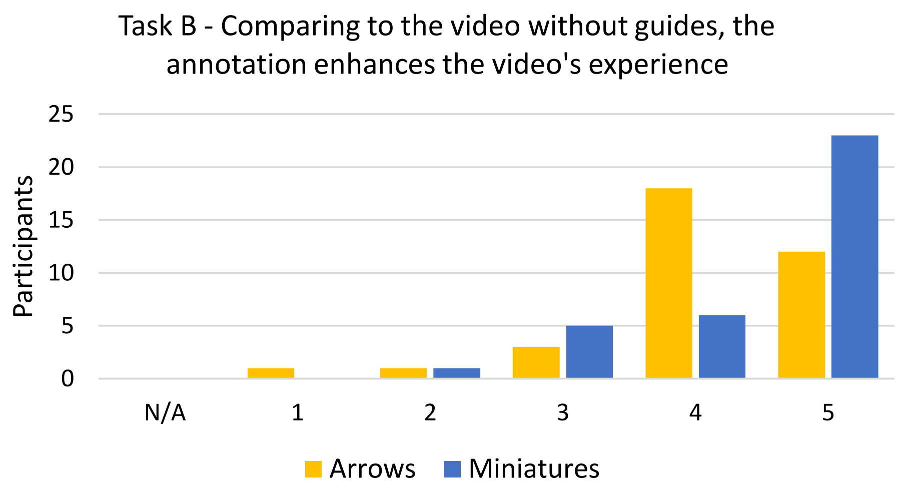
Question B.2	1st Quartile	Median	3rd Quartile
No Guides	2	3	4
Arrow	3	4	5
Miniatures	4	5	5

Figure A.10: Task B: Question B.2



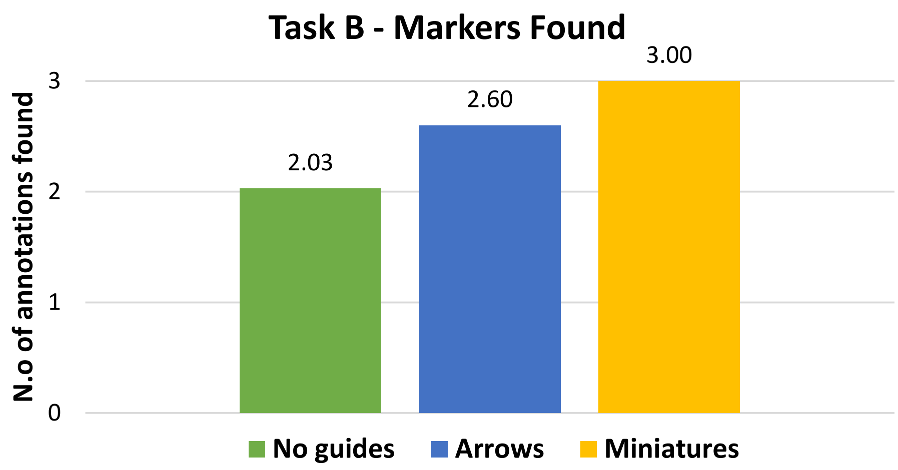
Question B.3	1st Quartile	Median	3rd Quartile
Arrow	4	4	5
Miniatures	4	5	5

Figure A.11: Task B: Question B.3



le	Question B.4	1st Quartile	Median	3rd Quartile
	Arrow	4	4	5
	Miniatures	4	5	5

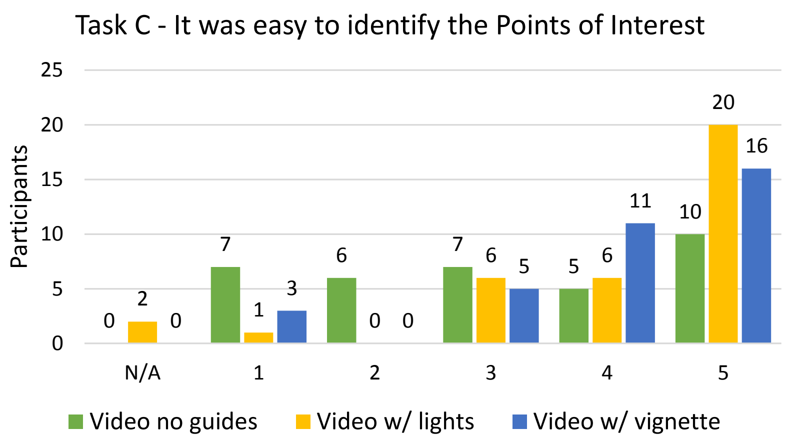
Figure A.12: Task B: Question B.4



Markers Found	1st Quartile	Median	3rd Quartile
No Guides	2	2	3
Arrow	2	3	3
Miniatures	3	3	3

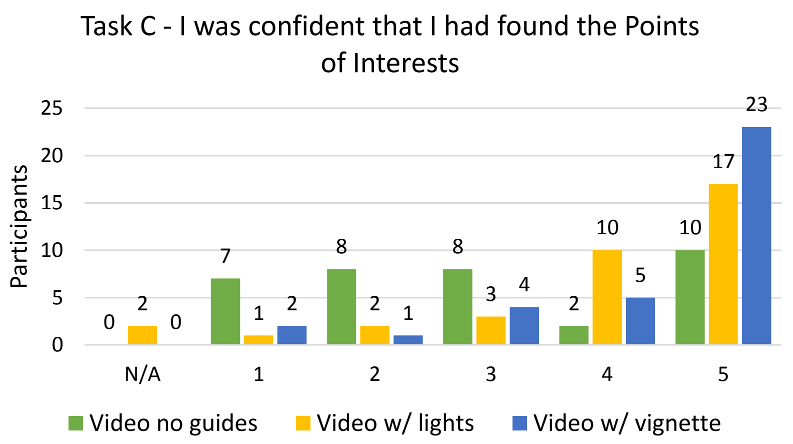
Figure A.13: Task B: Markers Found

A.4 Task C Graphs



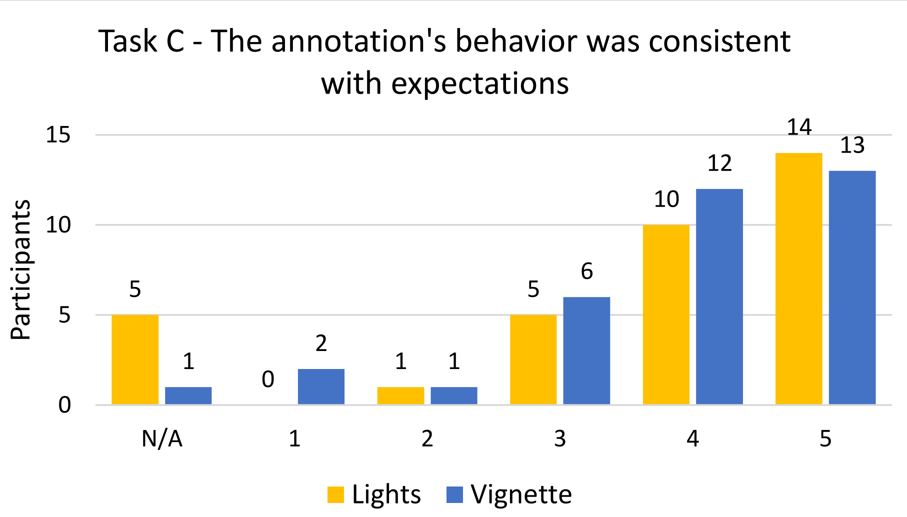
Question C.1	No Guides	Lights	Vignette
1st Quartile	2	4	4
Median	3	5	4
3rd Quartile	5	5	5

Figure A.14: Task C: Question C.1



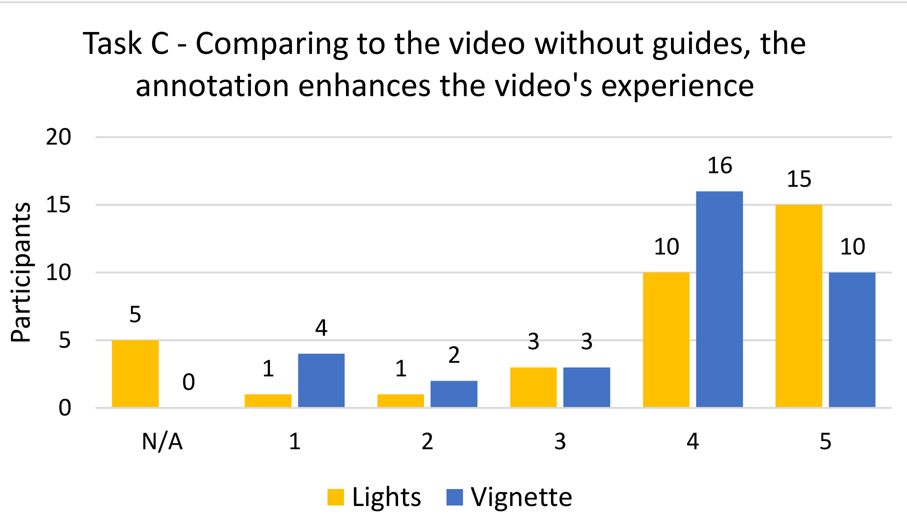
Question C.2	No Guides	Lights	Vignette
1st Quartile	2	4	4
Median	3	4	5
3rd Quartile	5	5	5

Figure A.15: Task C: Question C.2



Question C.3	Lights	Vignette
1st Quartile	3	3
Median	4	5
3rd Quartile	5	5

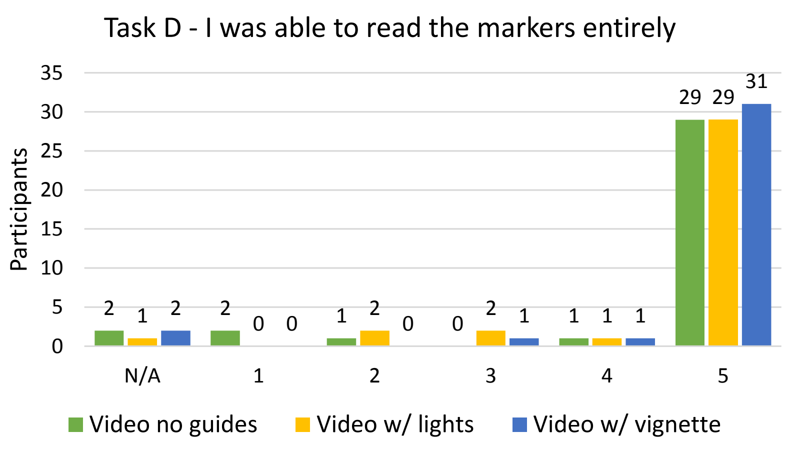
Figure A.16: Task C: Question C.3



Question C.4	Lights	Vignette
1st Quartile	3	4
Median	4	4
3rd Quartile	5	5

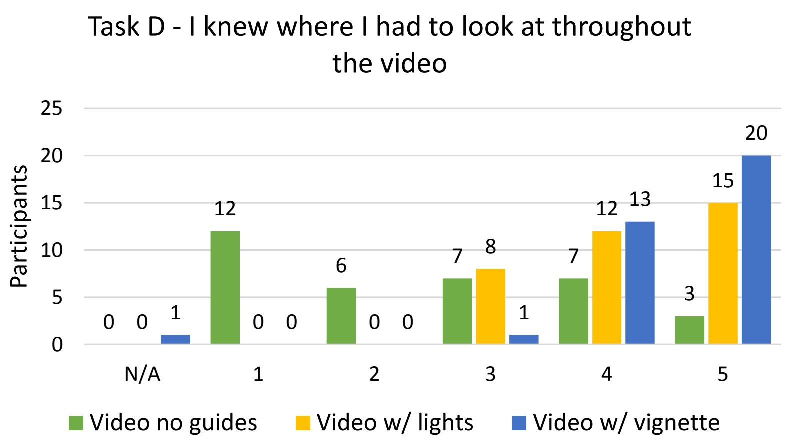
Figure A.17: Task C: Question C.4

A.5 Task D Graphs



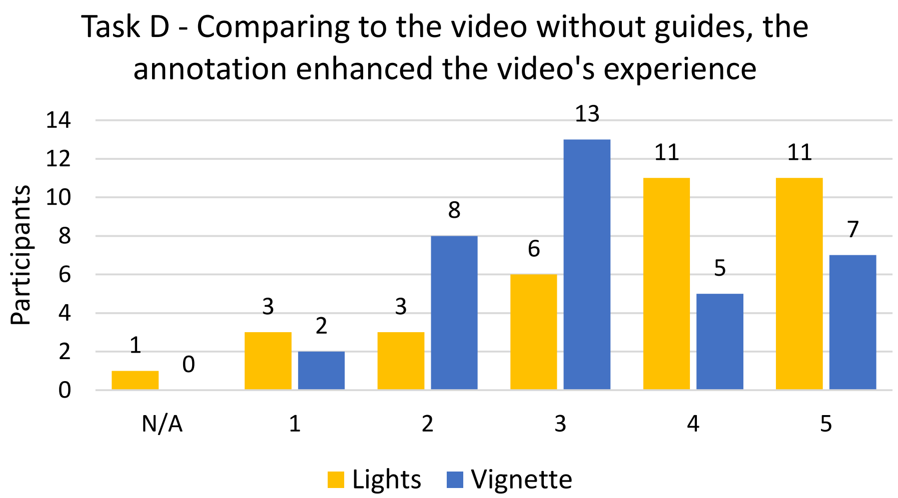
Question D.1	No Guides	Lights	Vignette
1st Quartile	5	5	5
Median	5	5	5
3rd Quartile	5	5	5

Figure A.18: Task D: Question D.1



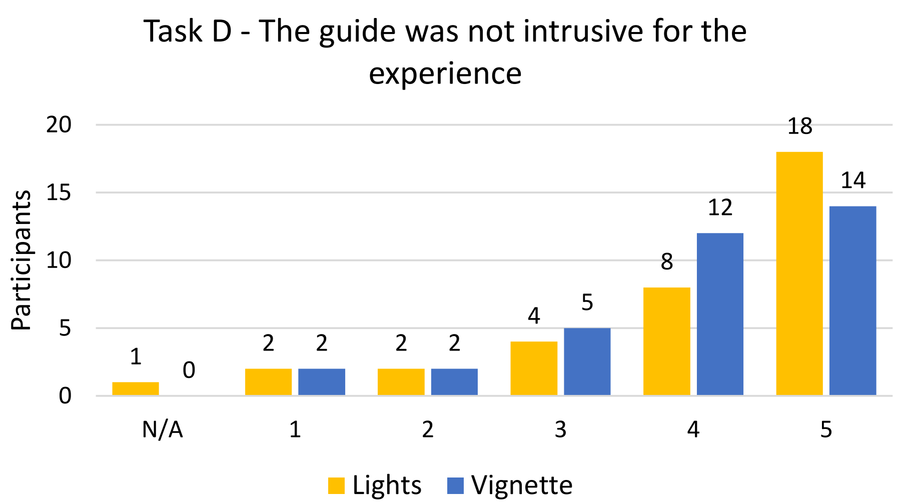
Question D.2	No Guides	Lights	Vignette
1st Quartile	1	4	4
Median	2	4	5
3rd Quartile	4	5	5

Figure A.19: Task D: Question D.2



Question D.3	Lights	Vignette
1st Quartile	4	4
Median	5	4
3rd Quartile	5	5

Figure A.20: Task D: Question D.3



Question D.4	Lights	Vignette
1st Quartile	3	2
Median	4	3
3rd Quartile	5	4

Figure A.21: Task D: Question D.4

Appendix B

Usability Questionnaires

AV360 - Immersive Video Player

*Obrigatório

1. **User ID** *

2. **Test ID** *
   Marcar apenas uma oval.
   - #1
   - #2
   - #3

Sessão de teste AV360 - Dados de utilizador

3. **Data de nascimento** *
   Exemplo: 15 de dezembro 2012

4. **Género** *
   Marcar apenas uma oval.
   - Feminino
   - Masculino
   - Outra:

5. **Experiência com dispositivos de Realidade Virtual** *
   Por favor, escolhe a opção apropriada para cada dispositivo. Inexperiente = nunca utilizou; Um pouco experiente = experimentou algumas vezes; Muito experiente = usou várias vezes, está confiante no seu uso
   Marcar apenas uma oval por linha.

	Inexperiente	Um pouco experiente	Muito experiente
Android Cardboard			
Oculus Rift			
HTC Vive			

Sessão de teste: Tarefa 1

Após cada vídeo:

6. **Sabia para onde olhar para encontrar a anotação.** *

Após o aviso para explorar o vídeo.

Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
1.º vídeo						
2.º vídeo						
3.º vídeo						

7. **Foi fácil encontrar a anotação.** *

Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
1.º vídeo						
2.º vídeo						
3.º vídeo						

Após todos os vídeos:

8. **O desempenho do guia foi consistente com as expectativas da tarefa.** *

Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Setas						
Miniaturas laterais						

9. **Comparando com o vídeo sem guias, os objetos visuais melhoraram a experiência do vídeo.** *

Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Setas						
Miniaturas laterais						

10. **Comentários**

11. **Anotações vistas no 1.º vídeo:**
Marcar tudo o que for aplicável.

- [ ] Kilos
- [ ] Catarina Rodrigues
- [ ] Mariana PTKS

12. **Anotações vistas no 2.º vídeo:**
Após o aviso para explorar o vídeo.
Marcar tudo o que for aplicável.

- [ ] Nuno Sarmento
- [ ] Bárbara R.
- [ ] Dylan Silva

13. **Anotações vistas no 3.º vídeo:**
Após o aviso para explorar o vídeo.
Marcar tudo o que for aplicável.

- [ ] Royal Studio
- [ ] Andy Calabozo
- [ ] Catarina Rodrigues

Sessão de teste: Tarefa 2

Após cada vídeo:

14. **Sabia para onde olhar para ver anotações novas.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
1.º vídeo	○	○	○	○	○	○
2.º vídeo	○	○	○	○	○	○
3.º vídeo	○	○	○	○	○	○

15. **Foi fácil encontrar as anotações.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
1.º vídeo	○	○	○	○	○	○
2.º vídeo	○	○	○	○	○	○
3.º vídeo	○	○	○	○	○	○

Após todos os vídeos:

16. **O desempenho do guia foi consistente com as expectativas da tarefa.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Setas						
Miniaturas laterais						

17. **Comparando com o vídeo sem guias, os objetos visuais melhoraram a experiência do vídeo.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Setas						
Miniaturas laterais						

18. **Comentários**

Sessão de teste: Tarefa 3

A preencher pelo observador:

19. **Pontos de interesse vistos no 1.º vídeo:**
Marcar tudo o que for aplicável.

- [ ] Pista do gelo
- [ ] Carrosséis
- [ ] Bilheteira

20. **Pontos de interesse vistos no 2.º vídeo:**
Marcar tudo o que for aplicável.

- [ ] Rampa de gelo
- [ ] Montanha-russa
- [ ] Casa do Pai Natal

21. **Pontos de interesse vistos no 3.º vídeo:**
Marcar tudo o que for aplicável.

- [ ] Pista de Gelo
- [ ] Carrinhos de choque
- [ ] Autocarros

Sessão de teste: Tarefa 3

Após cada vídeo:

22. **Foi fácil encontrar os pontos de interesse referidos pela narração.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
1.º vídeo	◯	◯	◯	◯	◯	◯
2.º vídeo	◯	◯	◯	◯	◯	◯
3.º vídeo	◯	◯	◯	◯	◯	◯

23. **Foi fácil diferenciar as áreas de interesse da área restante.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
1.º vídeo	◯	◯	◯	◯	◯	◯
2.º vídeo	◯	◯	◯	◯	◯	◯
3.º vídeo	◯	◯	◯	◯	◯	◯

Após todos os vídeos:

24. **O desempenho do guia foi consistente com as expectativas da tarefa.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Escurecimento	◯	◯	◯	◯	◯	◯
Luzes laterais	◯	◯	◯	◯	◯	◯

25. **Comparando com o vídeo sem guias, os objetos visuais melhoraram a experiência do vídeo.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Escurecimento	◯	◯	◯	◯	◯	◯
Luzes laterais	◯	◯	◯	◯	◯	◯

26. **Comentários**

Sessão de teste: Tarefa 4

A preencher pelo observador:

27.
Informação obtida do 1.º vídeo:
Marcar apenas uma oval por linha.

	Correcto	Falso	Não soube responder/não tinha a certeza
No início do vídeo, o coro estava a cantar?			
O maestro estava de pé a dar indicações?			
Estava alguém a tocar piano?			
Qual o nome do coro?			
O coro era o Coro Clássico Universitário do Porto?			
O coro era o Coro Orfeão Universitário do Porto?			
Qual era o nome da música?			
O nome da música era "Hossana In Excelsis"?			
O nome da música era "Ding Dong Merrily on High"?			

28.
Informação obtida do 2.º vídeo:
Marcar apenas uma oval por linha.

	Correcto	Falso	Não soube responder/não tinha a certeza
No início do vídeo, o coro estava a cantar?			
O maestro estava de pé a dar indicações?			
Estava alguém a tocar piano?			
Qual o nome da pessoa ao piano?			
A pessoa ao piano chama-se Rui Soares?			
A pessoa ao piano chama-se Rui Santos?			
Qual era o nome do compositor?			
O nome do compositor era John Adams?			
O nome do compositor era John Rutter?			

29. **Informação obtida do 3.º vídeo:**
Marcar apenas uma oval por linha.

	Correcto	Falso	Não soube responder/não tinha a certeza
No início do vídeo, o coro estava a cantar?			
O maestro estava de pé a dar indicações?			
Estava alguém a tocar piano?			
Qual o nome do maestro?			
O maestro chama-se António Pereira?			
O maestro chama-se António Ferreira?			
Qual era o nome da música?			
O nome da música era "Et Magnificat"?			
O nome da música era "Et Misericordia"?			

Sessão de Teste: Tarefa 4

Após cada vídeo:

30. **Consegui ler todas as informações das anotações que viu.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
1.º vídeo						
2.º vídeo						
3.º vídeo						

31. **Sabia para onde olhar ao longo do vídeo.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
1.º vídeo						
2.º vídeo						
3.º vídeo						

Após todos os vídeos:

32. **O desempenho do guia foi consistente com as expectativas da tarefa.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Escurecimento						
Luzes laterais						

33. **Os objetos visuais não eram intrusivos para a experiência do vídeo.** *
Marcar apenas uma oval por linha.

	Não aplicável	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Escurecimento	◯	◯	◯	◯	◯	◯
Luzes laterais	◯	◯	◯	◯	◯	◯

34. **Já tinhas conhecimento anterior sobre o que te foi perguntado?**
Marcar tudo o que for aplicável.

- [ ] Coro
- [ ] 1.ª música
- [ ] Organista
- [ ] 2.ª música
- [ ] Maestro
- [ ] 3.º música

35. **Comentários**

Sessão de teste: Perguntas finais

36. **Em relação à experiência geral com o sistema, responde às seguintes perguntas:** *
Marcar apenas uma oval por linha.

	1 - Discordo totalmente	2	3	4	5 - Concordo totalmente
Era confortável de usar	◯	◯	◯	◯	◯
Era divertido de usar	◯	◯	◯	◯	◯
Era fácil de usar	◯	◯	◯	◯	◯
Tinha confiança a usar o sistema	◯	◯	◯	◯	◯
O comportamento geral do sistema era consistente	◯	◯	◯	◯	◯
Repetiria a experiência	◯	◯	◯	◯	◯

37. **Comentários**

Appendix C

User Experiences' Consent Forms

<u>DECLARAÇÃO DE CONSENTIMENTO</u>

(Baseada na declaração de Helsínquia)

No âmbito da realização da tese de Mestrado do Mestrado Integrado de Engenhar Informática e Computação da Faculdade de Engenharia da Universidade do Port intitulada **Visualising and Interacting with Dynamically Annotated 360-degree Wel based Videos**, realizada pela estudante **Teresa Carla de Canha e Matos**, orientada pe Prof. Rui Rodrigues e sob a co-orientação do Prof. Rui Nóbrega, eu abaixo assinad ____, declaro que compreen a explicação que me foi fornecida acerca do estudo no qual irei participar, nomeadamen o carácter voluntário dessa participação, tendo-me sido dada a oportunidade de fazer a perguntas que julguei necessárias.

Tomei conhecimento de que a informação ou explicação que me foi prestada verso os objetivos, os métodos, o eventual desconforto e a ausência de riscos para a minha saúd e que será assegurada a máxima confidencialidade dos dados.

Explicaram-me, ainda, que poderei abandonar o estudo em qualquer momento, se que daí advenham quaisquer desvantagens.

Por isso, consinto participar no estudo e na recolha de imagens necessária respondendo a todas as questões propostas.

Porto, __ de ___________ de 201_

____

(Participante ou seu representante)

Declaração de Consentimento de Direitos de Imagem

Este documento pretende formalizar o pedido de filmagem, bem como a cedência dos direitos de imagem associados, para desenvolvimento de um vídeo 360º a ser utilizado no contexto do projeto de investigação **AV360** (Augmented Video 360).

Autorizo a equipa de investigação do projeto **AV360** a filmar a minha imagem, bem como a difundi-la no contexto de investigação acima mencionado.

A presente autorização é concedida a **título gratuito.**

E, para que assim conste, assina-se a presente declaração.

Porto, ____ de _____________ de _____.

ASSINATURA:

____

INESCTEC