Utrecht University

MSc Thesis

Master of Game and Media Technology

Investigating the Impact of Timeline Length and Control Panel Vertical Location in 360-Degree Video Players on User Experience and Behavior

Author:
Shiyi Chen[1]
(0537772)

Supervisor:
Dr. Wolfgang Hürst[1]

Second Reader:
Dr. H.J. (Hanna) Hauptmann[1]

[1]Utrecht University, Utrecht, The Netherlands

August 13, 2023

Abstract

Video players for 360-degree videos that are watched in VR HMDs (virtual reality head-mounted displays) commonly follow the interaction design of video players for regular videos that are watched on flat, 2D screens. In the latter case, interaction elements such as buttons to start, pause or stop a video as well as timelines enabling users to scroll through them are usually placed horizontally at the bottom of the video player window where they least obstruct the view of the video. Timelines usually extend to most if not all of the window's width to enable users to easily navigate through a video at a reasonable level of detail.

Yet, immersive VR HMDs are not bound by the physical limitation of a window's width and height but provide a full surround 360-degree display. Therefore, the optimal placement and size of these interface elements is not clear. How wide should a timeline be? Where should it be placed? And should it stay at a fixed position or move with the viewer's field of view when they are turning their heads?

In this thesis, we thoroughly investigate and experimentally evaluate the influence of timeline length and control panel vertical location in a 360-degree video player on user experience and behavior. The results show that while these two factors do not significantly affect the effectiveness of the user experience, the comfort and convenience setting leads to higher emotional preference. Moreover, we discover that user behavior in interacting with the timeline is closely tied to the duration of the jump.

Contents

1 Scientific Paper

This section contains the main results of this thesis in the form of a scientific paper. The paper is self-contained, serving as a full report in the ACM paper format. More detailed information is reported in the Appendix.

Investigating the Impact of Timeline Length and Control Panel Vertical Location in 360-Degree Video Players on User Experience and Behavior

Shiyi Chen
s.chen12@students.uu.nl
Information and Computing Sciences
Utrecht University
Utrecht, Netherlands

Abstract

Video players for 360-degree videos that are watched in VR HMDs (virtual reality head-mounted displays) commonly follow the interaction design of video players for regular videos that are watched on flat, 2D screens. In the latter case, interaction elements such as buttons to start, pause or stop a video as well as timelines enabling users to scroll through them are usually placed horizontally at the bottom of the video player window where they least obstruct the view of the video. Timelines usually extend to most if not all of the window's width to enable users to easily navigate through a video at a reasonable level of detail.

Yet, immersive VR HMDs are not bound by the physical limitation of a window's width and height but provide a full surround 360-degree display. Therefore, the optimal placement and size of these interface elements is not clear. How wide should a timeline be? Where should it be placed? And should it stay at a fixed position or move with the viewer's field of view when they are turning their heads?

In this thesis, we thoroughly investigate and experimentally evaluate the influence of timeline length and control panel vertical location in a 360-degree video player on user experience and behavior. The results show that while these two factors do not significantly affect the effectiveness of the user experience, the comfort and convenience setting leads to higher emotional preference. Moreover, we discover that user behavior in interacting with the timeline is closely tied to the duration of the jump.

1 Introduction

Over the past few years, virtual reality (VR) technology has grown rapidly and gained more attention. This unique technology offers users a sense of immersion and presence. In the field of education, VR has demonstrated significant potential, particularly in the form of 360-degree videos, which provide users with an immersive educational experience that is inaccessible in a realistic environment [13]. Unlike traditional 2D videos, 360-degree videos are captured using specialized panoramic cameras, enabling users to choose their viewing perspectives and immerse themselves in a more interactive visual experience [4, 18]. While 360-degree videos can be viewed on conventional flat screens, the use of VR Head-mounted devices (HMDs) significantly enhances the experience. As a result, dedicated 360-degree video players optimized for VR HMDs have been developed to offer users a truly immersive and interactive viewing experience. [13]. Snelson's review emphasizes the combination of educational 360-degree videos with immersive virtual reality (VR) experiences [17], revealing their potential to enhance the learning experience by creating a sense of immersion, presence, and increased student engagement. However, as the demand for immersive video content continues to grow, ensuring a satisfactory user experience becomes crucial during usage. Existing research falls short in exploring the interactive aspects provided by 360-degree video players [1, 13, 17]. Therefore, our study aims to investigate the impact of interaction design on user experience and behavior in 360-degree video players.

Ensuring a positive user experience is essential for enhancing the effectiveness and usability of 360-degree video players. Prior research has identified potential challenges associated with the use of 360-degree VR videos in educational contexts, including distraction and increased cognitive load [13, 17]. Furthermore, educational materials frequently demand active engagement from students. Similar to the use of textbooks or slides, students often need to jump back and forth, skip some unimportant parts, and review important parts again and again. Thus, we can assume that a 360-degree video player employed within an educational context should enable smooth navigation, intuitive controls, and comfortable viewing to fulfill these demands and optimize the learning experience and outcome.

However, there is a scarcity of research focusing on interaction design specifically tailored for 360-degree video players, making it challenging to identify designs that provide an enhanced user experience. Our study focuses on the control panel of 360-degree video players, which encompasses an interactive timeline, play/pause buttons, and fast-forward/rewind buttons. Designing the length of the timeline and the vertical location of the control panel poses unique challenges in VR environments. The timeline's length affects user operation speed and accuracy, while the control panel's vertical location influences users' exploration of video content. Consequently, determining the optimal timeline length and control panel vertical location is important for delivering an exceptional user experience.

We therefore present a study investigating the impact of timeline length and control panel vertical location on the accuracy, task completion time, and comfort of user interactions during video playback. Additionally, we will observe and analyze user behavior while interacting with the 360-degree video player. Our research is particularly relevant for the design of 360-degree video players that require a lot of interaction (such as the usages we mentioned earlier for educational purposes), and can also provide practical guidance for general 360-degree video player design. By contributing insights into the relationship between timeline length, control panel vertical location, and user experience, our findings expand the existing body of knowledge. Furthermore, through observing user behavior, we gain valuable insights into interaction patterns and preferences, which will inform future design decisions.

2 Related Work

2.1 Design Guidelines for Virtual Reality and 3D Applications

In recent years, with the development of VR

One key difference between VR (or 3D) environments and ordinary 2D settings is the incorporation of depth cues, such as motion parallax, relative scale, and lighting, to enhance immersion when designing user interfaces [1, 5]. But at the same time, the visual enhancement experience can be equally uncomfortable for the user's eyes. To counteract visual fatigue caused by the close proximity of the head-mounted display (HMD) to the human eyeball, studies suggest placing attention-demanding objects, like menus, within 0.5-1 meter from users, striking a balance between visibility and fatigue [1, 5]. Additionally, the horizontal positioning of these objects is also important. In Kolhe et al. 's research, they divide HMDs' field of regard (FOR) into 5 zones (see Figure 1). Among these zones, the *comfort zone* allows users to comfortably explore content, while the *peripheral zone* is unsuitable for extended content viewing. Therefore, they suggest put important objects in the *comfort zone*, and prevent to put objects in the *peripheral zone* to avoid potential discomfort [1, 10]. To optimize the arrangement of content like attention-demanding objects, guidelines recommend surrounding the user with the content while avoiding fixed graphic user interface (GUI) elements that obstruct the view [1, 5, 11]. Besides, appropriate spacing between objects is also essential to prevent accidental touches [1].

Regarding feedback in VR environments, guidelines propose two types: the feedback provided by the system to the user and the feedback from the user's body in the virtual environment. For the first one, it means the system must offer

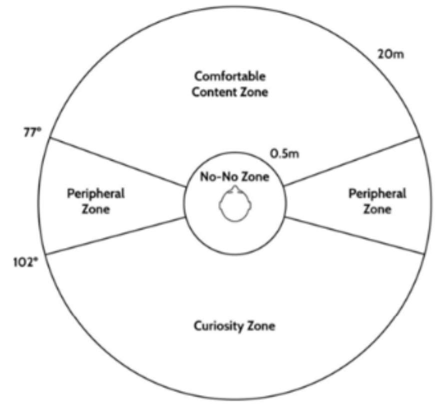


Figure 1. Five content zones [10].

feedback to ensure users perform the expected operations when interacting with the UI [1, 4, 12], for example, when the pointer hovers or clicks on a button, the color of the button changes accordingly. For the second one, since users cannot see their own body parts in the VR environment, providing feedback like a virtual hand (or controller) can help users orient themselves in the virtual world. Besides, in order to ensure it doesn't obstruct other objects, some guidelines suggest that make the virtual hand translucent when it is not in use [1, 5].

In addition to taking into account the recommendations mentioned in the design guidelines, considering ergonomics in VR environments is crucial to enhancing the user experience, especially when using VR HMD. For instance, according to Figure 2, normal human eyes have the binocular vision area covering from L62° to R62°, but when using an Oculus Quest HMD, the horizontal field of view (FOV) is limited to 90° [5, 7]. So, the FOV in HMD is smaller than that in human eyes. In order for the user to have quick access to interactive components, most components should be placed in a range that should be less than or equal to the horizontal FOV of the HMD, which is 90°. The vertical FOV is also divided into several parts. In which the maximum angle of upward rotation of the human eyeball without head rotation is 25° and the maximum angle of downward rotation is 30°, which means, it is easier for human eyes to rotate downward than upward [5]. It is also worth noting that the human eyeball naturally tilts downward by about 15° in a relaxed state, not horizontally as we usually think of it [5].

Overall, the design guidelines and ergonomic guidelines above mention some interesting points that are relevant in the context of 360-degree video as well. First, our player design needs to take into account the differences between 2D and 3D environments – a design that is suitable for 2D environments is not necessarily suitable for 3D environments. Secondly, since the 3D environment has one more dimension than the 2D environment, we also need to consider how to

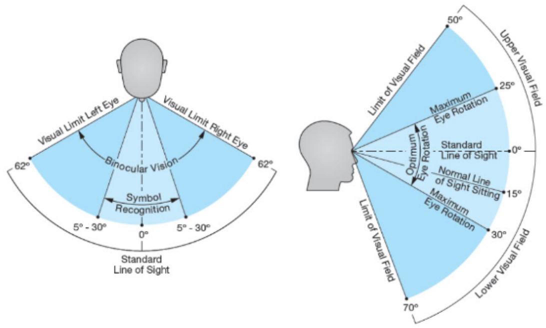


Figure 2. Horizontal and vertical FOV from Meta Quest's developer manual [5].

use this extra spatial dimension to give the user a better experience, rather than making it an extra burden. Finally, since our research focuses on 360-degree video players in HMDs, we also have to consider ergonomic recommendations. Because HMD devices, in general, tend to put an extra burden on the user's head and neck, the proper design must minimize these burdens as much as possible.

2.2 Studies on 360-degree Video Players

Immersive 360° videos and VR offer exciting opportunities for engaging user experiences. However, they also present challenges. In this section, we will focus on the challenges of cybersickness and effective subtitle implementation, which appear to be the most relevant challenges in the context of our work on 360-degree video interaction. Several recent studies have explored these issues to enhance user experiences in this emerging technology landscape.

Cybersickness, a form of motion sickness experienced during virtual experiences, has been a major concern in immersive 360° videos and VR. A study by Groth et al. investigated the use of visual techniques, including blur and opaque occlusion, to modify the ground truth condition and reduce cybersickness. Results showed both techniques proved effective in reducing cybersickness, with opaque occlusion slightly outperforming blur. Additionally, the unmodified ground truth condition was found to be the most sickness-inducing. Furthermore, the study identified that gender plays a significant role in susceptibility to cybersickness. Overall, visual techniques show promise in mitigating cybersickness and improving user comfort during immersive experiences [8].

Subtitles play a crucial role in providing information to users, especially in VR and 360-degree video experiences where the viewers' gaze can roam freely [2, 3, 14]. Two studies specifically focused on subtitle placement and design to enhance user experience in this context.

In the first study, researchers compared dynamic and static subtitle methods in cinematic VR. Dynamic subtitles, which change positions as the conversation progresses, were found to be more effective in reducing simulator sickness and task workload. They offered a more natural and immersive experience. However, the study also identified challenges in handling moving protagonists who are speaking and speaker identification for hearing-impaired individuals, calling for further investigation in these areas. [14]

In the second study, researchers explored subtitle behavior for 360° video and VR through user testing. They designed and implemented four subtitle behaviors and found that placing subtitles near the speaker proved to be the most favorable. This approach minimizes disruption to viewers' natural gaze patterns, allowing them to fully explore the scene while still accessing essential information. The study also identified several discriminatory themes, including effort, missing out, obstruction, distraction, and immersion, which provide valuable insights for improving subtitle implementation and user experience in this domain [2, 3].

However, we found that the existing research is still in the exploratory stage and lacks findings that can optimize the design of 360-degree video players. Many of the current designs simply transplant player interfaces from a 2D environment to a 3D environment, overlooking the need for adjustments to align with the unique characteristics of the 3D setting [1, 9, 15]. In light of this, we draw parallels with studies on subtitle design in 360-degree video. Subtitles, like the control panel of a player, serve a similar function: they offer information independently from the video content and can be accessed promptly when needed. To maintain immersion, their presence is minimized when not in use. The primary distinction lies in the active interactivity required for the control panel, compared to subtitles, which are passively received. Consequently, there is a pressing need for a better user experience in terms of interaction.

In combination, these studies offer valuable insights into the continuous efforts aimed at improving user experiences in immersive 360° videos and VR. By addressing cybersickness and optimizing subtitles, researchers and content creators can craft more pleasant, captivating, and inclusive experiences within this exciting realm of technology.

2.3 Virtual Reality in Educational Use Cases

To date, 360-degree video and VR have been used in many industries, and education is no exception. Among these, medical and healthcare education has emerged as two of the prominent areas of use. This section presents a summary of studies that analyze the applications and potential benefits of these immersive technologies in education. The findings of these studies underscore the potential of 360-degree video and VR to enhance learning experiences and offer valuable guidance for future research and design.

Two literature reviews which are conducted by Pirker et al. and Shadiev et al., analyzed existing research articles. The results of both reviews indicate a growing interest in these technologies and their potential for educational use.

They highlighted some benefits reported in various studies, including increased student engagement, improved learning outcomes, and heightened motivation. Moreover, these technologies enable students to explore environments that are difficult to access in real life. However, the studies also emphasize the need for further research as the technology is still largely exploratory and lacks standardized tools, processes, and metrics. Additionally, they note that most of the current applications have been in higher education, suggesting the necessity for exploring their potential in elementary education like K-12 education [13, 16]. Snelson et al.'s review also yielded similar results, with the added insight that while immersive technologies provide a sense of immersion, they can also distract students. Hence, the specific content of the videos should be carefully considered in educational settings [17].

In contrast, Violante et al.'s study focused on how 360-degree video can serve as an interactive virtual technology for immersive learning experiences, primarily in engineering education. Their research introduced a methodology based on educational principles and student engagement constructs for designing effective 360-degree interactive videos. The results demonstrated the effectiveness of this approach, with students expressing high levels of engagement and a deeper understanding of the subject matter through the use of 360-degree videos. The authors also emphasized that these videos can enhance student engagement and deliver memorable learning experiences, which can be applied across different educational domains [18].

In conclusion, these studies collectively highlight the potential of 360-degree video and VR technologies to enhance the educational learning experience. They provide valuable insights into the benefits, challenges, and diverse applications of these immersive technologies in various educational environments. However, it is crucial to acknowledge that these technologies are still in the exploratory phase, necessitating further research in various areas to create engaging and effective learning experiences for learners of all ages. As we mentioned in the introduction, we believe that in order for the 360-degree video to reach its full potential in learning and education, users must not be deterred by the player itself, and it must provide a comfortable, convenient, and intuitive experience that allows users to easily explore the content and fully focus on it.

3 Methodology

3.1 Research Questions

The control panel design of a 360-degree video player is critical to the user experience. The control panel (see Figure 3), whether in a 2D or 3D player, usually contains play/pause buttons, fast-forward and backward buttons, a timeline slider that can be clicked and dragged, and a timeline. The main focus of this study is on optimal timeline length and control

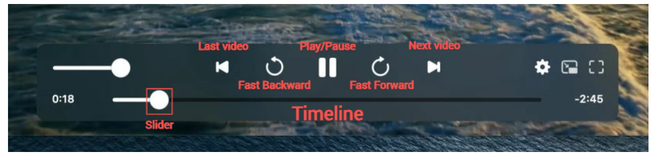


Figure 3. An example of player's control panel.

panel placement, as this is particularly important in some contexts that require a lot of interaction, such as educational contexts. The main purpose of our study is to assess the impact of these two parameters on user experience. In addition, we will observe and analyze user interaction behavior with the timeline to further understand how these interaction components are best implemented. Therefore, we summarized these goals into the following two research questions:

- **RQ1:** How do the timeline length and the control panel vertical location impact the user experience in a 360-degree video player?
- **RQ2:** What are the discernible user behaviors exhibited while interacting with a 360-degree video player?

By addressing these questions, we aim to provide valuable insights into the relationship between the timeline length, control panel vertical location, and user experience. Furthermore, our observation of user behaviors will contribute to understanding interaction patterns and preferences, serving as a valuable reference for future design decisions.

3.2 Specific Use Case

In this study, we choose the educational setting as our specific use case to examine the influence of timeline length and control panel vertical location on user experience and behavior. As previously mentioned in Section 2.3, several studies have highlighted the potential of using 360-degree videos in the educational context, particularly in fields like medical and healthcare education. This educational approach offers students a unique and immersive learning experience, enabling them to explore environments that are typically inaccessible or challenging to access physically [13, 17].

Furthermore, there are of course many other use cases where users tend to passively watch 360-degree videos without much interaction. However, in the educational context, students actively engage with the videos, repeatedly watching specific clips, exploring the video content, and quickly navigating through the material based on their own needs or the instructions of their teachers. These interactive behaviors align more closely with the focus of our study.

Therefore, the participants in our study are students aged 18 and above who have already pursued or are currently pursuing higher education. They are not only the right target users for our use case, but this age group also demonstrates a higher level of receptiveness to new technologies and possesses prior knowledge of VR, and is thus also naturally open to exploring and adapting to new experiences. To ensure better control of variables and collect precise experimental data, our study will simulate a classroom scenario where a teacher instructs students to explore the video content. The assigned tasks for the participants in the experiment will simulate instructions given by the teacher, thereby ensuring that all participants have similar interactions with a shared purpose throughout the experiment.

3.3 Experimental Design

3.3.1 Materials. For this study, we established an experimental system utilizing an existing framework for 360-degree video players. The framework was developed using Unity and C# specifically for HMDs. During the experiment, the system was executed on the experimenter's laptop, and the screen was synchronized with the Oculus Quest headset using Quest Link.

The system consists of a comprehensive player, a tutorial, task instructions, and an automated log function to record user interactions. In order to keep the experimental environment closer to real life, the player used in the experiment was completely restored to the existing player design. The control panel contains play/pause buttons, fast-forward and rewind buttons, and an interactive timeline (see Figure 3). The two ends of the control panel display two times, the left is the current time of the video, and the right is the total duration of the video. The time on the left is updated in real-time as it plays, and when the participant drags the slider of the timeline, it is updated in real-time based on the location of the slider. Besides, we did not put a time scale on the timeline. Although this design would help participants to be clear about the time point on the timeline, this design is not very commonly used by most of the players. During the experiment, we disabled the fast-forward and fast-backward buttons of the player to encourage participants to concentrate their interactions primarily on the timeline. Interaction with the system is achieved through the HMD's head input and controller inputs. Figures 4 and 5 displayed below provide examples of the interface shown in this system.

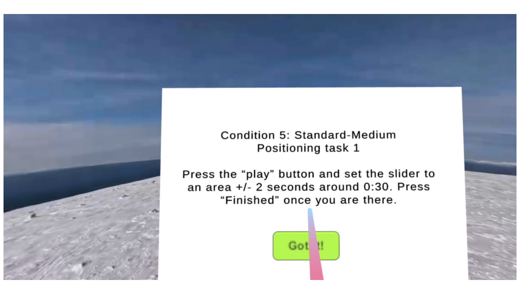


Figure 4. An example of task instruction in the system.

To minimize the impact of learning effects and ensure that participants in each experimental condition were not influenced by prior exposure to the video content, we chose

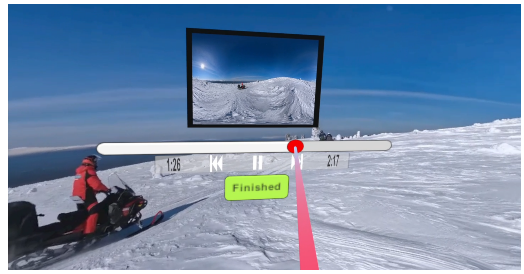


Figure 5. An example of the player's control panel in the system with the slider being dragged by the user.

10 distinct 360-degree videos for the experiment. Among these, 9 videos were allocated for the experiment itself, while one video was reserved for the tutorial session. The chosen videos were obtained from reliable sources on the Internet and were recorded using fixed cameras, with content on culture, tourism, or science.

To ensure consistent time granularity across the same timeline lengths, the duration of each video was standardized to approximately 2 minutes. In the experiment, all the videos are played in the same order and the task content of each video is fixed. For detailed information regarding the specific videos utilized in the experiments and the corresponding task instructions associated with each video, please check Appendix B.1.3.

3.3.2 Experimental Conditions. In this study, our primary focus is to investigate the effects of the timeline length (referred to as **length**) and the vertical location of the control panel (referred to as **vertical location**) on user experience and behavior. For each variable, we selected three parameters to explore their impact.

In relation to the length, we choose three distinct parameters: **Short, Medium**, and **Long** (see Figure 6), based on the FOV of Oculus Quest and the FOV of the human eye. The **Long** parameter encompasses a FOV ranging from L45° to R45°, aligning with the FOV of the Oculus Quest [7]. The **Medium** parameter spans from L30° to R30°, representing the range of human vision's symbol recognition zone [5]. To maintain a consistent 15-degree gap between the left and right extents, we specifically designated the **Short** parameter as L15° to R15°.

For the vertical location, we considered the **Lower, Standard**, and **Higher** locations (see Figure 7), taking into account the vertical FOV of the human eye. It is worth noting that previous studies have indicated that looking upwards imposes a greater burden on the human neck and eyes compared to looking downwards, requiring more effort to complete the upward movement [1, 5]. Therefore, we excluded the range above horizontal in our design. The **Higher** location corresponds to 0°, representing the horizontal view angle [5]. The **Standard** location aligns with a 15° downward view angle, which is the natural-relaxed state of the human eye and is mentioned in many VR design guidelines and also widely used in existing VR interaction designs [3, 5, 6]. Lastly, the **Lower** location represents the maximum angle at which the human eye can be turned downward while the head is fixed [5].

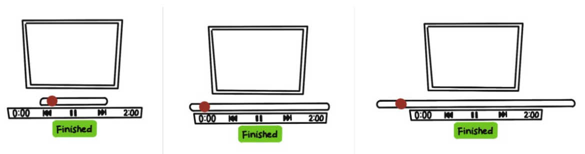


Figure 6. How **Short, Medium** and **Long** length parameters are represented in the experiment.

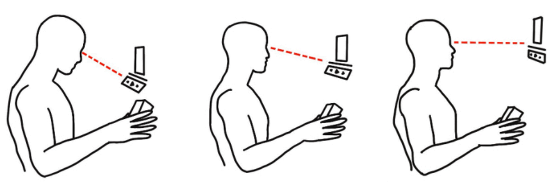


Figure 7. How **Lower, Standard** and **Higher** vertical location parameters are represented in the experiment.

Furthermore, we are interested in exploring the interrelationship between these two variables. We believe that different combinations of parameters may yield varying results. Hence, we combined each set of three parameters for both the length and vertical location, resulting in a total of nine experimental conditions (see Table 1).

Table 1. 9 experimental conditions

Length \ Vertical Location	Lower	Standard	Higher
Short	L-S	S-S	H-S
Medium	L-M	S-M	H-M
Long	L-L	S-L	H-L

Due to limitations in experimental time and the number of participants, we designed a within-subject experiment, where all participants experienced the same nine conditions. To mitigate potential learning effects, we employed a balanced Latin square design to determine the order in which participants performed the conditions. The specific order of experiments for each participant can be found in Appendix B.1.4.

3.3.3 Experimental Tasks.

The experiment comprises two types of tasks: the **positioning task** (Tpos) and the **search task** (Tsr). These tasks were selected to simulate instructions commonly given by teachers in class. The positioning task simulates the scenario where the teacher instructs students to navigate to a specific time point in the video, while the search task simulates an instruction to find a particular scene.

For the positioning task, participants were deemed correct if their response fell within a range of -/+2 seconds of the given time point. In the case of the search task, as the target clip appeared in all videos for a duration of 10-20 seconds and was presented from an easily accessible perspective, participants were considered correct if their response fell within the duration of the clip.

To explore the distinct effects of different time navigation methods (i.e., grab slider sliding or click a specific time on the timeline) and directions (left to right or right to left), two variations of the positioning task were implemented for the experiment. Prior to each task, the task instruction was displayed directly in front of the user within the system. The positioning task instruction provided a specific time point, while the search task instruction included a brief textual description of the target clip and a hint indicating whether the clip appeared in the first or second half of the video. Here are the examples of three tasks in the tutorial:

- Tpos 1: *Press the "play" button and set the slider to an area +/- 2 seconds around 1:45. Press "Finished" once you are there.*
- Tpos 2: *Press the "play" button and set the slider to an area +/- 2 seconds around 1:20. Press "Finished" once you are there.*
- Tsr: *Please find this scene in the first half of the video: two cats sitting in front of a fish tank, one of the cats is staring at the fish tank, and another gray cat is looking at you.*

3.4 Experiment Procedure

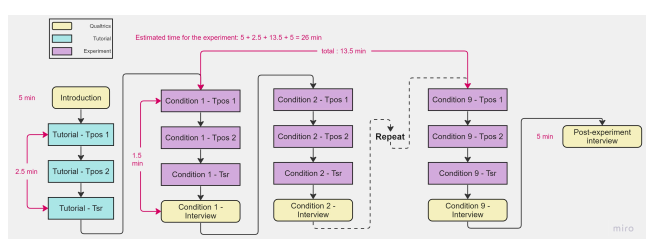


Figure 8. The chart of experiment procedure.

The overall experimental procedure is depicted in Figure 8, illustrating the sequential steps followed in the experiment. As this study adopts a within-subject design, all participants undergo the same experimental process.

Before the experiment starts, a brief introduction is provided to the participants. This introduction includes presenting an information form about the study, obtaining the participants' consent by signing a consent form, collecting participants' general information, and providing an overview of the experiment. Once it is confirmed that the participant agrees to partake in the experiment and has no questions regarding the procedures or content, the experimenter assists the participant in putting on the HMD and starts the experiment.

The participants first go through a tutorial session that follows the same arrangement of each condition in the formal experiment. The purpose of the tutorial is to introduce each task and help participants in familiarizing themselves with the experiment procedures and the system. After completing the tutorial, the main part of the experiment starts. The main part encompasses 9 distinct conditions, each condition comprising two positioning tasks and one search task. At the end of each condition, that is, after the participant completes the three tasks in this condition, they are asked to rate the difficulty level of the current condition's setting.

Each of the 9 conditions adheres to the same organization, and upon completion of all nine conditions, participants are instructed to remove their HMDs and participate in a brief post-experiment interview to gather additional insights.

3.5 Data Gathering

The data collected in the experiment encompassed two main types: quantitative and qualitative data. The quantitative data included measures such as correctness, accuracy, time spent, number of grabs/clicks, and duration of grabs/clicks. To provide more detailed insight, we gathered the following information:

- Correctness: This refers to whether participants completed the tasks correctly or not. Correct answers were labeled as 0, while incorrect answers were labeled as 1 in the interaction log.
- Accuracy: This indicates the deviation of participants' answers from the correct answers. Similar to correctness, correct responses were labeled as 0. For incorrect responses, the value corresponded to the number of seconds by which the participants' answers deviated from the correct answers.
- Time spent: This refers to the time participants spent on each task. In our experiments, this data was obtained through a timer integrated into the experimental system. The timer started when participants pressed the "play" button, and it stopped when participants completed the task and pressed the "finished" button. Then, the recorded time was then logged in the interaction log.
- Number of grabs/clicks: This indicates the count of interactions, especially grabs or clicks, made by participants on the timeline during each task.
- Duration of grabs/clicks: This refers to the time interval participants jumped on the timeline for each grab or click action.

All of the above data were automatically logged by the experimental system and saved in a text file, organized according to each participant's assigned number.

In addition to quantitative data, qualitative data were collected through questionnaires (see Appendix B.1.6 for questionnaire details). The questionnaires covered aspects such as the difficulty level of the tasks, cybersickness and discomfort levels, participants' preferences for timeline length and vertical location, and other feedback. The difficulty level was rated on a scale from 0 to 10, where 0 represented "super easy" and 10 indicated "super hard." Cybersickness was assessed using three yes-no questions about headache, dizziness, and nausea, while discomfort was evaluated on a scale from 0 (not at all likely) to 10 (extremely likely). Participants' preferences for timeline length and vertical position were gathered through two multiple-choice questions. In the post-interview section at the end of the experiment, participants provided additional feedback, which was categorized as "other feedback."

4 Results

4.1 Quantitative Data

In this experiment, 18 participants were recruited, each engaging in 9 conditions consisting of three tasks: two positioning tasks and one search task. Consequently, we collected a dataset comprising 486 sets of interaction logs. In the following section, we analyzed these 486 logs, categorizing them into various groups and obtaining quantitative results.

4.1.1 Task Correctness. Task correctness, represented by a Boolean value (0 for "correct" and 1 for "incorrect"), and its relationship with variables such as timeline length, control panel's vertical location, and conditions. The analysis involved a substantial sample size of 162 logs, and the chi-square test was used due to the non-normal distribution of the binary data. The results indicated that although the Long timeline length group showed potentially higher operational correctness (correctness Long = 93.2%)., there was no significant difference in correctness between the Short, Medium, and Long groups (p=0.41). Similarly, while the Lower vertical location group exhibited higher correctness (correctness Lower = 92.6%), there was no significant difference between the Lower group and the other two groups (p=0.52). Additionally, when considering the combination of independent variables in different conditions, no substantial variation in correctness was observed. Overall, the statistical analysis did not reveal any significant correlation between task correctness and the studied variables.

4.1.2 Task Accuracy. In regard to the accuracy of each task, we used the offset value in the interaction logs to assess it, which measures the time deviation between participants' responses and the correct time point. A larger offset value indicates greater deviation, while 0 represents a correct response. The normality assumption of the offset values was tested using the Shapiro-Wilk test, which indicated violations of normality with $p<0.05$. Nevertheless, due to the lack of a more suitable alternative and the robustness of ANOVA in the presence of normality violations, a two-way repeated measures ANOVA was employed for data analysis.

When testing for the normality assumption, we found some extreme values in the results. These extreme values may be due to some participants misremembering the target time in the task. In order to prevent these values influence the analysis of the results, the Three-Sigma Rule was applied, leading to the removal of two records from the task accuracy analysis.

Table 2. The results (p-value) of ANOVA on task accuracy. Here, $p_{vertical-location:length}$ represents the ANOVA result of the interaction between 2 independent variables.

Task	$p_{vertical-location}$	p_{length}	$p_{vertical-location:length}$
Tpos1	0.69	0.12	0.06
Tpos2	0.30	0.95	0.27
Tsr	0.52	0.62	0.40

Table 2 above shows the results of ANOVA. The analysis showed no significant differences in offset values among the three tasks (Tpos1, Tpos2, and Tsr) based on vertical location or length. Additionally, the interaction between these independent variables did not yield any significant differences across all three tasks.

Examining the mean offset values revealed that the Tpos1 task achieved 100% accuracy in five conditions: Higher-Length, Lower-Medium, Lower-Short, Standard-Length, and Standard-Medium. Conversely, the Higher-Medium and Standard-Short conditions exhibited relatively lower accuracy. For the Tpos2 task, participants achieved 100% accuracy in the Lower-Long and Standard-Medium conditions, while the Lower-Short condition showed the lowest accuracy. In the Tsr task, only the Higher-Medium condition achieved 100% accuracy, while the Standard-Short condition displayed the lowest accuracy.

Overall, the results of the analysis showed no significant correlation between the accuracy of the task and either the length of the timeline, the vertical location of the control panel, or the interaction between the two.

4.1.3 Task Time Spent. We explore the relationship between the time spent on different tasks and their vertical location and length. The time spent data was extracted from the interaction logs, representing the duration between pressing the "Play" button to start video playback and pressing the "Finish" button to complete the task.

The Shapiro-Wilk test indicated a violation of the normality assumption for the Tpos1, Tpos2, and Tsr groups. Despite this, a two-way repeated measures ANOVA was employed, considering the lack of better alternatives but ANOVA still has some robustness in handling normality violations. Notably, extreme values were not excluded from the data set due to the absence of task completion time restrictions.

Task	$p_{vertical-location}$	p_{length}	$p_{vertical-location:length}$
Tpos1	0.24	0.23	0.54
Tpos2	0.43	0.07	0.27
Tsr	0.96	0.32	0.95

Table 3. The results (p-value) of ANOVA on task time spent. Here, $p_{vertical-location:length}$ represents the ANOVA result of the interaction between these independent variables.

The analysis revealed no significant differences in time spent between each task and vertical location, length, or interaction between these two. Table 3 shows the results of the ANOVA of time spent.

Examining the mean values, the Tpos1 group showed that Lower-Long had the highest average time spent, while Higher-Medium had the lowest. In the Tpos2 group, time spent did not significantly vary across conditions, with Higher-Short having the highest average time spent and Higher-Medium having the lowest. For the Tsr group, the overall average time spent was significantly higher compared to the other two groups, with Higher-Medium having the highest average time spent and Standard-Short having the lowest.

In conclusion, the Higher-Medium setting generally allows for faster completion of the positioning task but requires more time for the search task.

4.1.4 Grab/Click Action. This section focuses on analyzing participants' interaction behavior with the timeline during the experiment, specifically regarding their grab and click actions on the slider. The study examines the number of grabs and clicks recorded in the interaction logs, as well as the time duration of the slider movement with each grab or click. The objective is to determine if the participants' choice of grab or click is related to different conditions or duration.

Regarding the number of grabs and clicks, according to Figure 9, the number of grabs exceeded clicks across all conditions. The number of grabs decreased as the timeline length increased, indicating greater accuracy in grab actions with longer timelines. However, there was no substantial correlation between the number of grabs and the vertical location. Besides, the number of clicks was consistently lower than the number of grabs, with the highest number of clicks

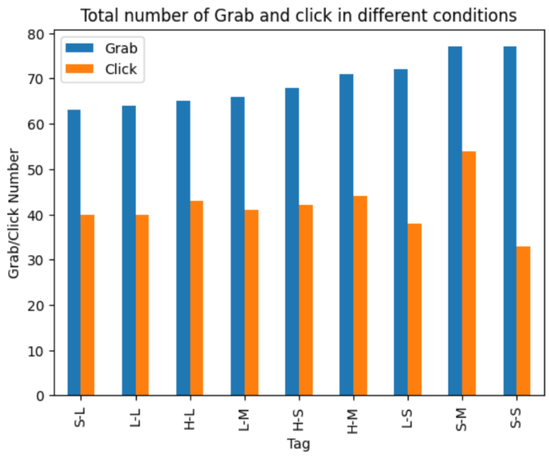


Figure 9. Total number of Grab and Click in different conditions.

in the Standard-Medium condition and the lowest in the Standard-Short condition.

When investigating the relationship between the number of grabs/clicks and conditions in different tasks, we found that participants tended to use more grabs in positioning tasks (Tpos1 and Tpos2), while in the search task (Tsr), they would use click more in some conditions. Figure 10 showed a higher number of grabs and clicks in the Tsr group compared to the Tpos1 and Tpos2 groups, which could be attributed to the presence of a clear target time in Tpos1 and Tpos2 compared to the exploratory of Tsr. In addition, the player used in our experiment will show the time of the current slider when the participant grabs the slider (even if you don't release it), and the click will only know the time of jumping when you release the mouse. Thus, we can say that the grab action gives the participant "predictable" feedback in advance, whereas the click does not. In order to accurately jump to the target time given by Tpos, participants are more inclined to choose grab, which can give them feedback in advance. For Tsr, where the targets are larger ranges and the correct answer is more focused on the video content, participants will use more clicks.

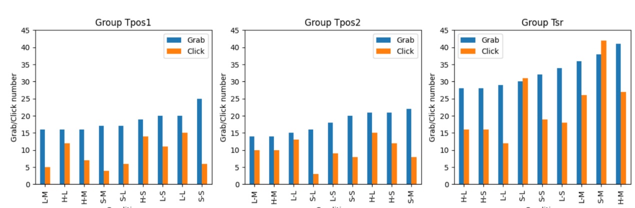


Figure 10. Total number of Grab and Click in different conditions grouped by different tasks.

Additionally, we also explored the influence of slider movement distance, which we called "duration", on participants' grab and click actions. Regarding grab action, we found that the average grab duration ranged from 30s to 40s across different conditions, and there was no substantial difference in duration based on vertical location (see Figure 11). Surprisingly, we found the Standard setting had the lowest mean value, which is contrary to our initial assumptions.

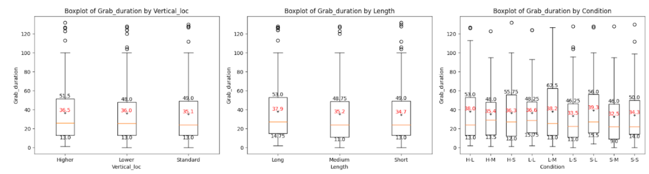


Figure 11. Boxplots for grab duration grouped by vertical location, length, and conditions.

Furthermore, as the timeline length increased, the duration also increased, we think it likely due to the decrease in time granularity on longer timelines. When participants grab the same distance in the timeline (e.g., 1cm), a shorter timeline will have fewer details, which will lead to content loss. So, we analyzed the average grab number over different lengths (see Figure 12) and the correctness at different lengths (see Figure 13). From the result, we can see the participants tended to perform more actions on shorter timelines to achieve higher accuracy, whereas longer timelines allowed them to finish the task in "one-step" actions, which explains the finding above.

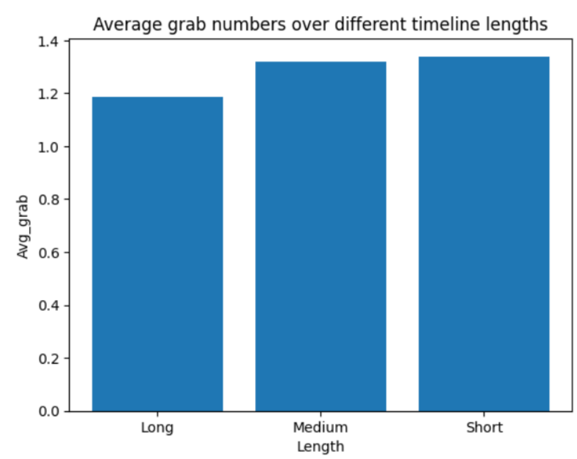


Figure 12. Average grab numbers over different timeline lengths.

Regarding click action, based on the average click duration in different conditions, we observed that the Standard-Long condition had the shortest average click duration, while the Higher-Medium condition had the longest (see Figure 14). Surprisingly, the Standard vertical location exhibited the lowest average duration, contrary to the initial hypotheses.

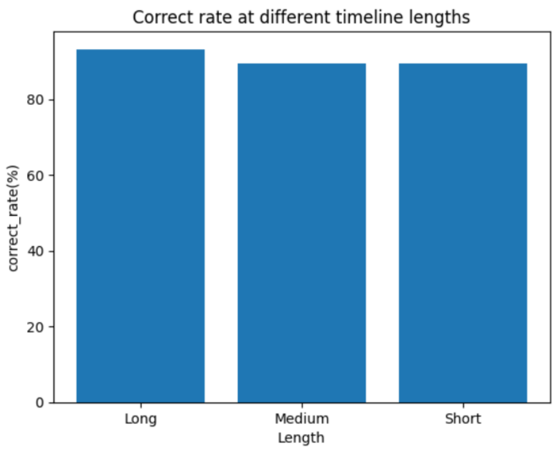


Figure 13. Correctness at different timeline lengths.

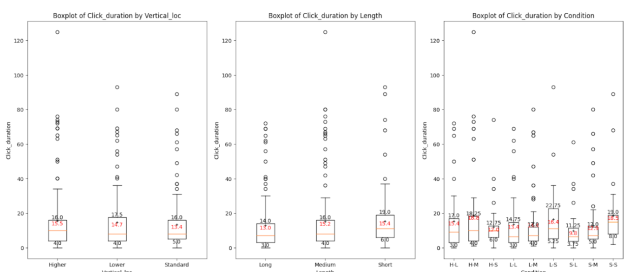


Figure 14. Boxplots for click duration grouped by vertical location, length, and conditions.

Additionally, shorter timelines were associated with longer click duration, which was attributed to the increased time granularity on shorter timelines. When they click the same distance (e.g., 1cm), a shorter timeline implies a longer duration.

Comparing grab and click actions based on different vertical locations, it was found that grab actions generally had a longer average duration than click actions. Participants preferred to click actions for a duration that less than 15 seconds and leaned towards grab actions for a duration between 15 and 50 seconds (see Figure 15).

Despite the initial prediction that the Standard vertical location would result in a longer duration due to its assumed comfort, the analysis revealed a different pattern (see Figure 16). While grab actions were more frequent on Standard, click actions did not follow the same trend. Moreover, the accuracy of click actions on Standard was lower compared to other vertical positions, which was not fully explained by the quantitative data. Further qualitative data analysis may be required to gain deeper insights.

4.2 Qualitative Data

4.2.1 Participants Information. In our experiment, we involved a total of 18 participants, with an equal split of

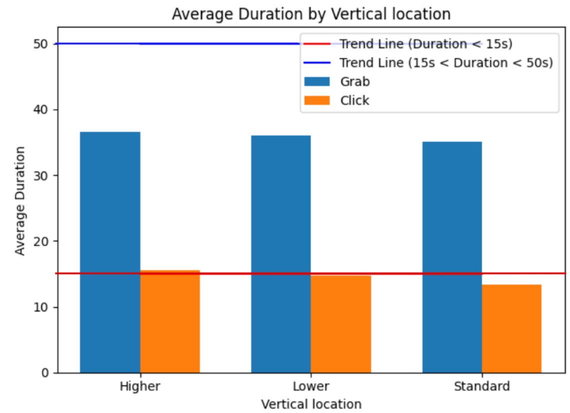


Figure 15. Average duration by different vertical locations (with trend line duration < 15s and 15s < duration < 50s).

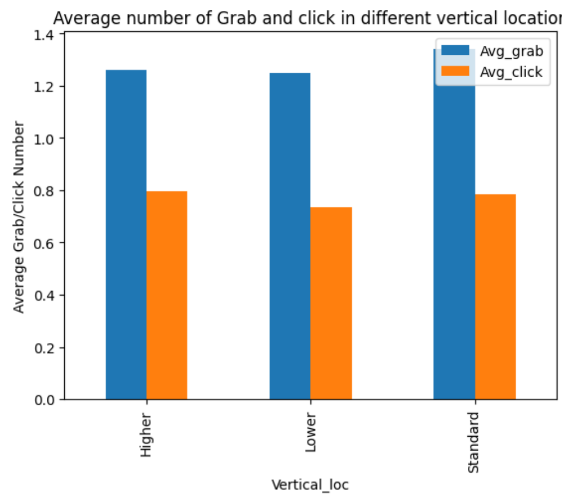


Figure 16. Average number of Grab and Click in different vertical locations.

9 male and 9 female participants, each accounting for 50% of the total. In terms of age distribution, 12 participants were between 18 and 24 years old, making up 66.67% of the sample, while 6 participants were aged between 25 and 34 years, representing 33.33% of the participants.

Regarding VR experience, the majority of participants (11 individuals) reported using VR a few times (61.1%), followed by 3 participants who indicated occasional use (16.7%), and 2 participants who had no prior experience with VR (11.1%). As for 360-degree video experience, 13 participants reported having experienced it (72.2%), with 8 of them using a Head-Mounted Display (HMD) device (61.5%) and 3 trying it on a Mobile device (23.1%). On the other hand, 5 participants stated that they had no experience with 360-degree videos (27.8%).

Overall, the participant group mainly consisted of young individuals with introductory experience in VR and 360-degree videos, which aligns with the intended target user profile for the study.

4.2.2 Cybersickness and Discomfort. This section focuses on assessing the participants' overall experience of cybersickness and discomfort during the experiment. In terms of cybersickness, participants were asked about their experience of headaches, dizziness, and nausea, and the results indicated that most participants reported no significant cybersickness. Additionally, we collected discomfort level ratings from the participants on a scale from 0 to 10, and the majority reported experiencing a low level of discomfort, with only a few indicating a moderate level. Figure 17 displays all the ratings. Overall, the participants demonstrated relatively lower levels of cybersickness and discomfort throughout the experiment.

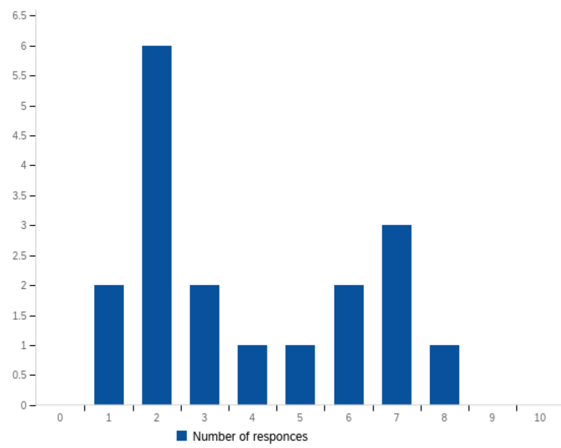


Figure 17. Discomfort level during the experiment.

4.2.3 Difficulty Level of Different Conditions. To assess participants' perceived difficulty in different conditions, we asked them to rate the difficulty level for each condition on a scale of 0 to 10, where 0 means "super easy" and 10 means "super hard". We calculated the mean difficulty level for each condition and found that the Higher-Medium condition was rated as the least difficult, while the Lower-Long condition was rated as the most challenging. Regarding the vertical location, the average difficulty level for the Lower condition was higher than the Higher and Standard conditions. Similarly, regarding the length, the average difficulty level for the Long condition was higher than that of the Short and Medium conditions.

To confirm our findings, we conducted an ANOVA test on the average difficulty levels, which revealed a significant difference (p=0.0008). The post hoc comparison using Tukey's Honestly Significant Difference test showed significant distinctions between the Lower-Long group and the Higher-Medium, Standard-Medium, and Standard-Short groups. In conclusion, participants perceived the Higher-Medium condition as the least difficult and the Lower-Long condition as the most challenging, with a notable difference between the ratings of these two conditions.

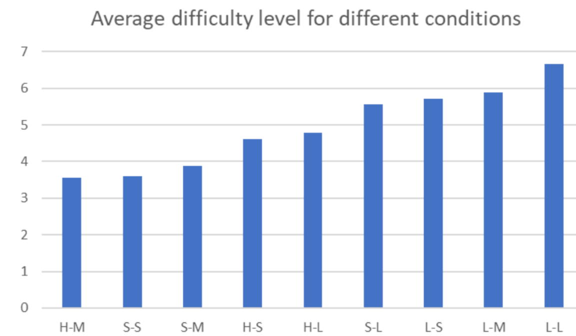


Figure 18. Average difficulty level for different conditions.

4.2.4 Preference of the Length and the Vertical Location. In order to know participants' preferences on length or vertical location, we asked them about their preferred setting via two multiple-choice questions, in which they could choose more than one setting. Additionally, we also collected the discomfort experience due to length or vertical location settings. The following paragraphs show the results.

Regarding timeline length preference, we received a total of 23 responses, the majority of participants (65.22%) expressed a preference for Medium length timelines, followed by 21.74% for Long timelines, and 13.04% for Short timelines. However, 72.22% of participants reported discomfort related to the timeline length. The main issues raised were dissatisfaction with the accuracy of Short timelines, difficulty in locating and grasping the slider with Long timelines due to the control panel's rotation with head movements, and inconvenience in navigating video content and performing search tasks with Long timelines.

Regarding vertical location preference, we received a total of 24 responses, 58.33% of participants preferred the Standard vertical location, while 33.33% favored a Higher location, and only 8.33% preferred a Lower location. However, 61.1% of participants reported discomfort associated with the vertical location. The main reasons mentioned were neck pain from lowering their heads with the Lower location and difficulty in spotting the Lower timeline.

Overall, the study found that the majority of participants preferred Medium length timelines and the Standard vertical location. However, a subset of participants showed a preference for Long timelines or Higher vertical locations. Additionally, the discomfort was primarily associated with extreme settings, which impacted participants' operations.

4.2.5 Overall Feedback.

In addition to the feedback mentioned earlier, we conducted brief interviews with the participants to gather their overall comments, suggestions, and feedback. This section summarizes the insights obtained from these interviews.

Regarding discomfort, participants commonly expressed discomfort associated with the weight of the HMD devices throughout the experiments, which was consistently reported across different experimental conditions to varying degrees. They believed that this discomfort could accumulate over time. To address this, several participants recommended the use of lighter HMD devices to alleviate this discomfort.

Participants also provided interesting feedback regarding their preferences and comfort levels concerning the timeline length and vertical location. Some participants found one particular setting to be "significantly more uncomfortable/dislike", while perceiving little difference between the other two settings. Specifically:

- The Short timeline length caused considerable discomfort, but participants did not perceive a significant difference between the Medium and Long lengths.
- The Long timeline length was deemed excessively extended and unfavorable. However, participants found the Short length to be better than expected, requiring minimal action to complete tasks. The Medium length was highly regarded as a preferred choice.
- Length and height had minimal impact on positioning tasks, but longer timelines were considered worse in search tasks.
- While longer timelines allowed for more precise operations, medium-length timelines provided a better understanding of the overall content.
- The Lower vertical location caused discomfort in the neck, while the Higher location was beneficial for the neck but required lifting the hands to operate. The Medium location was seen as a trade-off between the two.

Consistent with the previous preference survey, participants were able to clearly indicate their most preferred setting among the three options. However, they did not perceive a distinct difference between the other two settings. Their preference for these settings depended on personal usage habits or level of experience.

Participants also offered suggestions to improve the experiment, highlighting additional factors that could impact the results:

- The textual content of the tasks was lengthy and not easily memorized. Participants sometimes needed to confirm details with the experimenter, resulting in increased time consumption.
- Some participants found the player's control panel too close and suggested placing it farther away for easier operation.

5 Discussion

5.1 RQ1: User Experience

This study focuses on two main aspects of user experience: effectiveness, measured by quantitative data, and user emotions, captured through qualitative data. The experimental results indicate no significant relationship between length, vertical location, or their interaction on the accuracy and time consumption of operations. However, participants' feedback reveals that different length and vertical location settings do influence user experiences. Preference-wise, most of the participants favored Medium length and Standard vertical location, and follow-up interviews corroborated this. But Short length and Lower vertical location are more likely to cause discomfort. Long and Higher settings, which ranked second in preference, were generally found to be dependent on different situations. Nevertheless, several participants identified one setting as notably more challenging, with little difference between the other two.

Therefore, while the length and vertical location may not significantly impact effectiveness in user experience, the Medium length and Standard vertical location were perceived as more comfortable from an emotional standpoint.

5.2 RQ2: User Behavior

In this study, we observed users' behavior while using the 360-degree video player, specifically their interaction behavior with the timeline during the experiments. Interaction behavior, in this context, refers to when users choose to grab the slider or click on the timeline while exploring the video. The results indicate that, overall, participants preferred grabbing the slider, especially when they had a clear goal for jumping to a specific time point. Moreover, as the timeline's length increased, the number of grabs decreased, but the distance jumped increased. This behavior can be attributed to the impact of the timeline's time granularity. With a longer timeline, the time granularity became smaller, providing users with clearer details and facilitating more accurate location of their desired points, leading to increased correctness in their operations. Additionally, longer timelines result in shorter click duration, which is also related to time granularity. Longer timeline, smaller corresponding duration for the same 1cm jump on the timeline.

Furthermore, concerning the vertical location, the Standard setting exhibited the lowest duration. However, the current results do not provide a clear explanation for this phenomenon, necessitating further research to gain deeper insights.

5.3 Limitation and Future Works

In this study, we limited our investigation to length and vertical location via 6 fixed parameters in the 360-degree video player. However, some participants expressed the need for more flexibility in setting the length and vertical location, as they believed different tasks may require different configurations. Therefore, future research could explore more personalized settings for length and vertical location and analyze users' preferences for these parameters based on their actual behaviors.

Furthermore, it is important to note that this study specifically focuses on the user experience of a 360-degree video player in the educational use case. Results may differ in other use cases, such as medical or sports use cases. Therefore, future studies could expand the investigation to examine the user experience of 360-degree video players in various other use cases.

Lastly, several participants reported discomfort primarily due to the weight of the device used during the experiment. This discomfort may have influenced the overall results. For future research, employing more comfortable and lightweight devices could help mitigate this impact and provide a more accurate assessment of the user experience.

6 Conclusion

This study aimed to investigate the influence of timeline length and vertical position of the control panel on user experience and behavior. The results demonstrate that while these factors do not significantly impact user experience in terms of effectiveness, they do affect the emotional and behavioral aspects of user interactions. Specifically, we found that a medium-length timeline with a standard-height setup offers the most comfortable user experience. On the contrary, extreme settings, such as long timelines and low heights, can lead to user discomfort. Long timelines may hinder maneuverability, and low heights can cause discomfort to the head and neck, making these settings unfavorable.

Regarding user behavior, we observed that users' choice of interaction actions with the timeline is linked to the clarity of their goal and the duration of the jump. Users preferred grabbing the timeline slider when their goal was clear, whereas they utilized clicking for more precise adjustments to the jump position. Moreover, users tended to opt for grabbing the slider when the time needed to jump was longer.

In conclusion, the study highlights the importance of medium-length and standard-height settings for a comfortable user experience, and it sheds light on the relationship between user behavior and timeline interactions based on clarity of goals and jump duration.

References

[1] Samuel Alves, Arthur Callado, and Paulyne Jucá. 2020. Evaluation of graphical user interfaces guidelines for virtual reality games. In *2020 19th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*. IEEE, 71–79.

[2] Andy Brown, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong, and Maxine Glancy. 2017. Subtitles in 360-degree Video. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. 3–8.

[3] Andy Brown, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong, and Maxine Glancy. 2018. Exploring subtitle behaviour for 360 video. *White Paper WHP* 330 (2018).

[4] Mimmi Cromsjö and Martin Christensson. 2019. Design of video player and user interfaces for branched 360 degree video.

[5] Oculus developer. 2022. Vision | Oculus developer. https://developer.oculus.com/resources/bp-vision/

[6] Google developers. 2018. Design guidelines - Google VR - google developers. https://developers.google.com/vr/develop/experimental/6dof-design-guidelines

[7] Maxim Eylenbosch. 2022. *Defining Differences in Movement and View in Tactical First-Person Shooter Games between Virtual Reality and Personal Computer*. Ph. D. Dissertation. https://doi.org/10.13140/RG.2.2.18412.03205

[8] Colin Groth, Jan-Philipp Tauscher, Nikkel Heesen, Steve Grogorick, Susana Castillo, and Marcus Magnor. 2021. Mitigation of cybersickness in immersive 360 videos. In *2021 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*. IEEE, 169–177.

[9] Chris J Hughes and Mario Montagud. 2021. Accessibility in 360° video players. *Multimedia tools and applications* 80, 20 (2021), 30993–31020.

[10] Prakash R Kolhe, GP Khetri, and NK Deshmukh. 2012. Study of standard assumptions of graphical user interface (GUI) based on usability, adaptability and security factors. *International Journal of Advanced Research in Computer Science and Software Engineering* 2, 12 (2012), 1–7.

[11] Teresa Matos, Rui Nóbrega, Rui Rodrigues, and Marisa Pinheiro. 2018. Dynamic annotations on an interactive web-based 360 video player. In *Proceedings of the 23rd International ACM Conference on 3D Web Technology*. 1–4.

[12] Filip Nilsson. 2017. Upper body ergonomics in virtual reality: An ergonomic assessment of the arms and neck in virtual environments.

[13] Johanna Pirker and Andreas Dengel. 2021. The potential of 360 virtual reality videos and real VR for education—a literature review. *IEEE computer graphics and applications* 41, 4 (2021), 76–89.

[14] Sylvia Rothe, Kim Tran, and Heinrich Hußmann. 2018. Dynamic subtitles in cinematic virtual reality. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video*. 209–214.

[15] Saeed Safikhani, Michael Holly, and Johanna Pirker. 2020. Work-in-Progress—Conceptual Framework for User Interface in Virtual Reality. In *2020 6th International Conference of the Immersive Learning Research Network (iLRN)*. IEEE, 332–335.

[16] Rustam Shadiev, Liuxin Yang, and Yueh Min Huang. 2022. A review of research on 360-degree video and its applications to education. *Journal of Research on Technology in Education* 54, 5 (2022), 784–799.

[17] Chareen Snelson and Yu-Chang Hsu. 2020. Educational 360-degree videos in virtual reality: A scoping review of the emerging research. *TechTrends* 64, 3 (2020), 404–412.

[18] Maria Grazia Violante, Enrico Vezzetti, and Pietro Piazzolla. 2019. Interactive virtual technologies in engineering education: Why not 360° videos? *International Journal on Interactive Design and Manufacturing (IJIDeM)* 13, 2 (2019), 729–742.

A Literature Review

In this section, relevant previous work and their relation to this thesis project are outlined in detail. Section A.1 briefly introduces 360-degree video and its features and some existing 360-degree video players. Section A.2 focuses on interface designs that might be relevant for the work in this thesis, including some existing VR design guidelines, attractive interaction designs in VR games, and captioning behavior designs in 360-degree environments. Section A.3 analyzes the difficulties and challenges of VR UI design pointed out by some existing studies. Section A.4 summarizes the results mentioned in the literature review, the future research directions that can be focused on, and introduce the research question addressed in this thesis.

A.1 Introduction of 360-degree Video

360-degree video, also known as immersive video, panoramic video, or spherical video, is a type of video shot in multiple directions by 360-degree (omnidirectional) cameras or multiple cameras simultaneously. In contrast to traditional video, where the viewpoint is limited to the direction the camera is pointing, in 360-degree video, viewers can watch any viewpoint [7, 36]. The critical feature of the 360-degree video is that the viewer is placed at the center of the video and feels like "experiencing" the content, creating a sense of presence [7, 15].

360-degree video is typically played through a player that, similar to traditional video players, has interactive elements like a play/pause button, forward and rewind buttons, and a timeline with a slider. **Interactive elements** in this case means the UI elements that the user can interact with by head, hand, or controllers. There are two main types of 360-degree video players in the market: 2D platform players (using desktop or mobile devices) and players in 3D environments (using VR HMDs). An example of 2D platform players is *YouTube* [42], where people can now watch 360-degree videos on computers or mobile devices easily. Also have the same interaction function as the traditional video player, which makes user easy to use. But as Gutiérrez-Caneda et al. said, immersive content such as 360-degree videos are mainly conceived for VR HMDs (glasses), and even if they can be viewed on desktop or mobile devices, such an experience does not bring a great sense of presence [15]. Video players in 3D environments such as *Skybox* [31] and *GizmoVR* [13], where users can control the video with the controllers or mouse to manipulate the video and rotate their heads to change the viewpoint. These players are characterized by better adaptation to the 3D environment of 360-degree video. For example, instead of dragging the video to change the current field of view, people can just turn their heads, which is considered more intuitive.

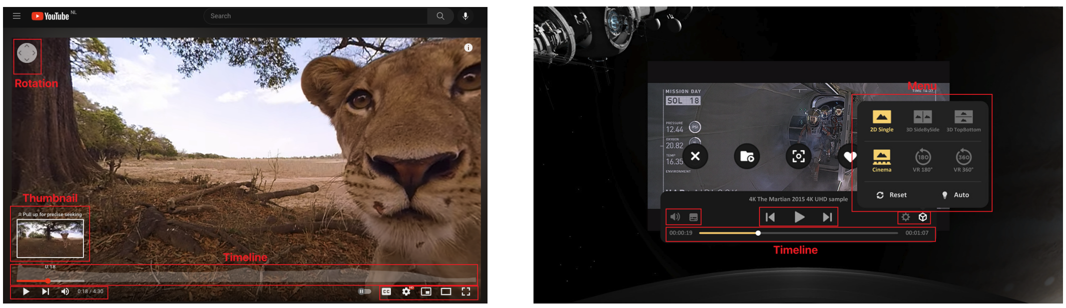


Figure 1: Screenshot of YouTube player on the mobile device [42] (left) and Skybox player in VR environment [31] (right)

A.2 Related Works of Virtual Reality and 360-degree Video

A.2.1 Design Guidelines for Virtual Reality Applications

We generally refer to a graphical user interface (GUI) as a way of interacting with electronic devices using images rather than text commands [18]. Traditional GUIs consist mainly of interactable components like menus, windows, and buttons, using devices such as mice and keyboards for input and monitors for output. However, interaction in a VR environment is different from traditional mobile or desktop devices because it usually works in a 3D environment. We need to design 3D user interfaces(UI) that are more compatible with the 3D environment [21, 2]. In this context, a 3D UI is an interface that allows three-dimensional interactions [21]. Natural 3D UI (3D NUI) is widely used in existing VR applications, mainly because the movement and manipulation of this type of UI are similar to real-world actions, and users do not need much expertise to interact with VR applications [2]. An example of this UI type is the In-hand menu mentioned in A.2.2.

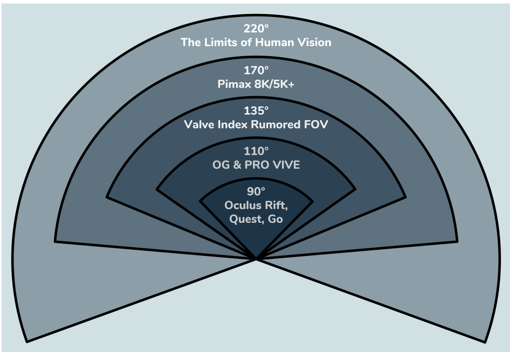


Figure 2: FOV of most VR HMDs on the market compare to human vision [11]

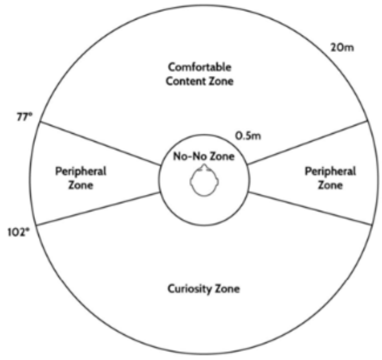


Figure 3: Content zone [18]

Generally speaking, the field of regard (FOR) in an HMD is a spherical or cylindrical screen surrounding the user and displaying images. However, since the HMD market is not highly standardized, different HMD devices will have different fields of view (FOV), generally between 80° to 120° [2] (see Figure 2). [1] and [6] divided the FOR of HMD into five zones, as shown in Figure 3. Among them, the **comfort zone** is the area that users use to view and rotate their heads comfortably and is suitable for placing most of the important contents. The **peripheral zone** is the area where the head is maximally rotatable and unsuitable for long-term contents. The **curiosity zone** is where the user must make some effort to rotate the body and see. The **no-no zone** is less than 0.5 meters away from the user/camera. It is not recommended to place any content in this area, as it is too close to the user's eyes and will cause eye discomfort. The **background zone** is a distance greater than 20 meters from the user/camera. Outside this range, the user cannot use stereo depth perception [1, 6]. A similar horizontal visual range is mentioned in Meta Quest's developer manual, on top of which this manual also provides a vertical visual range, as shown in Figure 4, with a comfortable vertical motion range between -30° to +25° [37].

In the existing market, most VR applications use a HMD as output, which is a wearable device that provides users with an immersive experience by blocking the view of the physical world and rendering computer-generated stereoscopic images [10]. To help developers design a UI that is better suited to the 3D environment, several HMD manufacturers and game engines offer design manuals and guidelines. A study by Alves et al. summarized and evaluated some of the existing manuals and guidelines, and concluded the following recommendations [2]:

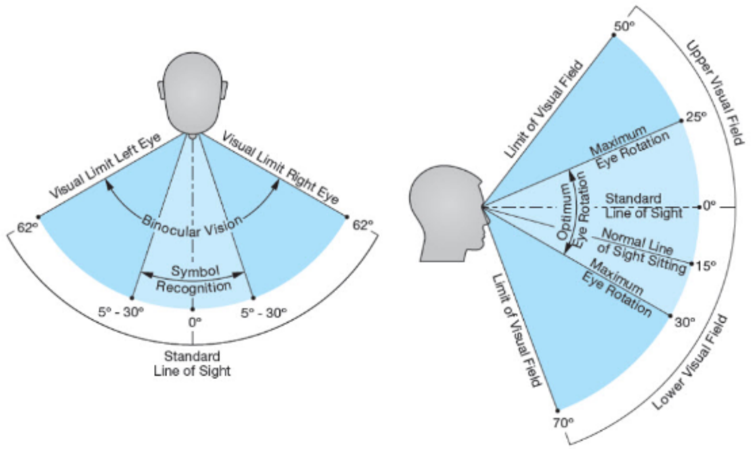


Figure 4: Horizontal and vertical FOV from Meta Quest's developer manual [37]

1. Depth cues such as motion parallax, relative scale, and illumination can be used in the GUI to increase immersion.
2. Place objects (e.g., menus) that require prolonged user attention or attention within 0.5-1 meter of the user. Too far away will result in the user not being able to see the content; too close will cause visual fatigue.
3. Integrating GUI elements into environments or characters can improve immersion.
4. Avoid content in peripheral areas. Encourage users to explore the environment, but too large a range of head movement can cause neck discomfort.
5. Use appropriately sized and easy-to-read text in the UI.
6. Use wearable menus. Wearable menus refer to the integration of menus into the user's virtual hands for a more natural interaction and to avoid occlusion.
7. Use a surrounding GUI to help users have a better reading experience.
8. Provide visual feedback from the GUI to ensure that users take the actions they expect.
9. Use proper scale and spacing of interactive elements to avoid accidental touch.
10. Avoid virtual hands from blocking interactive elements. The hover state of the virtual hand can be set to semi-transparent.
11. Avoid fixing GUI elements in the view. Fixed elements not only block the user's view, but also cause discomfort and nausea when used for a long time.

In a follow-up evaluation, Alves et al. found that points 1, 2, 5 and 9 of their previous recommendations were particularly important, while point 11 was considered a "less important feature", which they suggest may be because most respondents did not experience much discomfort with static menus [2].

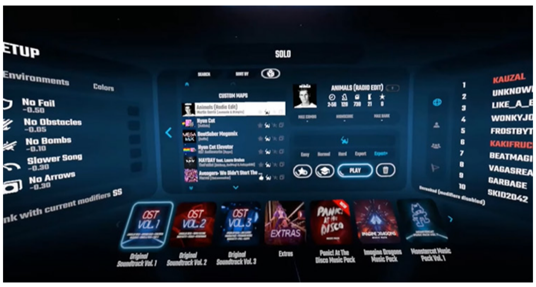
In addition, the size and position of interactive elements are stated in some design guidelines. These guidelines generally recommend appropriate scaling of interactive elements and require a certain distance between elements [2, 8, 9, 37]. Some guidelines also mention the need to limit hand interaction [9]. For example, in Leap Motion's guideline, it is detailed that it is usually recommended the UI elements be interacted with by the index fingertips and that the actual size of the target for one finger be greater than or equal to 20 mm [9]. Such a requirement ensures that the interactive elements are not obscured by the virtual hand or cursor and that they do not accidentally touch the target next to them.

Regarding controller settings, Google's design guidelines mention that the pointer light angle can be lowed by 15 degrees to obtain a more comfortable and ergonomic experience and can reduce the possibility of controller/hand occlusion [8].

A.2.2 Interaction Design of Virtual Reality Games

Although this thesis mainly focuses on the interaction design of 360-degree videos in VR environments. However, games are still the most popular category in the VR field [26]. Many diverse 3D UIs and ergonomics-related studies have emerged over the years, and many comprehensive ideas have been built in the growing consumer market [32]. These studies and ideas can inspire 360-degree video players' UI and interaction design since they are used in the same devices and environments as VR games. Steed et al. introduced several innovative UI designs in their study of 3D UI design in VR games, like the multi-screen non-diegetic menu from *Beat Saber* and the body-centred menu from *Lone Echo* [32].

The **Body-centred menu** is a design that uses the user's body as the basis for placing menus and controls, which is an idea that has been well-received in past research and practice [32]. The user opens the menu primarily by raising the wist and adjusting the menu's size by moving the position of the wrist relative to the head. This design helps users to find the menu more intuitively, and the are similar to the user's movements in real life, which can bring a sense of realism [2]. However, this design will likely cause user exhaustion under prolonged use and needs to be considered by designers. Steed et al. also mention an In-hand menu design, similar to the Body-Centred Menu mentioned above, which typically uses the dominant hand to execute actions and the non-dominant hand to determine a coordinate space [32]. The menu in *Tilt Brush* used this design, with the non-dominant hand having a tool panel and executing actions via the controller or the joystick in the dominant hand.



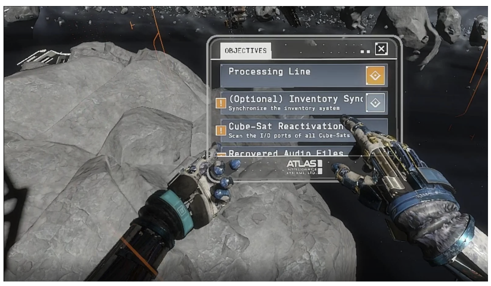


Figure 5: Multi-screen non-diegetic menu from Beat Saber [34] (left) and the body-centred menu from Lone Echo [17] (right)

Both of the designs mentioned earlier are **Diegetic** interfaces. The terminology of game UI proposed by Fagerholt and Lorentzon categorized UI components into four categories according to the degree of connection between narrative and game geometry: Diegetic, Non-Diegetic, Spatial, or Meta [12]. Raffaele et al. evaluated VR first-person games based on Fagerholt and Lorentzon's terminology. The results showed that **Diegetic** interfaces could bring the highest immersion in VR first-person games, followed by **Spatial** interfaces. **Meta** and **Non-Diegetic** interfaces, on the other hand, can break the user's focus and thus lose some important information in the game [26].

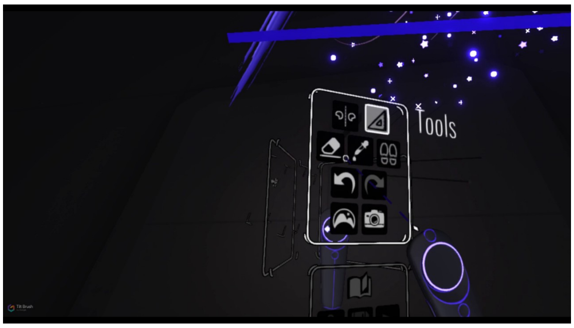


Figure 6: In-hand menu from Tilt Brush [33]

Similar to VR games, especially VR first-person games, 360-degree video as a video genre viewed in first-person, the UI should not be designed in such a way that it breaks the focus of the user's interest and makes them lose some important information. Secondly, although immersion is less important in a 360-degree video interaction than it in VR games, as an advantage of VR, our UI and interaction design should consider this. Therefore, in 360-degree video players, we can adopt Diegetic and Spatial designs such as Body-Centred Menus.

A.2.3 Subtitle Design in 360-degree Video

Similar to the interaction design of a 360-degree video player, subtitle design in the 360-degree video also plays an essential role in an enjoyable user experience. Subtitles are a method of presenting audio content in textual form to the viewer in a stream. Subtitles not only aid in the accessibility design of video players, but many people also use subtitles for different reasons [5, 4].

In a traditional video player, designers follow the guideline for subtitles recommendations and place subtitles near the bottom of the screen. However, in a 3D environment like 360-degree video, defining the "bottom of the screen" becomes complicated. In addition, because of the characteristics of 3D environments, designers in such players also need to take into account additional challenges such as occlusion issues, immersion, locating, and virtual reality sickness [5, 4, 27]. In a study of subtitles design for 360-degree videos by Andy Brown et al., they proposed four caption designs based on these challenges [4]:

- **Evenly spaced:** divides 360 degrees into three 120-degree regions evenly and displays three identical subtitles at 0, 120, and 240-degree positions.
- **Follow head immediately:** the subtitles follow the head movement using the HUD mechanism, which can be moved in horizontal and vertical axes.
- **Follow with lag:** the subtitle will make a smooth movement after the viewer's head rotates beyond the 30-degree threshold; otherwise, it remains stationary.
- **Appear in front, then fixed:** the subtitle is placed in the center of the field of view and can be turned off/displayed at any time. It is fixed at the position every time it is displayed.

The user study results of [4] showed that all four designs were easy to use and achieved good user experience performance, but **Follow head immediately** had the best overall user experience. Regarding user behavior, some users still felt that the subtitle design tended to obscure the video content and cause motion sickness. In addition, the authors identified the two most essential attributes of subtitles in the test: ease of location and freedom to explore the scene.

As mentioned earlier, the interactive control panel design of the player has similar requirements to the subtitle design in terms of user experience. Compared to subtitle design, however, the interaction design does not need clarity and readability [27], and users spend little time looking at the interactive elements. Still, it does require more precise actions by the user. Based on the results of Brown et al.'s study, we can summarize the following requirements which also work for interactive control panel design:

- **Avoid obscuring:** avoid interactive elements obscuring video content.
- **Easy Locating:** the location of the interactive element is easy to find, controlling the user's effort to find the interactive element.
- **Avoidance of virtual reality sicknesses:** Virtual reality sickness refers to exposure to virtual environments that cause motion sickness-like symptoms (e.g., dizziness, nausea, sweating, etc.) and is a major factor affecting the VR user experience [20].

A.3 Design and Interaction Challenges for 360-degree video

A.3.1 Cybersickness

VR sickness, also known as cybersickness, is a significant problem in 3D environments and the biggest threat to the success of VR. Unfortunately, there is no complete solution to this problem for now [22, 5, 4, 20]. Existing studies have tended to decrease this problem by reducing exposure time or reducing visual stimuli to reduce the symptoms like dizziness and nausea [22, 14]. For example, Groth et al.'s study prevented cybersickness symptoms by applying blur and opaque occlusion to 360-degree video to constrain the FOV. The experiment result shows that this method can reduce the symptoms of cybersickness. In addition, they also mentioned that opaque occlusion of the peripheral FOV is more appropriate for passive viewing of 360-degree videos with intense motion [14]. As mentioned earlier, motion sickness is a vestibular symptom caused by postural incoordination, and cybersickness symptoms are similar to motion sickness. However, they are triggered by a different cause, with the latter occurring more often due to visual stimulation accompanied by a lack of vestibular stimulation than the former, which requires only vestibular stimulation [20]. In [20], the author identified three leading theories of cybersickness: the sensory conflict theory, the poison theory, and the postural instability theory. Subsequent studies have provided evidence for each of these theories. Hence, it is only possible to determine which theory is partially correct. In addition, the effects of cybersickness may vary across individuals [20].

Overall, the goal of our study is not to completely eliminate the effects of cybersickness but to build on previous successful designs, like using some methods to reduce the visual stimuli, to mitigate the suffering of cybersickness for viewers, and provide them with a better user experience.

A.3.2 Ergonomics/Comfortable User experience

In many previous user tests related to VR and 360-degree video applications, users have mentioned a certain degree of "discomfort". The "discomfort" here is different from the cybersickness that brings dizziness and nausea but refers more to the discomfort in the physical aspects, such as arm pain or neck pain. Such reactions may be due to prolonged non-biological movements, such as a large head turn or moving the arms [2, 29, 27]. Such extreme movements may be due to the nonergonomic design of movements. Because the HMD has a large FOR, the designer may design the interaction interface too large or too close to the user, resulting in the user having to operate in an uncomfortable position. In addition, because the user's head can also be used as an input in the HMD [2], some applications may rely too much on head input and cause the user to turn their head for a long time. Therefore, user comfort also needs to be taken into account when designing a suitable VR/360-degree video UI.

A.3.3 Impose A 2D Solution into A 3D Environment

Any immersive VR requires the user to maintain a constant presence in the application, which is an essential feature of immersive VR [28]. However, most existing GUI designs for VR applications force a solution from a 2D environment to present in a 3D environment [28, 2, 16]. For example, the UI of YouTube VR[42] takes the desktop design and places it in the 3D environment without much change in all the interaction design to adapt to the new 3D environment. Such a design would result in an uncomfortable reading/viewing experience because of the lower resolution of the application due to the computing power of the HMD. In addition, the 2D menu system and UI design are independent of the 3D environment, which can destroy the immersion [28].

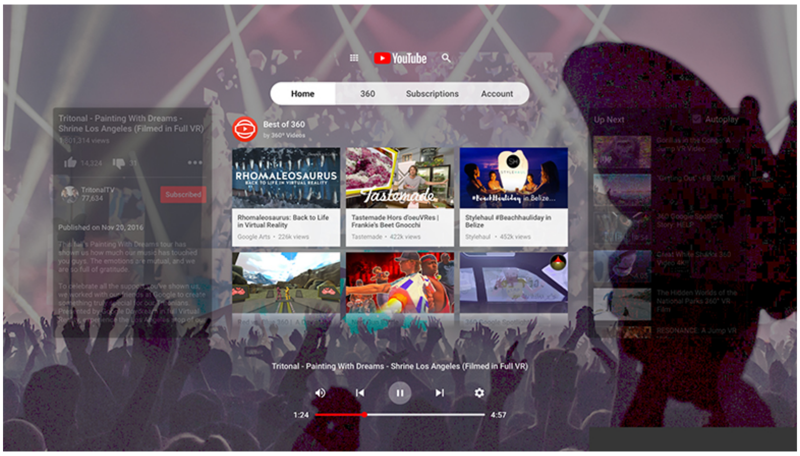


Figure 7: Example of YouTube VR interface [42]

Compared to the traditional 2D environment, the 3D environment has an additional depth cue [2]. Even a single depth cue can pose many problems for user interaction behavior. Within a certain distance, the user's vision can judge the distance of the object in front of them, and the wrong depth cue can lead to a more difficult operation for the user. In addition, because the FOR in HMD is extensive, some objects may be placed in places the user cannot see, thus affecting the operation.

Therefore, as mentioned in [28], a suitable VR UI should satisfy the following:

- Fit with the environment of the application/game.
- Refer to the interaction in daily life.
- Interactive elements can be easily accessible.
- The UI system can be accessed anywhere or in a specific location in the application.
- Consider usability guidelines.
- Encourage players to explore the environment.

The above requirements can also be used as a reference for the environment of the 360-degree video we focus on.

A.4 Conclusion

In this study, we will focus on the educational usage of 360-degree VR videos. Generally speaking, when viewing 360-degree VR videos, the viewer passively watches them and rarely interacts with them. However, in educational scenarios where the video is used, teachers will ask learners to interact more with the video. For example, “Please pause here and let us take a closer look at this building on the left side of the screen” or “Please skip to 1:30, where you can see this monkey picking on another monkey”. This also happens in educational scenarios using 360-degree VR video. According to [29], the technology of 360-degree video has brought significant development to the education field, showing some learning scenarios nearly realistically and allowing users to explore in a relatively safe environment. However, in regular teaching activities, just like we mentioned before, teachers may ask students to pause or replay several times to see the key points or to re-watch details they might have missed. Such scenarios require students to frequently interact with 360-degree videos in various ways.

Previous research[2, 28, 5, 4, 26, 32]has provided many innovative UI designs in VR environments. However, the main focus of these designs is to give the player a more immersive gaming experience, so the interface design would be more in line with the gaming world. While our research focuses on applications in the educational domain, with this in mind, our high-level goal is to make the system more accessible for students to use in the classroom. The standard 2D video player interactive interface is already prevalent, and people are more familiar with such a standard system, so we will keep the same design framework in our system. Also, immersive experience is less relevant in 360-degree video interaction. Unlike VR games, as long as the viewer actively interacts with the video, immersion in the content will be interrupted.

Nevertheless, the viewer wants to avoid the presence of the interaction control panel blocking their view while watching. As [5, 4, 27] mentioned, users prefer to place the interaction control panel at the bottom-center in traditional video players of the screen, which minimizes the blocked area. However, in a 3D environment, the screen is not in a fixed window, and the viewers’ FOV changes as their heads rotate, so we cannot say precisely where the ”bottom-center” of the screen is. We have no idea where is the best range of ”bottom”. Besides, the size of the video is not equal to the viewer’s FOV. There are more options for the size of the interaction control panel. Existing research and design guidelines recommend button size and spacing [2, 8, 37]. However, no studies or guidelines mentioned the length of the timeline. Therefore, we hope our research will provide a more comfortable, accurate, and fast interaction design recommendation for the 360-degree video player in educational usage.

Briefly, our study will focus on the **vertical location** of the interaction control panel and the **length** of the timeline in the 360-degree VR video player. First, the vertical position of the panel. The panel’s position can be divided into two aspects, one relative to the virtual world and another relative to the user. The position is relative to the video is also presented in many existing guidelines [9, 2]. The guidelines specify a comfortable depth range (0.5m - 20m) of video projection. For the position of the interaction control panel relative to the user (or camera), some guidelines and studies mention that users prefer to place interactive elements, such as the timeline, in “the center of the bottom”. It is easy to understand “center” in a 3D environment. Like most existing players and designs, we will keep the horizontal position of the panel directly in front of the user’s eyes/camera and follow the horizontal movement of the head. However, defining the “bottom” for 3D environments is challenging. Some guidelines suggest going 15 degrees below eye level, while some subsequent experiments[4] have shown that below 15 degrees can still cause some degree of obscuring and discomfort. Meanwhile, there is also some ergonomic requirement of the human head’s range of motion mentioned in the guidelines. For example, Figure 4 mentioned the normal line of sight sitting is 15 degrees and cannot be lower than 70 degrees. And ergonomically speaking, keeping your head down is a little easier than up. Because on a human vertical view field, with the eye and head rotation, we can reach -70 degrees below but only +50 degrees above. Therefore, it is worth exploring the best vertical location of the interaction control panel.

The interaction control panel of existing video players consists of several elements, including some interactive buttons (such as pause/play, forward/rewind) and an interactive timeline. Some of the existing VR glasses, HMDs, and game engine companies have given developer guidelines about the button size requirements [9, 2]. However, these guidelines rarely mention the requirements for timeline length. In a 3D environment, the viewer's FOV will not simply be limited to a fixed window, while the FOR differs from the viewer's FOV. The timeline has more space to become longer to achieve smaller granularity. However, how long it should be is also considerable. An excessively long timeline can cause people to put more effort into interaction, which may bring about arm or neck discomfort. It is also possible that the length of the timeline overextends the FOV of the viewer, causing them not able to know the current time position. An excessively short timeline may make time more granular and less accurate as the viewer interacts with the timeline. Therefore, finding an appropriate timeline length is also a point we can explore more deeply.

In summary, the vertical location of the interaction control panel and the length of the timeline should be further investigated to find a suitable interface design for educational usage. To verify if an interface design is suitable in an educational context is the education outcome. For a good learning outcome, people need to be able to interact easily and quickly with the player. For example, users might gain higher accuracy when they can access concrete positions more easily and quickly.

Therefore, this study investigates how an interface design will affect users' performance on various tasks regarding user experience, task accuracy, and speed. To do this, we will design a two-step experiment using different vertical locations of the interaction control panel and timeline lengths, then measure the performance. We hope this research will shed light on improving users' learning experience by introducing a better interface design.

B Methodology Details

B.1 Experiment Plan

B.1.1 Material

To conduct the experiment, we prepared the first-generation Oculus Quest VR Headset (with controllers), a tablet computer, and fixed chairs. The experimental system was developed using Unity and ran on a laptop, with real-time transfer to the Oculus Quest via Quest Link. Participants interacted with the system using controllers and ray casting. The questionnaire implemented through the Qualtrics platform is filled out using the tablet. Throughout the experiment, participants were seated in fixed chairs chosen to restrict physical movement while allowing easy reorientation to the video's "front". Besides, it is also a suitable choice for an educational context like a classroom.

For this experiment, we developed a Unity-based 360-degree video player incorporating essential features for the study. The player included a video playback function with image and sound components, a "Finished" button, tutorials, and task descriptions specific to the experiment. Additionally, an interaction log automatically captured participants' interactions, such as clicks/grabs on the slider, button presses, and task completion time.

To ensure the experimental results remain unaffected by extraneous variables and to minimize participants' adaptation time to the player, we replicated the design of some existing 360-degree video players. This involved creating a user interface with the following interactive elements (see Figure 8):

- Play/pause button
- Clickable timeline with a draggable slider for adjusting the playback position
- Fast forward and fast backward buttons (disabled during the experiment to focus on observing participants' interactions with the timeline)
- Thumbnails that appear above the timeline while the slider is being dragged
- "Finished" button to indicate task completion

Figure 8: Screenshot of video player interface

In terms of player dimensions, we adopted the settings used in the study by Brown et al. on Subtitles in 360-degree Video [5, 4]. The camera is fixed at the center of a 12.5m radius, equirectangular, non-stereoscopic large sphere screen. The videos are texture-mapped onto the inside surface of the sphere screen, allowing the camera view to be rotated with 3 degrees of freedom (DOF) based on input from the Oculus Quest. Furthermore, the control panel is positioned 4m away from the camera (see Figure 9).

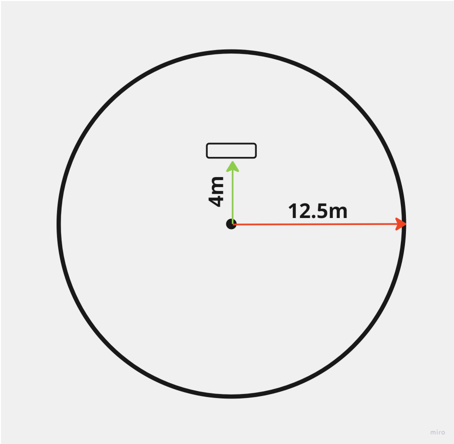


Figure 9: The position of the spherical screen and camera, control panel: the black dot indicates the camera position; the green line indicates the distance from the camera to the control panel is 4m; the red line indicates the distance from the camera to the screen is 12.5m

In addition to the aforementioned hardware and software materials, our experiment required several additional materials, including ten 360-degree videos (one for the tutorial and nine for the formal experiment), a questionnaire, and a consent form. To ensure consistency, all videos were selected to have a similar length of approximately 2 minutes, thereby ensuring that under the same timeline length, the timeline has the same granularity across all videos. And all videos are shot with fixed cameras, with audio, and the content is popular science or introduction. Besides, the details of the questionnaire and consent form will be described in sections B.1.6 and B.2.3.

B.1.2 Experiment Conditions

In this study, our focus revolves around two independent variables: the **length** of the timeline and the **vertical location** of the control panel. For each independent variable, we carefully selected three parameters.

Regarding the length of the timeline (hereinafter referred to as "**length**"), we designated **Short, Medium,** and **Long** parameters. These choices were determined by considering the horizontal field of view (FOV) of humans (see Figure 4), combined with the FOV of the Oculus Quest (see Figure 2). Refer to the following Table 1 for a comprehensive description.

Parameters of Length	Range	Reasons
Short	L15°- R15°	To maintain a consistent 30°difference between the chosen parameters, we selected the shortest timeline within the range
Medium	L30°- R30°	Symbol recognition zone of human vision (according to Figure 4)
Long	L45°- R45°	Maximum FOV of the Oculus Quest (according to Figure 2)

Table 1: Length's three parameters and the reasons for their selection

For the vertical location of the control panel (hereinafter referred to as "**vertical location**"), we also took into account the vertical FOV of humans (see Figure 4). Considering that raising the head may cause discomfort to the neck, we excluded vertical locations above the horizontal plane. Consequently, we selected three parameters: **Lower, Standard,** and **Higher**. Further details can be found in Table 2 below.

Parameters of Vertical Location	Range	Reasons
Lower	-30°	The maximum downward rotation of the human eye without turning the head (according to Figure 4)
Standard	-15°	Normal line of sight sitting (according to Figure 4)
Higher	0°	Standard line of sight (according to Figure 4)

Table 2: Vertical location's three parameters and the reasons for their selection

Given that the primary objective of this experiment is to identify the optimal combination of length and vertical location, we combined the two sets of three parameters for the independent variables. This resulted in nine distinct experimental conditions (hereinafter referred to as "**conditions**"), as outlined in Table 3.

Length \ Vertical Location	Lower	Standard	Higher
Short	L-S	S-S	H-S
Medium	L-M	S-M	H-M
Long	L-L	S-L	H-L

Table 3: 9 experimental conditions

It is important to highlight that the size of the control panel remained constant across different conditions in order to ensure that it did not influence the experimental results. Only the independent variables, which are the length and vertical location, changed across different conditions.

To mitigate the influence of learning effects on the experimental outcomes, we employed a balanced Latin square approach to assigning the order in which participants would test the nine conditions. The experimental order of participants will be elaborated on in section B.1.4.

B.1.3 Experiment Tasks

This section provides an overview of the two types of tasks conducted in the experiment, specifically designed to simulate actions in an educational context. The tasks are as follows:

- Positioning Task (Tpos): Participants manipulate the timeline to accurately jump to the target time specified in the task instruction (e.g., Jump to 1:30). A time range of +/-2 seconds from the target time is considered a correct completion.
- Search Task (Tsr): Participants locate the target scene based on the provided scene description and approximate range (e.g., finding a scene of a monkey nursing its baby in the first half of the video). Any time in the range of the target scene range is considered a correct completion.

Each condition in the experiment includes two Tpos (Tpos1 and Tpos2) and one Tsr. Tpos1 requires participants to jump from left to right, covering a longer distance, while Tpos2 requires participants to jump from right to left, covering a shorter distance. The purpose of having two Tpos is to investigate potential differences in participants' behavioral patterns (e.g., choose grab or click the timeline) for different jump directions. Regarding to Tsr, the participants are asked to find a certain scene in the video as soon as possible. The instruction of Tsr provides a description of the target scene and a cue indicating whether it appears in the first or second half of the video. To ensure participants can quickly locate the target scene across all three timeline lengths, we specified that the scene appears for at least 10-20 seconds within a 2-minute video and is positioned in front of the participant's view without significant head or body movement.

Throughout the experiment, participants are required to complete a total of 30 tasks, including the initial three in the tutorial. To simulate real-world educational scenarios, task instructions are provided solely in text form, without any accompanying pictures, audio, or video cues, emulating students wearing VR HMDs in a classroom environment and following textual instructions.

Each task follows the flow presented in Figure 10. Participants manually click the "Play" button to initiate video playback and start the timer, and then click the "Finished" button to end the task and stop the timer once they confirm task completion. Tpos1 and Tsr videos start from 0:00, while Tpos2 videos start from the target time of Tpos1.

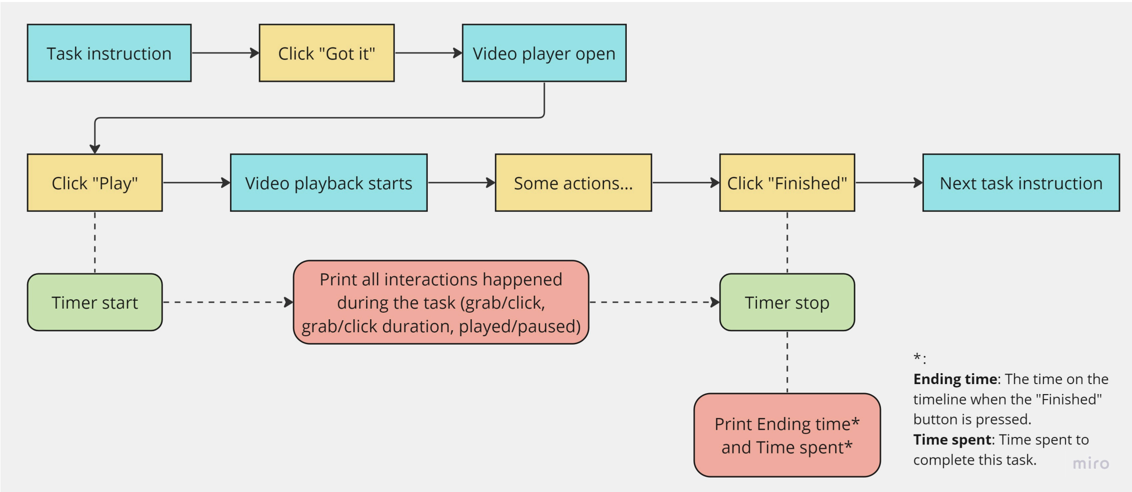


Figure 10: The flow of each task

To mitigate the impact of learning effects, nine different videos are utilized across the nine conditions, with an additional video dedicated to the tutorial. Table 4 provides details about the video names, duration, task descriptions, task instructions, and target times/ranges.

B.1.4 Experiment Procedure

This section provides a detailed account of the experimental procedure and the order in which participants engaged in the experiment.

As mentioned earlier, the experiment comprises a tutorial and nine conditions, utilizing a within-subject design where all participants partake in each condition.

As shown in Figure 11, each participant begins with an introduction delivered by the experimenter. This includes signing the experiment consent form, an overview of the experiment, and gathering basic participant information. After ensuring that participants have no questions regarding the experiment's content or purpose and willingly volunteer to participate, we assist them to wear the VR head-mounted display (HMD), preparing to start the tutorial. The tutorial replicates the conditions of the experiment, aiming to familiarize participants with the operation, understand the experiment process, and adapt to the VR environment. Upon completion of the tutorial, we pause to ask if participants have any questions regarding the task or need adjustments to the VR HMD. If there are no concerns, we proceed to the formal part of the experiment.

The formal experiment is divided into nine conditions and is followed by a post-experiment in-

Video number	Video name /total time	Task instructions
Tutorial	Catcafe [38] 2:00	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 1:45. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 1:20. Press "Finished" once you are there. Tsr: Please find this scene in the first half of the video: two cats sitting in front of a fish tank, one of the cats is staring at the fish tank, and another gray cat is looking at you. (right answer: 0:30-0:40)
1	Costa_Rica [30] 2:14	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 0:45. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 0:33. Press "Finished" once you are there. Tsr: Please find this scene in the second half of the video: a green round bench in the lavishly decorated hall of the National Theatre of Costa Rica. (right answer: 1:11-1:23)
2	Farm [3] 2:00	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 0:50. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 0:15. Press "Finished" once you are there. Tsr: Please find this scene in the second half of the video: a worker in a yellow apron is putting vegetables from a sink onto a conveyor belt. (right answer: 100-116)
3	Mexico [40] 1:53	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 1:24. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 1:00. Press "Finished" once you are there. Tsr: Please find this scene in the first half of the video: a street artist is painting a couple's portrait. (right answer031-039)
4	Manpupuner2 [39] 2:17	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 2:10. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 2:00. Press "Finished" once you are there. Tsr: Please find this scene in the first half of the video: a group of people on snowmobiles cross the border marker (a black circular sign standing on the ground with Russian letters on it). (right answer: 0:12-0:26)
5	Campus [41] 2:00	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 0:30. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 0:15. Press "Finished" once you are there. Tsr: Please find this scene in the second half of the video: a group of students is doing an experiment in the laboratory. (right answer: 1:29-1:45)
6	Spain [23] 1:59	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 0:45. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 0:30. Press "Finished" once you are there. Tsr: Please find this scene in the second half of the video: a man dressed in black standing by the Segovia Aqueduct (Aqueduct de Segovia) overlooking the city. (right answer: 1:18-1:28)
7	TheUnitedArabEmirates [24] 2:01	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 1:35. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 1:20. Press "Finished" once you are there. Tsr: Please find this scene in the first half of the video: an Arab man in blue showing a man in a mask around the city. (right answer: 0:49-1:00)
8	Oman [25] 2:07	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 0:50. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 0:35. Press "Finished" once you are there. Tsr: Please find this scene in the second half of the video: a group of tourists visiting the local people's house. (right answer: 1:28-1:35)
9	Manpupuner1 [39] 2:00	Tpos 1: Press the "play" button and set the slider to an area +/- 2 seconds around 0:48. Press "Finished" once you are there. Tpos 2: Press the "play" button and set the slider to an area +/- 2 seconds around 0:27. Press "Finished" once you are there. Tsr: Please find this scene in the second half of the video: a man is banging on a drum in the snow. (right answer: 1:11-1:29)

Table 4: Details of videos used in the experiment

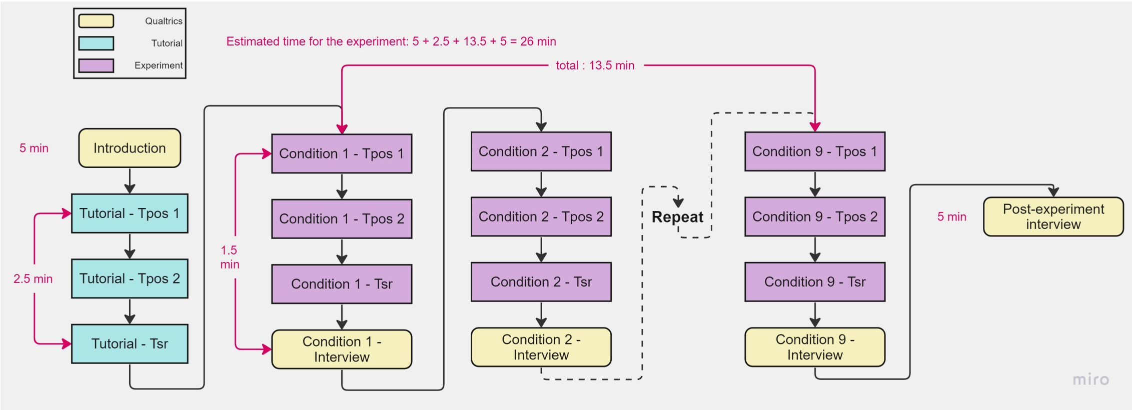


Figure 11: The flow chat of experiment procedure

terview. Each condition consists of three tasks (two Tpos and one Tsr) along with a brief rating question. To ensure efficiency, the experimenter directly poses the rating questions after each condition, allowing participants to respond verbally without needing to remove their VR HMDs. Once all nine conditions are completed, participants can remove the VR HMD and proceed to the post-experiment interview. The entire experiment is expected to last less than 30 minutes, and participants retain the freedom to pause the experiment or withdraw from participation at any time.

Because all participants are involved in nine conditions, a balanced Latin-square design was employed to allocate participants to the experimental order. This approach aims to minimize the influence of learning effects. While the order of the videos remained fixed, the order of the conditions was randomized using the balanced Latin square method. In addition, by balancing the Latin square design, we calculated that at least 18 participants were needed to complete the experiment. For a detailed overview of the experimental order, please refer to Table 5. It is important to note that the Standard-Medium setting was used in the tutorial. This choice was made due to its proximity to the existing standard setting, ensuring that the setting was not overly extreme and would facilitate ease of learning. The tutorial tasks mirror those of the subsequent conditions. To prevent any potential bias in the experimental data, a different video was used in the tutorial.

Video / Participant Number	1 Costa_Rica	2 Farm	3 Mexico	4 Manpupuner2	5 Campus	6 Spain	7 TheUnited AranEmirates	8 Oman	9 Manpupuner1
1	L-S	L-M	H-L	L-L	H-M	S-S	H-S	S-M	S-L
2	L-S	H-L	L-M	H-M	L-L	H-S	S-S	S-L	S-M
3	L-M	L-S	L-L	H-L	S-S	H-M	S-M	H-S	S-L
4	H-L	L-S	H-M	L-M	H-S	L-L	S-L	S-S	S-M
5	L-M	L-L	L-S	S-S	H-L	S-M	H-M	S-L	H-S
6	H-L	H-M	L-S	H-S	L-M	S-L	L-L	S-M	S-S
7	L-L	L-M	S-S	L-S	S-M	H-L	S-L	H-M	H-S
8	H-M	H-L	H-S	L-S	S-L	L-M	S-M	L-L	S-S
9	L-L	S-S	L-M	S-M	L-S	S-L	H-L	H-S	H-M
10	H-M	H-S	H-L	S-L	L-S	S-M	L-M	S-S	L-L
11	S-S	L-L	S-M	L-M	S-L	L-S	H-S	H-L	H-M
12	H-S	H-M	S-L	H-L	S-M	L-S	S-S	L-M	L-L
13	S-S	S-M	L-L	S-L	L-M	H-S	L-S	H-M	H-L
14	H-S	S-L	H-M	S-M	H-L	S-S	L-S	L-L	L-M
15	S-M	S-S	S-L	L-L	H-S	L-M	H-M	L-S	H-L
16	S-L	H-S	S-M	H-M	S-S	H-L	L-L	L-S	L-M
17	S-M	S-L	S-S	H-S	L-L	H-M	L-M	H-L	L-S
18	S-L	S-M	H-S	S-S	H-M	L-L	H-L	L-M	L-S

Table 5: Experimental order of participants

B.1.5 Data Gathered

In our experiments, we collected two main types of data: quantitative data and qualitative data. Here we listed the details about these two types of data:

- Quantitative data:
  - Correctness: This data indicates whether participants completed the task correctly and is represented by a Boolean value (0 for correct, 1 for incorrect).
  - Accuracy: It measures the deviation of the participant's answer (a timestamp) from the correct answer. Represented by an integer value, with 0 indicating a correct answer and a larger number representing a greater time offset.
  - Time spent: This data records the duration taken by participants to complete each task, represented in seconds.
  - Number of Grab/Clicks: It captures the count of grabs or clicks made by participants during each task, indicated by an integer value.
  - Grab/Click duration: This data represents the interval of time between each grab or click made by participants, measured in seconds.
- Qualitative data:
  - Difficulty level: Participants' subjective perception of the difficulty level associated with each condition. Collected using a numerical scale ranging from 0 to 10, where 0 represents super easy and 10 represents super hard.
  - Cybersickness and discomfort level: It assesses the level of discomfort experienced by participants during the experiment. Collected through a numerical scale ranging from 0 to 10, where 0 represents no discomfort and 10 represents severe discomfort.
  - Preference for length and vertical location: Participants' preferences and discomfort regarding the three parameters of length/vertical location. Collected through multiple-choice and short-answer questions in a questionnaire.
  - Overall feedback: Participants' general feedback on the experiment as a whole. Gathered through post-experiment interviews.
  - Participant information: This includes participants' gender, age, and level of experience with VR and 360-degree video. Collected via questionnaires.

Regarding data collection, quantitative data were obtained through Unity's debug log. The experimental system we utilized incorporated a function to automatically record interaction logs. All participant interactions within the experiment were printed through Unity's debug log function and saved in a text file for subsequent analysis. The log includes details such as the start/end of the timer, timer results, timeline timestamps when participants grabbed/clicked the timeline, release timestamps, task completion timestamps (when participants pressed the "Finished" button), play/pause button press, and corresponding feedback. An example of an interaction log is illustrated in Figure 12 below.

```
[2023-05-01 00:49:37] [Timer] Timer start!
[2023-05-01 00:49:39] [Timeline] press: 0:01
[2023-05-01 00:49:43] [Timeline]slider grabbed.
[2023-05-01 00:49:43] [Timeline] release: 1:44
[2023-05-01 00:49:45] [Timer] Time spent in this task: 00:07
[2023-05-01 00:49:45] [Timeline] Ending Time = 1:45
[2023-05-01 00:49:48] [Timer] Timer start!
[2023-05-01 00:49:50] [Timeline] press: 0:02
[2023-05-01 00:49:53] [Timeline]slider grabbed.
[2023-05-01 00:49:53] [Timeline] release: 0:30
[2023-05-01 00:49:55] [Timer] Time spent in this task: 00:06
[2023-05-01 00:49:55] [Timeline] Ending Time = 0:32
```

Figure 12: An example screenshot for the interaction log

In addition, to ensure the reproducibility of all collected quantitative data, we will record the system screen and participants' movement during the experiment. These video recordings are solely intended for subsequent analysis and will not be shared publicly. To uphold confidentiality, all collected data will be anonymized, stored locally, and deleted after the analysis is concluded.

Regarding data analysis, we will classify the data based on its characteristics and employ various methods for analysis. For quantitative data, we will utilize the chi-square test to assess the statistical significance of correctness, validating our conclusions. To analyze accuracy and time spent, which involve multiple variables and repeated measurements, we will employ two-way repeated measures ANOVA. The number of grabs/clicks will be shown in charts. Regarding grab/click duration, we will initially evaluate the overall patterns using box plots and subsequently employ a more detailed approach to analyze our observations. As for qualitative data, we will first use the chart to show the data and then apply subsequence analysis. All qualitative data will be anonymized to ensure confidentiality. More analysis details will be shown in the following section C.

B.1.6 Questionnaire

In this experiment, we employed a custom questionnaire designed specifically to gather the required data for our study. The questionnaire was developed by incorporating elements from the System Usability Scale (SUS) and making modifications to a questionnaire used in a previous study by Vermast [35] and Laudisa [19]. Given the similarities between Vermast and Laudisa's research and our own, we drew inspiration from both questionnaires to create a comprehensive instrument for our experiment.

This questionnaire (including the consent form) was distributed through the Qualtrics platform. In the experiment, the experimenter will use a tablet to collect the responses to the questionnaire. In the questionnaire, Q1 refers to the participant number, which is filled in by the experimenter. Q2 serves as the information sheet for the experiment, providing essential details to the participants. Q3 is the consent form, ensuring the participants' voluntary participation. Q4-8 capture personal information and gather data on participants' experience with VR and 360-degree video. Q9-26 consist of difficulty level questions specific to each condition. Each condition includes two questions: the first records the condition as filled in by the experimenter, while the second employs a 0-10 scale for difficulty level. Q27-35 comprise the post-experiment interviews. Specifically, Q27-30 address cybersickness and discomfort-related concerns. Q31-32 pertain to timeline length-related questions, while Q33-34 pertain to the control panel's vertical location-related questions. Q35 allows participants to provide any additional feedback. Below are the questions in the questionnaire. The details of the information sheet and consent form will be elaborated in Section B.2.3.

Experiment Questionnaire

Q1 Participant number (Filled by researcher)

____

Q2 **Research Participant Information Sheet**

Skip To: End of Survey If Research Participant Information Sheet 360-degree Video Player Interaction Design for VR HMDs 1... = I disagree with it and decide not to partake

Q3 **Consent form for participation in the research project**

Skip To: End of Survey If Consent form for participation in the research project 360-degree Video Player Interaction Design... = I disagree with it and decide not to partake

Q4 What is your gender?

- ○ Male
- ○ Female
- ○ Other/prefer not to say

Q5 What is your age?

- ○ 18-24
- ○ 25-34
- ○ 35-44
- ○ 45-54
- ○ 55-64
- ○ 65-74
- ○ 75-84
- ○ 85 or older
- ○ Prefer not to say

Q6 How familiar are you with Virtual Reality (VR)?

○ Never used it

○ Used it a few times (e.g., tried it a few times)

○ Use it every now and then (e.g., once every few months)

○ Use it regularly (e.g., more than once a month)

○ Use it frequently (e.g., more than once a week)

Q7 Do you have any experience with 360-degree videos?

○ Yes

○ No

Skip To: End of Block If Do you have any experience with 360-degree videos? = No

Q8 If yes, on which devices have you watched/do you watch 360-degree videos?

○ Flat screen (desktop PC, laptop, TV)

○ Mobile screen (phone, tablet)

○ Head-mounted devices (VR headsets such as Oculus Rift, Oculus Quest, HCT Vive)

○ Other, please specify: ____

Difficulty level rating (repeating in all 9 conditions)

Q9 Condition (Filled by researcher)

____

Q10 By using this setting, how hard was it to solve these tasks (on a scale from 0 = super easy, to 10 = super hard)

- ○ 0
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

Q27 Did you experience any headaches during or shortly after the VR experiment?

- ◯ Yes
- ◯ Maybe
- ◯ No

Q28 Did you experience any dizziness during or shortly after the VR experiment?

- ◯ Yes
- ◯ Maybe
- ◯ No

Q29 Did you experience any nausea during or shortly after the VR experiment?

- ◯ Yes
- ◯ Maybe
- ◯ No

Q30 Please rate the uncomfortable level on the following scale.

- ◯ 0
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8

○ 9

○ 10

Q31 For the different timelines, which length would you prefer in a final interface (multiple answers are possible)?

☐ Long

☐ Medium

☐ Short

Q32 Did you experience any discomfort because of the length of the timeline in some conditions?

○ No

○ Yes. Please give a short description of your discomfort:

____

Q33 Which vertical placement of the interaction elements would you prefer in a final interface (multiple answers are possible)?

☐ Horizontal/Higher

☐ Middle/Standard

☐ Bottom/Lower

Q34 Did you experience any discomfort because of the vertical placement of the interaction elements?

○ No

○ Yes. Please give a short description of your discomfort:

____

Q35 Do you have any other comments or feedback?

____

B.2 Other Details

B.2.1 Experimental Environment

In this experiment, we established specific requirements for the experimental environment, aiming for a quiet and spacious room with a fixed chair. As there was no dedicated experimental facility available, we had to identify suitable environments that met our criteria for each round of participant testing.

During the experiment, participants were equipped with a VR HMD and comfortably seated in a designated fixed chair. They held a controller in their right hand to interact with the virtual environment. An experimenter was present on the sidelines to supervise and monitor the proceedings. To capture the participant's movements, video recording was carried out using the experimenter's cell phone. Additionally, we utilized the built-in video recording software of the Windows operating system to capture the interior view of the experimental system.

B.2.2 Ethical Review, Risk Assessment, and Safety Measures

This study has undergone an ethical review by Utrecht University to ensure the protection and well-being of participants. Prior to commencing the experiment, participants were provided with comprehensive information regarding any potential risks involved. They were fully informed about the entire experimental procedure.

Regarding personal information and data security, we have implemented strict measures to safeguard the collected data. All data obtained during the experiment will be used solely for the purposes of this study. It will be securely stored on the researcher's local device, protected by a password. Furthermore, all data will be permanently deleted within four months following the completion of the study. Any personal information gathered during the experiment will be fully anonymized to ensure confidentiality.

Specific details regarding the potential risks associated with the study are outlined in the consent form, which participants signed prior to their participation. The main identified risks include cybersickness and discomfort stemming from the use of the VR HMD. Participants were explicitly informed that they had the right to pause or withdraw from the experiment at any time if they experienced any form of discomfort.

B.2.3 Participant Information Sheet and Consent Form

In this section, we show the information sheet used in the experiment as well as the consent form.

Research Participant Information Sheet

Investigating the Impact of Timeline Length and Control Panel Vertical Location in 360-Degree Video Players on User Experience and Behavior

1. Introduction

This study evaluates the optimal placement and size of the interactive elements of a 360-degree video player in VR HMDs – that is, how long timelines should be and where the interactive elements should be located. The experiment will be conducted in a quiet indoor space. Participants are asked to wear and operate a VR headset during the experiment.

2. Who will carry out the study?

This study is carried out by me, Shiyi Chen (s.chen12@students.uu.nl) as part of my master thesis under the supervision of Wolfgang Hürst (huerst@uu.nl).

3. How will the study be carried out?

In this study, you will wear the Oculus Quest VR headset to experience nine designs of a 360-degree video player. For design you will be asked to perform three tasks. After completing the tasks for one design, the researcher will ask you one question about it. After completing the tasks for all designs, the researcher will ask you a few general questions and give you the opportunity to make additional remarks. The experiment will take about 30 minutes. You can ask for a break anytime during the experiment.

4. What will we do with your data?

If you consent to this, a video recording will be made. This recording and all other data that is captured during this study (e.g., your answers to the questions and sensor data related to your interaction with the interface) will be stored on a secure university server. The recording will be transcribed so that participants' opinions are captured into text. The video will be securely deleted after transcription (within 4 months of the study). The transcribed text will be anonymized so that you will not be identifiable. The transcript will become part of my thesis and will also be stored in a data repository for use by other researchers and research users. My thesis, any publications based on this research, and the data repository will not include your name or any other individual information by which you could be identified.

5. What are your rights?

Participation is voluntary. We are only allowed to collect your data for our study if you consent to this. If you decide not to participate, you do not have to take any further action. You do not need to sign anything. Nor are you required to explain why you do not want to participate. If you decide to participate, you can always change your mind and stop participating at any time, including during the study. You will even be able to withdraw your consent after you have participated. However, if you choose to do so, we will not be required to undo the processing of your data that has taken place up until that time. The personal data we have obtained from you up until the time when you withdraw your consent will be erased (where personal data is any data that can be linked to you, so this excludes any already anonymized data).

6. Approval of this study

This study has been allowed to proceed by the Research Institute of Information and Computing Sciences on the basis of an Ethics and Privacy Quick Scan. If you have a complaint about the way this study is carried out, please send an email to: ics-ethics@uu.nl. If you have any complaints or questions about the processing of personal data in general, please send an email to the Faculty of Sciences Privacy Officer: privacy-beta@uu.nl. The Privacy Officer will also be able to assist you in exercising the rights you have under the GDPR. For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information at www.uu.nl/en/organisation/privacy.

7. More information about this study?

If you have any questions or concerns about this research please contact me, Shiyi Chen (s.chen12@students.uu.nl) or my supervisor Wolfgang Hürst (huerst@uu.nl).

8. Appendices:

- Consent Form
- Experiment Questionnaire

o I understand and want to continue with the consent form.

o I disagree with it and decide not to partake.

Consent form for participation in the research project

Investigating the Impact of Timeline Length and Control Panel Vertical Location in 360-Degree Video Players on User Experience and Behavior

Please read the statements below and tick the final box to confirm you have read and understood the statements and upon doing so agree to participate in the experiment.

I confirm that I am at least 18 years of age or older.

I confirm that the research project "Investigating the Impact of Timeline Length and Control Panel Vertical Location in 360-Degree Video Players on User Experience and Behavior" has been explained to me. I have had the opportunity to ask questions about the project and have had these answered satisfactorily. I had enough time to consider whether to participate.

I consent to the material I contribute being used to generate insights for the research project "Investigating the Impact of Timeline Length and Control Panel Vertical Location in 360-Degree Video Players on User Experience and Behavior".

I consent to audio recordings being used in this study as explained in the information sheet. I understand that I can request to stop recordings at any time.

I consent to video recordings being used in this study as explained in the information sheet. I understand that I can request to stop recordings at any time.

I consent to sensor recordings being used in this study as explained in the information sheet. I understand that I can request to stop recordings at any time.

I understand that if I give permission, the audio/video/sensor recordings will be held confidentially so that only Shiyi Chen and Wolfgang Hürst have access to the recording. The recordings will be stored in a secure way for up to 4 months after which period they will be securely destroyed, fully anonymized, and transcribed. In accordance with the General Data Protection Regulation (GDPR), I can have access to my recordings and can request them to be deleted at any time during this period.

I understand that personal data will be collected from me and that this information will be held confidentially so that only Shiyi Chen and Wolfgang Hürst have access to this data and are able to trace the information back to me personally. The information will be stored in a secure way for up to 4 months after which period it will be deleted, fully anonymized, and completely destroyed. In accordance with the General Data Protection Regulation (GDPR) I can have access to my information and can request my data to be deleted at any time during this period.

I understand that my participation in this research is voluntary and that I may withdraw from the study at any time without providing a reason, and that if I withdraw any personal data already collected from me will be erased.

I understand that my participation is not a requirement for my study and that participating or not will not impact me or my studies in any way.

I consent to allow the fully anonymized data to be used in future publications and other scholarly means of disseminating the findings from the research project.

I understand that the data acquired will be securely stored by researchers, but that appropriately anonymized data may in the future be made available to others for research purposes. I understand that the University may publish appropriately anonymized data in appropriate data repositories for verification purposes and to make it accessible to researchers and other research users.

I understand that I can request any personal data collected from me to be deleted.

o I confirm that I have read and understood the above statements and agree to participate in the study.

o I disagree with it and decide not to partake.

C Experiment Results

C.1 Quantitative Results

In this experiment, a total of 18 participants were recruited, and each participant took part in 9 conditions, encompassing three tasks: two positioning tasks and one search task. As a result, we gathered a data set comprising 486 sets of interaction logs (18 participants x 9 conditions x 3 tasks).

In the subsequent section, we conducted an analysis of these 486 logs, categorizing them into different groups, and obtained quantitative results.

C.1.1 Correctness of Tasks

The correctness of task completion is denoted by a Boolean value, with "correct" assigned as 0 and "incorrect" as 1, representing the percentage of participants who completed the task correctly. To explore the relationship between task correctness and variables such as the timeline length, the control panel's vertical location, and conditions, a comprehensive analysis was conducted from these three perspectives. Due to the non-normal distribution of the binary data, which consisted of a substantial sample size of 162 logs, the chi-square test was employed to assess the statistical significance of the hypotheses.

C.1.1.1 Length Table 6 illustrates the relationship between different timeline lengths and correctness. Participants achieved similar correctness in the Short and Medium groups, while the Long group exhibited higher correctness compared to the other two. This suggests a potentially higher operational accuracy in the Long group. However, a chi-square test on the correctness of all three groups indicated no significant difference between the Long and the other two groups (p=0.41). This outcome could be attributed to the influence of factors such as task type or the interaction between the length and vertical location.

Length	correct number	Total number	Correctness %
Short	145	162	89.5
Medium	145	162	89.5
Long	151	162	**93.2**

Table 6: Correctness s for different lengths

C.1.1.2 Vertical Location From Table 7, it is evident that the correctness was higher in the Lower group compared to the other two groups, implying that the Lower setting may be associated with higher operational accuracy. To test this assumption, a chi-square test was performed on the correctness rates of the three groups. However, the results indicated no significant difference in correctness rates between the Lower group and the other two groups (p=0.52). Similar underlying factors may contribute to this outcome.

Vertical Location	correct number	Total number	Correctness %
Lower	150	162	**92.6**
Standard	144	162	88.9
Higher	147	162	90.7

Table 7: Correctness s for different vertical locations

C.1.1.3 Different Conditions To further examine whether correctness is associated with the combination of the two independent variables, the interaction logs were categorized based on different conditions. Table 8 presents the correctness for each condition. It can be observed that participants achieved the highest correctness in the Higher-Long condition. Nevertheless, the chi-square test results indicated no significant difference in correctness across different conditions (p=0.25). Therefore, it can be concluded that there is no substantial variation in correctness among participants across different conditions, and the interaction effect did not affect the correctness.

Vertical Location	Length	correct number	Total number	Correctness %
Lower	Short	50	**54**	92.6
	Medium	50	54	92.6
	Long	50	**54**	92.6
Standard	Short	45	54	83.3
	Medium	50	54	92.6
	Long	49	54	90.7
Higher	Short	50	54	92.6
	Medium	45	54	83.3
	Long	52	54	96.3

Table 8: Correctness s for different conditions

In conclusion, the statistical analysis conducted revealed no significant correlation between participants' task correctness and the timeline length, control panel's vertical location, or different conditions.

C.1.2 Accuracy of Each Task

In this section, we assessed the accuracy of each task using the offset value in the interaction logs. The offset value represents the time difference between the participant's response upon pressing the "Finished" button and the correct time point. A larger offset value indicates a greater deviation, with 0 indicating a correct response. To examine the normality assumption of the offset values for the Tpos1 (positioning task 1), Tpos2 (positioning task 2), and Tsr (search task) groups, a Shapiro-Wilk test was conducted. The test results indicated violations of the normality assumption, with small statistics and p-values below 0.05. However, due to the lack of a more suitable alternative for our data and the robustness of ANOVA in the presence of normality violations, we proceeded with a two-way repeated measures ANOVA for data analysis.

Group	p-value
Tpos1	2.2995888573615365e-27
Tpos2	3.0152457816939717e-25
Tsr	5.821936106582731e-23

Table 9: Shapiro-Wilk test results for 3 groups ($p>0.05$ means data has normality)

Due to the presence of extreme values in the results (e.g., some participants misremembering the target time in the task), the Three-Sigma Rule was applied to identify and exclude these values from the analysis to ensure that they do not unduly influence the results. Figure 13 and Table 10 showed the result: one case in the Tpos1 group and one in the Tpos2 group were excluded from the accuracy analysis.

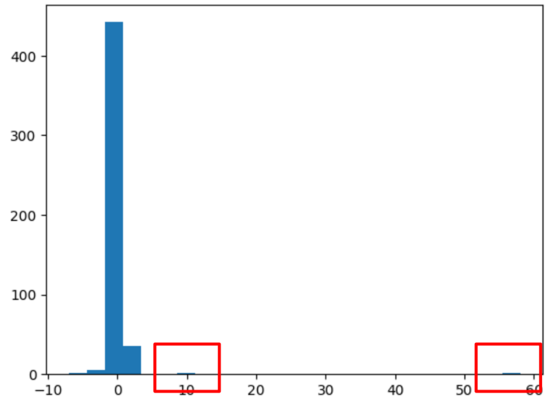


Figure 13: Results of the Three-Sigma Rule Screening Process

index	task	offset
108	Tpos1	58
196	Tpos2	9

Table 10: Records that need to be removed

Following the ANOVA analysis for each of the three data groups, the results are presented in the following tables, where "Factor A" corresponds to vertical location and "Factor B" corresponds to length:

	sum_sq	df	F	PR(>F)
Intercept	0.266678	1.0	0.591299	0.443118
FactorA	0.333333	2.0	0.369547	**0.691671**
FactorB	1.925926	2.0	2.135159	**0.121782**
FactorA: FactorB	4.130786	4.0	2.289778	**0.062349**
Subject	1.065229	1.0	2.361911	0.126423
Residual	68.101438	151.0	NaN	NaN

Table 11: ANOVA Results of Tpos1 group

	FactorA	FactorB	Subject	Mean
0	Higher	Long	242.500000	0.000000
1	Higher	Medium	242.500000	0.444444
2	Higher	Short	242.500000	0.111111
3	Lower	Long	242.500000	0.166667
4	Lower	Medium	250.352941	0.000000
5	Lower	Short	242.500000	0.000000
6	Standard	Long	242.500000	0.000000
7	Standard	Medium	242.500000	0.000000
8	Standard	Short	242.500000	0.444444

Table 12: Mean offset values of Tpos1 group

	sum_sq	df	F	PR(>F)
Intercept	0.154542	1.0	0.525879	0.469467
FactorA	0.703704	2.0	1.197289	**0.304862**
FactorB	0.033010	2.0	0.056164	**0.945404**
FactorA: FactorB	1.533352	4.0	1.304431	**0.270942**
Subject	0.000883	1.0	0.003003	0.956369
Residual	44.374934	151.0	NaN	NaN

Table 13: ANOVA Results of Tpos2 group

	FactorA	FactorB	Subject	Mean
0	Higher	Long	243.500000	0.111111
1	Higher	Medium	243.500000	0.166667
2	Higher	Short	246.235294	0.117647
3	Lower	Long	243.500000	0.000000
4	Lower	Medium	243.500000	0.055556
5	Lower	Short	243.500000	0.333333
6	Standard	Long	243.500000	0.277778
7	Standard	Medium	243.500000	0.000000
8	Standard	Short	243.500000	0.111111

Table 14: Mean offset values of Tpos2 group

	sum_sq	df	F	PR(>F)
Intercept	0.172050	1.0	1.504770	0.221836
FactorA	0.148148	2.0	0.647860	**0.524602**
FactorB	0.111111	2.0	0.485895	**0.616098**
FactorA: FactorB	0.469136	4.0	1.025779	**0.395924**
Subject	0.176414	1.0	1.542936	0.216094
Residual	17.379142	152.0	NaN	NaN

Table 15: ANOVA Results of Tsr group

	FactorA	FactorB	Subject	Mean
0	Higher	Long	244.500000	0.055556
1	Higher	Medium	244.500000	0.111111
2	Higher	Short	244.500000	0.000000
3	Lower	Long	244.500000	0.166667
4	Lower	Medium	244.500000	0.111111
5	Lower	Short	244.500000	0.055556
6	Standard	Long	244.500000	0.166667
7	Standard	Medium	244.500000	0.222222
8	Standard	Short	244.500000	0.333333

Table 16: Mean offset values of Tsr group

The analysis revealed no significant differences in offset values among the three tasks (Tpos1, Tpos2, and Tsr) based on vertical location or length. Furthermore, the interaction between these two independent variables did not yield any significant differences across all three tasks.

Upon examining the mean offset values, we found that the Tpos1 task exhibited a mean value of 0 for the conditions of Higher-Length, Lower-Medium, Lower-Short, Standard-Length, and Standard-Medium. This indicates 100% accuracy in these five conditions. Conversely, the Higher-Medium and Standard-Short conditions had higher mean offset values, suggesting relatively lower accuracy in these two conditions.

In the Tpos2 task, the conditions of Lower-Long and Standard-Medium displayed an average offset value of 0, indicating high accuracy. In contrast, the Lower-Short condition showed the largest average offset value, suggesting lower accuracy. Therefore, participants achieved 100% accuracy in the Lower-Long and Standard-Medium conditions while exhibiting the lowest accuracy in the Lower-Short condition.

Regarding the Tsr task, only the Higher-Medium condition exhibited an average offset value of 0, indicating optimal accuracy in completing the search task. On the other hand, the Standard-Short condition displayed the highest average offset value, implying lower accuracy. Thus, participants achieved 100% accuracy when using the Higher-Medium condition for the search task, while the lowest accuracy was observed in the Standard-Short condition.

C.1.3 Time Spent on Each Task

This section focuses on exploring the relationship between the time spent on different tasks and their corresponding vertical location and length. The time spent data extracted from the interaction logs were used for this analysis. Here, the "time spent" refers to the duration between pressing the "Play" button to initiate video playback and pressing the "Finish" button to finish the task.

To assess the normality assumption of the data, the Shapiro-Wilk test was conducted. The results indicated that the Tpos1, Tpos2, and Tsr groups exhibited low statistics and p-values below the significance level of 0.05, suggesting a violation of the normality assumption. However, similar to the previous section C.1.2, considering the lack of a more suitable alternative for our data context and the robustness of ANOVA in handling violations of normality, we proceeded with a two-way repeated measures ANOVA. Notably, the data set did not have any restrictions on task completion time, rendering the exclusion of extreme values unnecessary.

Group	p-value
Tpos1	2.721574968966099e-18
Tpos2	6.911286820621143e-11
Tsr	1.10447176578099e-12

Table 17: Shapiro-Wilk test results for 3 groups (p>0.05 means data has normality)

	sum_sq	df	F	PR(>F)
Intercept	702.948764	1.0	20.594386	0.000011
FactorA	98.111111	2.0	1.437187	**0.240805**
FactorB	101.777778	2.0	1.490899	**0.228444**
FactorA: FactorB	106.209877	4.0	0.777911	**0.541169**
Subject	0.112905	1.0	0.003308	0.954212
Residual	5188.220428	152.0	NaN	NaN

Table 18: ANOVA Results of Tpos1 group

	FactorA	FactorB	Subject	Mean
0	Higher	Long	242.500000	7.166667
1	Higher	Medium	242.500000	6.388889
2	Higher	Short	242.500000	9.611111
3	Lower	Long	242.500000	10.222222
4	Lower	Medium	242.500000	7.666667
5	Lower	Short	242.500000	8.833333
6	Standard	Long	242.500000	7.611111
7	Standard	Medium	242.500000	8.444444
8	Standard	Short	242.500000	9.944444

Table 19: Mean offset values of Tpos1 group

	sum_sq	df	F	PR(>F)
Intercept	612.943131	1.0	64.814512	2.205607e-13
FactorA	16.148148	2.0	0.853778	**4.278350e-01**
FactorB	50.333333	2.0	2.661202	**7.311883e-02**
FactorA: FactorB	49.740741	4.0	1.314935	**2.669415e-01**
Subject	1.165196	1.0	0.123211	7.260632e-01
Residual	1437.445915	152.0	NaN	NaN

Table 20: ANOVA Results of Tpos2 group

	FactorA	FactorB	Subject	Mean
0	Higher	Long	243.500000	6.888889
1	Higher	Medium	243.500000	5.388889
2	Higher	Short	243.500000	7.722222
3	Lower	Long	243.500000	6.333333
4	Lower	Medium	243.500000	6.000000
5	Lower	Short	243.500000	5.611111
6	Standard	Long	243.500000	5.555556
7	Standard	Medium	243.500000	6.222222
8	Standard	Short	243.500000	6.777778

Table 21: Mean offset values of Tpos2 group

	sum_sq	df	F	PR(>F)
Intercept	5775.694429	1.0	20.717229	0.000011
FactorA	22.703704	2.0	0.040719	**0.960110**
FactorB	638.111111	2.0	1.144442	**0.321130**
FactorA: FactorB	196.024691	4.0	0.175784	**0.950573**
Subject	31.762037	1.0	0.113929	0.736180
Residual	42375.626852	152.0	NaN	NaN

Table 22: ANOVA Results of Tsr group

	FactorA	FactorB	Subject	Mean
0	Higher	Long	244.500000	19.944444
1	Higher	Medium	244.500000	26.333333
2	Higher	Short	244.500000	18.388889
3	Lower	Long	244.500000	20.444444
4	Lower	Medium	244.500000	25.833333
5	Lower	Short	244.500000	21.444444
6	Standard	Long	244.500000	21.500000
7	Standard	Medium	244.500000	23.388889
8	Standard	Short	244.500000	17.000000

Table 23: Mean offset values of Tsr group

The result shows that there is no significant difference between each task's time spent and vertical location, length, or the interaction between these two. Furthermore, the mean values of the three groups were examined. In the Tpos1 group, Lower-Long had the highest average time spent, while Higher-Medium had the lowest average time spent. Within the Tpos2 group, the average time spent did not vary considerably across the different conditions, with Higher-Short having the highest average time spent and Higher-Medium having the lowest average time spent. Regarding the Tsr group, the overall average time spent was significantly higher compared to the other two groups, with Higher-Medium exhibiting the highest average time spent and Standard-Short exhibiting the lowest average time spent.

Thus, it can be concluded that, in general, the Higher-Medium setting enables faster completion of the positioning task but requires more time for the search task.

C.1.4 Grab/Click Action

This section aims to analyze the participants' interaction behavior on the timeline during the experiment, specifically focusing on their grab and click actions performed on the slider. We examined the number of grabs and clicks recorded in the interaction logs, as well as the time distance (unit by second) that the slider moved on the timeline with each grab or click. Our objective is to determine whether the participants' selection of grab or click is associated with different conditions or the move distance.

C.1.4.1 Grab/Click Numbers First, we conducted a count of the number of grabs and clicks in various conditions. As shown in Figure 14, the number of grabs exceeded the number of clicks across all conditions. Additionally, we observed a decreasing trend in the number of grabs as the length of the timeline increased. This suggests that participants exhibited greater accuracy in their grab actions as the timeline became longer. However, we did not find a substantial correlation between the number of grabs and the vertical location.

On the other hand, the number of clicks is all less than the grab number in the same condition, with the highest number of clicks observed in the Standard-Medium condition and the lowest number of clicks in the Standard-Short condition.

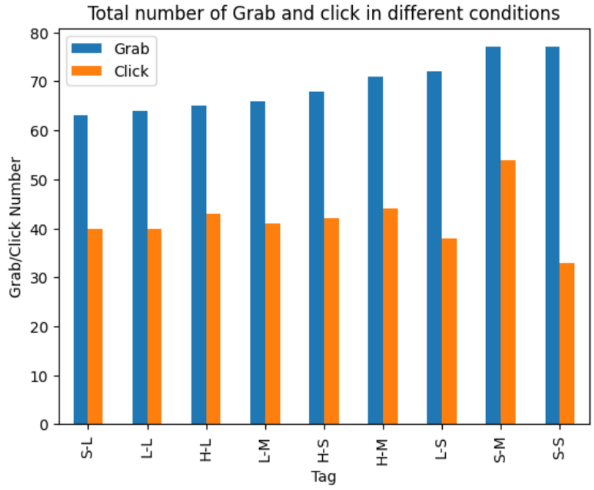


Figure 14: Total number of Grab and Click in different conditions

To investigate the relationship between the number of grabs and clicks in different conditions and task types, we segregated the data into three groups based on the task type. The results are presented in Figure 15 below.

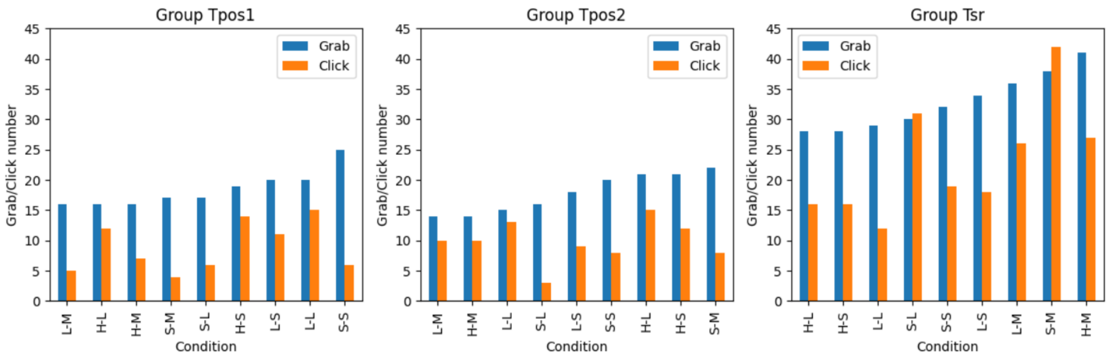


Figure 15: Total number of Grab and Click in different conditions grouped by different tasks

Upon analyzing Figure 15, it is evident that participants exhibited a higher number of grabs in the positioning tasks (Tpos1 and Tpos2). In the search task (Tsr), participants tended to grab more frequently in most conditions, except for the standard-long and standard-medium conditions. This suggests that participants have a preference for using grab actions when completing positioning tasks compared to the search task. Furthermore, the number of grabs and clicks in the Tpos1 and Tpos2 groups was relatively smaller compared to the Tsr group. This observation can be attributed to the fact that Tpos1 and Tpos2 had a clearly defined target time, whereas in Tsr, participants had to identify suitable scenes by exploring video content, resulting in more frequent grabs and clicks. Therefore, we can infer that when users watch 360-degree videos with the player, they will choose the "one-step" grab operation when they encounter a task with a clear goal; on the contrary, when they encounter a task without a clear goal and need to search for the video content, they will use the grab operation to complete the selection of the general range, and work with the more delicate click operation to make detailed adjustments.

C.1.4.2 The Relationship between Grab/Click and Duration In addition to analyzing the relationship between the number of grabs/clicks and different conditions and tasks, we also examined whether participants' grab/click actions were influenced by the distance the slider moved on the timeline, referred to as "duration". The results are presented below in Figures 16 and 19.

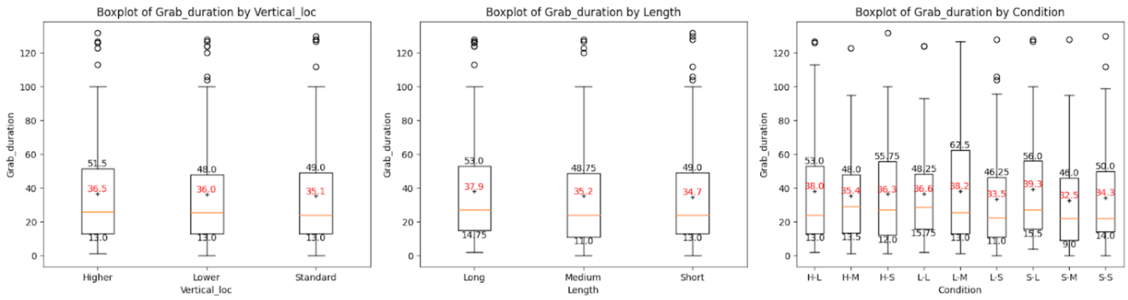


Figure 16: Boxplots for grab duration grouped by vertical location, length and conditions

From Figure 16, we observed that the average grab duration generally ranged between 30s to 40s, regardless of the condition. Regarding the vertical location, there was no substantial difference in duration. Surprisingly, the Standard setting had the lowest mean value, contrary to our initial assumptions. Similarly, for the length of the timeline, we found no substantial variation in duration. However, as the length of the timeline increased, the duration also increased. This phenomenon can be attributed to the decrease in time granularity as the timeline lengthens. When participants drag the same distance on the timeline (e.g., 1 cm), the time represented by this distance becomes shorter for a longer timeline. Consequently, participants experience less loss of detail and can execute more precise operations. As a result, participants tend to do more actions on shorter timelines to achieve higher accuracy, while they opt for a "one-step" approach on longer timelines. To validate our hypothesis, we conducted an analysis of the average number of grabs and correctness across different timeline lengths.

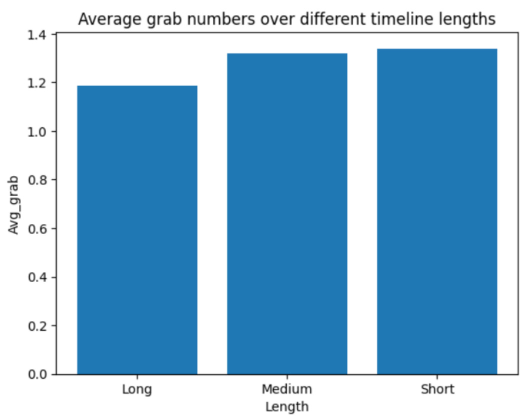


Figure 17: Average grab numbers over different timeline lengths

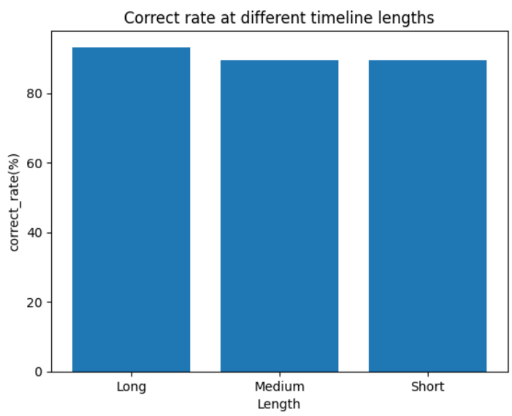


Figure 18: Correctness at different timeline lengths

The results showed that as the length of the timeline increased, the average number of grabs made by participants in each task decreased, and the correctness increased. This is consistent with our hypothesis. Therefore, we conclude that as the length of the timeline increases, people tend to perform fewer actions to achieve the "one-step" goal because they can see more details of the video when dragging the timeline. On a shorter timeline, people tend to repeat the task "a few more times" to ensure the accuracy of the operation because they cannot see many details.

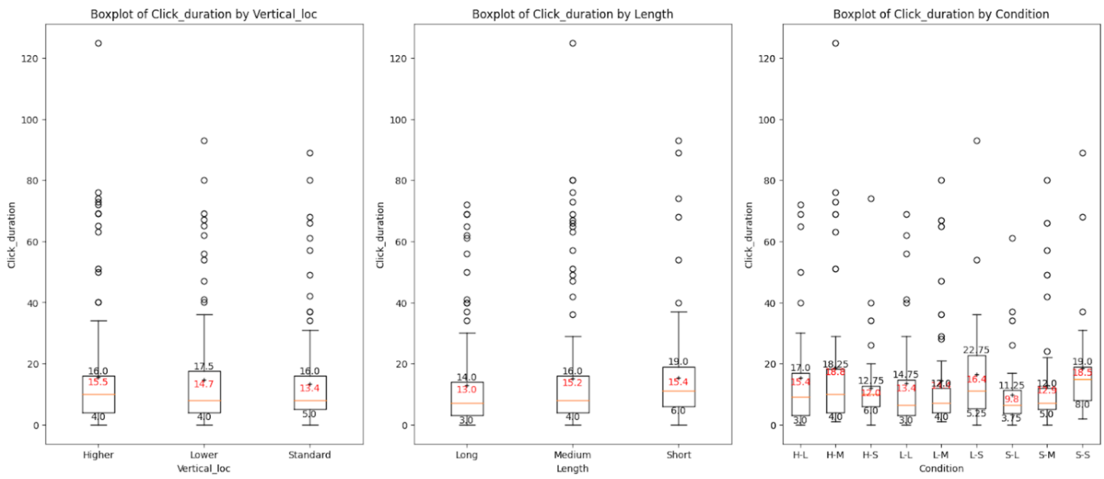


Figure 19: Boxplots for click duration grouped by vertical location, length, and conditions

Based on Figure 19, the average click duration is shortest for the Standard-Long condition and longest for the Higher-Medium condition. Regarding the vertical location, the Standard setting has the lowest average value. This finding contradicts our initial hypotheses. Furthermore, in terms of timeline length, we discovered that shorter timelines are associated with longer click duration. We speculate that this phenomenon arises due to the increased time granularity as the timeline becomes shorter. Therefore, when participants click on the timeline at the same distance, the shorter timeline results in a longer click duration.

Upon comparing the results between the grab and click groups based on different vertical locations, it is evident that the average duration for grab actions is generally longer than that for click actions. Specifically, when considering the experiment's video duration (2 minutes), participants exhibited a preference for click actions when the required jump duration was less than 15 seconds. Conversely, participants leaned towards grab actions for a duration between 15 and 50 seconds (see Figure 20).

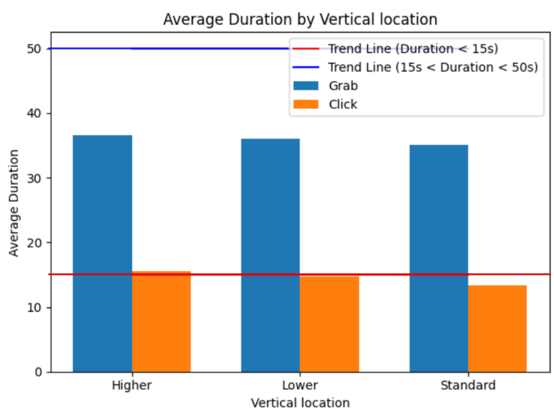


Figure 20: Average duration by different vertical locations (with trend line duration < 15s and 15s $<$ duration < 50s)

Furthermore, regardless of the grab or click action, the Standard vertical location consistently exhibited the lowest average duration. This unexpected finding contradicts our initial prediction that this position provided the most comfortable sensation for the neck and hands and would result in a longer duration. We hypothesized that participants might have been trying to achieve higher operational accuracy by employing a "few times" strategy that included fewer movements and longer duration.

However, the analysis revealed a different pattern. While grab actions do appear more frequently on Standard, click actions do not follow the same trend. In addition, the accuracy of click actions on Standard is lower compared to other vertical positions. Unfortunately, we cannot fully explain why this difference appears in terms of quantitative data, and further qualitative data analysis may provide insights.

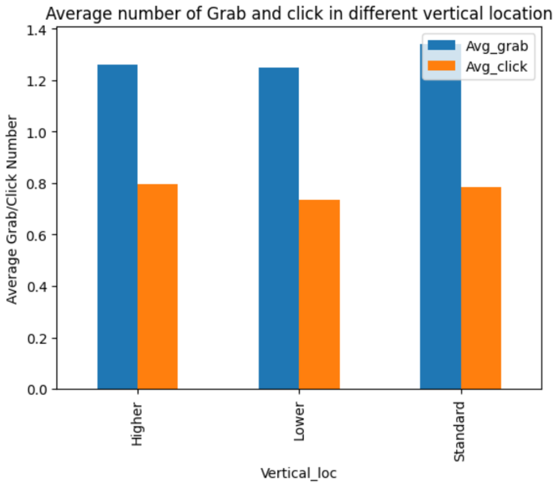


Figure 21: Average number of Grab and Click in different vertical locations

C.2 Qualitative Results

This section provides an in-depth analysis of the qualitative data obtained from the experiment, which encompasses questionnaire responses and participant interviews. The analysis revolves around various aspects, including participant information, cybersickness and discomfort levels, perceived difficulty of different conditions, preferences for vertical location and length settings, and overall feedback. By delving into these qualitative findings, our objective is to gain a profound understanding of participants' experiences and perceptions, thereby offering a comprehensive assessment of their engagement and the effectiveness of the experimental conditions.

C.2.1 Participant information

This section provides statistical graphs depicting the participants' demographics, including gender and age, as well as their relevant experiences with VR and 360-degree videos.

C.2.1.1 Participants' Demographics Figure 22 presents the gender distribution of the participants, revealing an equal representation of male and female participants, with 9 individuals in each group, accounting for 50% of the total. Furthermore, Figure 23 illustrates the age distribution, indicating that 12 participants were aged between 18 and 24 years, making up 66.67% of the sample, while 6 participants fell within the age range of 25-34 years, representing 33.33% of the participants.

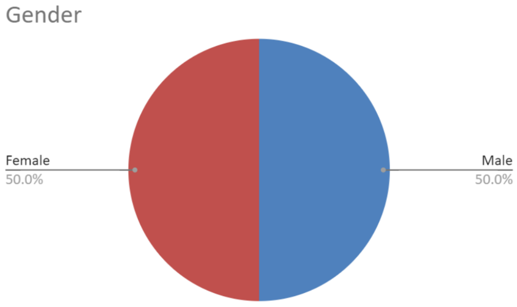


Figure 22: Participants' gender distribution

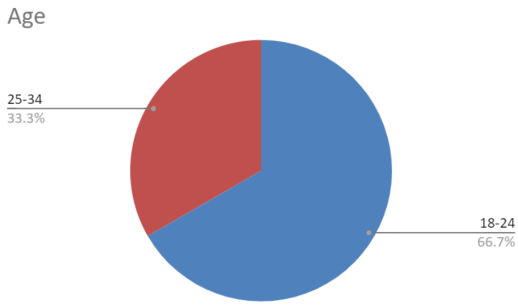


Figure 23: Participants' age distribution

C.2.1.2 Participants' Relevant Experience Levels Figure 24 provides insights into the participants' experience levels with VR. The majority of participants (11 individuals) reported using VR a few times (61.1%), followed by 3 participants who indicated occasional use (16.7%), and 2 participants who had no prior experience with VR (11.1%). Additionally, Figure 25 depicts the participants' experience with 360-degree videos and related platforms. Out of the total participants, 13 individuals reported having experienced 360-degree videos (72.2%), with 8 of them using a Head-Mounted Display (HMD) device (61.5%) and 3 of them trying it on a Mobile device (23.1%). On the other hand, 5 participants stated that they had no experience with 360-degree videos (27.8%).

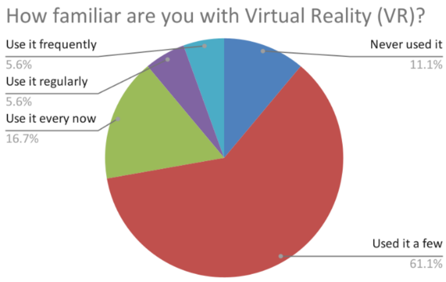


Figure 24: How participants are familiar with virtual reality

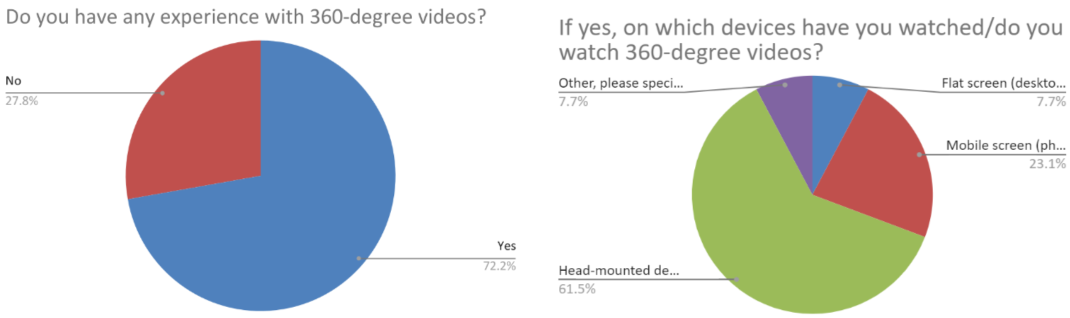


Figure 25: How participants are familiar with 360-degree videos and the devices they have used

Overall, the participant group consisted mainly of young individuals with an introductory experience in VR and 360-degree videos. This aligns with our target user profile for the given use case.

C.2.2 Cybersickness and discomfort level

This section focuses on assessing the participants' overall experience of cybersickness and discomfort during the experiment. Figure 26 provides insights into the participants' cybersickness levels, indicating that most participants reported no significant cybersickness. Furthermore, Figure 27 illustrates the participants' discomfort levels, revealing that most participants experienced a low level of discomfort, with only a few reporting a moderate level. Overall, the participants demonstrated relatively lower levels of cybersickness and discomfort throughout the experiment.

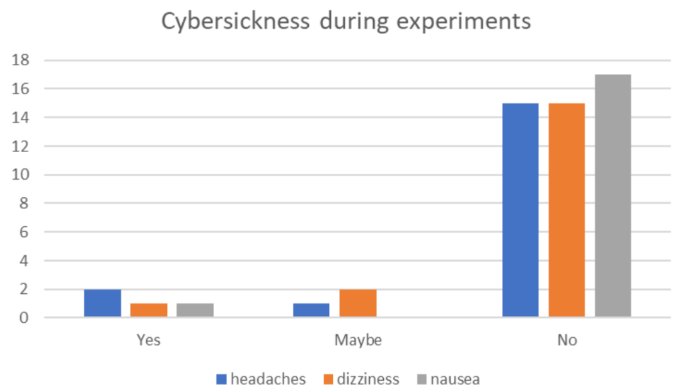


Figure 26: Cybersickness experience during the experiment

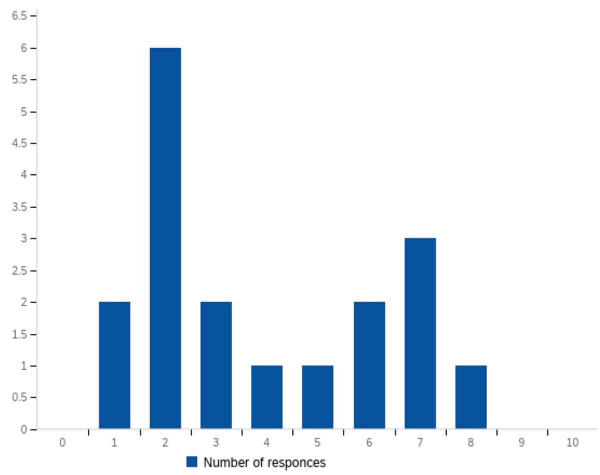


Figure 27: Discomfort level during the experiment

C.2.3 Difficulty level of different conditions

To assess the participants' perceived difficulty across different conditions, we categorized the difficulty level scores provided by all participants on a scale of 0 to 10, where 0 represents "super easy" and 10 represents "super hard" for each condition. We referred to these scores as difficulty levels and calculated the mean difficulty level for each condition. The results are presented in Figure 28, indicating that the Higher-Medium condition had the lowest mean difficulty level, while the Lower-Long condition had the highest. Regarding the vertical location, the average difficulty level for the Lower condition was higher than the Higher and Standard conditions. Similarly, regarding the length, the average difficulty level for the Long condition was higher than that of the Short and Medium conditions.

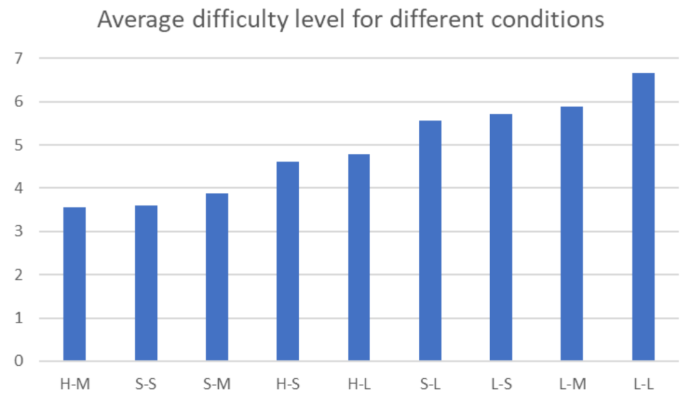


Figure 28: Average difficulty level for different conditions

To further validate our findings, we conducted an ANOVA on the difficulty levels of the 18 participants across the 9 conditions. The results revealed a significant difference among the 9 groups of difficulty levels (p=0.0008). To determine the specific groups that exhibited differential levels, we performed post hoc multiple comparisons using Tukey's Honestly Significant Difference test. These results are illustrated in Figure 29, demonstrating significant distinctions between the Lower-Long group and the Higher-Medium, Standard-Medium, and Standard-Short groups. Hence, we can conclude that participants perceived the Higher-Medium condition as the least difficult while considering the Lower-Long condition as the most challenging, and there was a notable disparity between the ratings of these two conditions.

C.2.4 Preference for length and vertical location

This section examines participants' preferences for the length of the timeline and the vertical location, as well as the discomfort experienced due to these factors.

C.2.4.1 Length Preference and Discomfort To gain insights into participants' preferences regarding different timeline lengths, a multiple-choice question was included in the questionnaire. The results, shown in Figure 30, reveal that out of the 23 responses received, 65.22% of participants expressed a preference for Medium length timelines. This was followed by 21.74% of participants indicating a preference for Long timelines, and 13.04% of respondents favoring Short timelines.

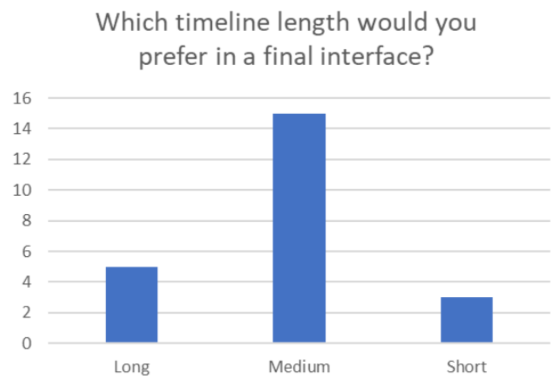


Figure 30: The preference for timeline length

Furthermore, an analysis was conducted on participants who experienced discomfort during the experiment due to the timeline length. The results showed that 13 participants (72.22%) reported discomfort. The main issues raised by participants were:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
HL	HM	-2.5621	0.1021	-5.3626	0.2384	False
HL	HS	-1.5065	0.75	-4.307	1.294	False
HL	LL	0.549	0.9995	-2.2515	3.3495	False
HL	LM	-0.2288	1.0	-3.0293	2.5717	False
HL	LS	-0.3954	1.0	-3.1959	2.4051	False
HL	SL	-0.5621	0.9994	-3.3626	2.2384	False
HL	SM	-2.2288	0.2382	-5.0293	0.5717	False
HL	SS	-2.5065	0.1191	-5.307	0.294	False
HM	HS	1.0556	0.9547	-1.7046	3.8158	False
HM	LL	3.1111	0.0148	0.3509	5.8713	True
HM	LM	2.3333	0.1713	-0.4269	5.0935	False
HM	LS	2.1667	0.2549	-0.5935	4.9269	False
HM	SL	2.0	0.3607	-0.7602	4.7602	False
HM	SM	0.3333	1.0	-2.4269	3.0935	False
HM	SS	0.0556	1.0	-2.7046	2.8158	False
HS	LL	2.0556	0.3231	-0.7046	4.8158	False
HS	LM	1.2778	0.8733	-1.4824	4.038	False
HS	LS	1.1111	0.9393	-1.6491	3.8713	False
HS	SL	0.9444	0.9767	-1.8158	3.7046	False
HS	SM	-0.7222	0.996	-3.4824	2.038	False
HS	SS	-1.0	0.967	-3.7602	1.7602	False
LL	LM	-0.7778	0.9934	-3.538	1.9824	False
LL	LS	-0.9444	0.9767	-3.7046	1.8158	False
LL	SL	-1.1111	0.9393	-3.8713	1.6491	False
LL	SM	-2.7778	0.0472	-5.538	-0.0176	True
LL	SS	-3.0556	0.0182	-5.8158	-0.2954	True
LM	LS	-0.1667	1.0	-2.9269	2.5935	False
LM	SL	-0.3333	1.0	-3.0935	2.4269	False
LM	SM	-2.0	0.3607	-4.7602	0.7602	False
LM	SS	-2.2778	0.1966	-5.038	0.4824	False
LS	SL	-0.1667	1.0	-2.9269	2.5935	False
LS	SM	-1.8333	0.4839	-4.5935	0.9269	False
LS	SS	-2.1111	0.2878	-4.8713	0.6491	False
SL	SM	-1.6667	0.6147	-4.4269	1.0935	False
SL	SS	-1.9444	0.4002	-4.7046	0.8158	False
SM	SS	-0.2778	1.0	-3.038	2.4824	False

Figure 29: The results of Tukey's Honestly Significant Difference test

- Dissatisfied with the Short timeline's accuracy because it required multiple adjustments to accurately point to the target location.
- The Long timeline is excessively extended, causing difficulties in locating and grasping the slider due to the entire control panel rotating with their head movements. This made it challenging to keep their head still and rely on eye movement alone.
- Navigating video content and performing search tasks with the Long timeline proved inconvenient. Some participants preferred grabbing the slider and turning their heads for navigation, but the simultaneous movement of the control panel with their head made it difficult to pinpoint specific timeslots, resulting in a cumbersome process.

Did you experience any discomfort because of the length of the timeline in some conditions?	The number of responses
No	5
Yes	13

Table 24: Discomfort due to the timeline length

Settings	Discomfort reasons
Short	Low accuracy makes it difficult to complete tasks
Long	Not very convenient overall
	The length is equal to the FOV, and it is difficult to catch the slider when it tends to be at both ends of the timeline
	Lower+Long makes things even worse

Table 25: Other discomfort reasons due to the timeline length

C.2.4.2 Vertical Location Preference and Discomfort Similar to the previous section, a multiple-choice question was included in the questionnaire to investigate participants' preferences regarding the vertical location. The results, presented in Figure 31, collected 24 responses. Among these, 58.33% of participants expressed a preference for the Standard vertical location, while 33.33% indicated a preference for a Higher vertical location, and only 8.33% of responses indicated a preference for a Lower vertical location.

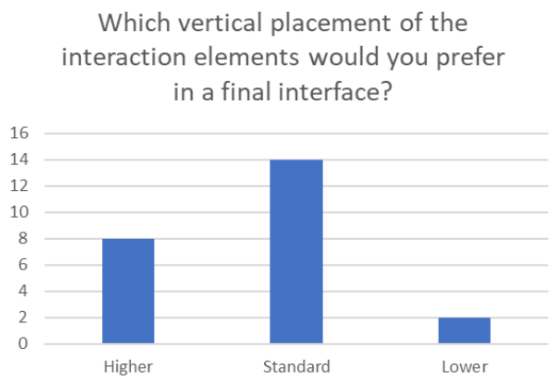


Figure 31: The preference for vertical location

Table 26 shows participants' feedback on the discomfort caused by the vertical location. 11 participants (61.1%) reported feeling discomfort. The main reasons mentioned were:

- The Lower vertical location forced them to lower their heads while using the system, leading to neck pain due to the weight of the HMD device.
- The Lower vertical location was too low, and spotting the timeline was challenging.

Did you experience any discomfort because of the vertical placement of the interaction elements?	The number of responses
No	7
Yes	11

Table 26: Discomfort due to the vertical placement

Settings	Discomfort reasons
Lower	Difficult to find the timeline
	Too low, it takes time to get used to the location
Higher	The thumbnail is blocking the screen
	Need to lift arms to operate, friendly to the neck, but not friendly to the arm
	Neck pain

Table 27: Other discomfort reasons due to the vertical location

Overall, our findings indicate that the majority of participants preferred the Medium length timelines and the Standard vertical locations. However, a subset of participants showed a preference for Long timelines or Higher vertical locations. Furthermore, our analysis of participants' reported discomfort revealed that the primary sources of discomfort were associated with extreme timeline lengths or vertical locations, affecting participants' ease of operation.

C.3 Overall feedback

In addition to the feedback mentioned above, we conducted brief interviews with the participants to gather their overall comments, suggestions, and feedback. This section summarizes the insights obtained from these interviews.

C.2.5.1 Discomfort The participants commonly expressed discomfort associated with the weight of the HMD devices throughout the experiments, which was consistently reported across different experimental conditions to varying degrees. They believed that this discomfort could accumulate over time. As a potential solution, several participants recommended the use of lighter HMD devices to alleviate this discomfort.

C.2.5.2 Length and Vertical Location Preference Participants provided interesting feedback regarding their preferences and comfort levels concerning the timeline length and vertical location. Notably, some participants found one particular setting to be "significantly more uncomfortable/dislikable", while perceiving little difference between the other two settings. The feedback includes:

- The short timeline length caused considerable discomfort, but participants did not perceive a significant difference between the medium and long lengths.
- The long timeline length was deemed excessively extended and unfavorable. Conversely, participants found the short length to be better than expected, requiring minimal action to complete tasks. The medium length was highly regarded as a preferred choice.
- Length and height had minimal impact on positioning tasks, but longer timelines were considered worse in search tasks.
- While longer timelines allowed for more precise operations, medium-length timelines provided a better understanding of the overall content.
- The lower vertical location caused discomfort in the neck, while the higher location was beneficial for the neck but required lifting the hands to operate. The medium location was seen as a trade-off between the two.

Consistent with the previous preference survey, participants were able to clearly indicate their most preferred setting among the three options. However, they did not perceive a distinct difference between the other two settings. Their preference for these settings depended on personal usage habits or level of experience.

C.2.5.3 Other Feedback Participants also offered suggestions to improve the experiment, highlighting additional factors that could impact the results:

- The textual content of the tasks was lengthy and not easily memorized. Participants sometimes needed to confirm details with the experimenter, resulting in increased time consumption.
- Some participants found the player's control panel too close and suggested placing it farther away for easier operation.

D References

[1] Mike Alger. “Visual design methods for virtual reality”. In: *Ravensbourne. http://aperturesciencellc.com/vr/VisualDesignMethodsforVR_MikeAlger.pdf* (2015).

[2] Samuel Alves, Arthur Callado, and Paulyne Jucá. “Evaluation of graphical user interfaces guidelines for virtual reality games”. In: *2020 19th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*. IEEE. 2020, pp. 71–79.

[3] BrightFarms. *BrightFarms Virtual Reality Greenhouse Tour*. Vimeo. 2018. URL: https://vimeo.com/266752213.

[4] Andy Brown et al. “Exploring subtitle behaviour for 360 video”. In: *White Paper WHP* 330 (2018).

[5] Andy Brown et al. “Subtitles in 360-degree Video”. In: *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. 2017, pp. 3–8.

[6] Peter Chu. “Vr design: Transitioning from a 2d to 3d design paradigm”. In: *San Francisco: Samsung Developer Connection* (2014).

[7] Mimmi Cromsjö and Martin Christensson. *Design of video player and user interfaces for branched 360 degree video*. 2019.

[8] *Design guidelines - Google VR - google developers*. URL: https://developers.google.com/vr/develop/experimental/6dof-design-guidelines.

[9] *Designing for hand tracking¶*. URL: https://docs.ultraleap.com/xr-guidelines/designing-hand-tracking.html.

[10] Rohann Dorabjee et al. “Back to the future: identifying interface trends from the past, present and future in immersive applications”. In: *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. 2015, pp. 540–544.

[11] Maxim Eylenbosch. “Defining Differences in Movement and View in Tactical First-Person Shooter Games between Virtual Reality and Personal Computer”. PhD thesis. June 2022. DOI: 10.13140/RG.2.2.18412.03205.

[12] Erik Fagerholt and Magnus Lorentzon. “Beyond the HUD-User Interfaces for Increased Player Immersion in FPS Games”. In: (2009).

[13] *GizmoVR Video Player*. URL: https://store.steampowered.com/app/768100/GizmoVR_Video_Player/.

[14] Colin Groth et al. “Mitigation of cybersickness in immersive 360 videos”. In: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2021, pp. 169–177.

[15] Beatriz Gutiérrez-Caneda, Sara Pérez-Seijo, and Xosé López-Garcıa. “Analysing VR and 360-degree video apps and sections. A case study of seven European news media outlets”. In: *Revista Latina de Comunicación Social* 75 (2020), pp. 149–166.

[16] Chris J Hughes and Mario Montagud. “Accessibility in 360° video players”. In: *Multimedia tools and applications* 80.20 (2021), pp. 30993–31020.

[17] Cale Hunt. *Lone Echo's beautiful graphics are a welcome home for its compelling story*. July 2017. URL: https://www.vrheads.com/lone-echo.

[18] Prakash R Kolhe, GP Khetri, and NK Deshmukh. “Study of standard assumptions of graphical user interface (GUI) based on usability, adaptability and security factors”. In: *International Journal of Advanced Research in Computer Science and Software Engineering* 2.12 (2012), pp. 1–7.

[19] Costanza Laudisa. “Timeline designs for 360-degree videos in VR”. MA thesis. 2023.

[20] Joseph J LaViola Jr. “A discussion of cybersickness in virtual environments”. In: *ACM Sigchi Bulletin* 32.1 (2000), pp. 47–56.

[21] Joseph J LaViola Jr et al. *3D user interfaces: theory and practice.* Addison-Wesley Professional, 2017.

[22] Sigbjørn Litleskare. "The relationship between postural stability and cybersickness: It's complicated–An experimental trial assessing practical implications of cybersickness etiology". In: *Physiology & Behavior* 236 (2021), p. 113422.

[23] ESCAPE NOW. *Escape Now to Spain's World Heritage Sites — Trailer — VR 360 Video.* Youtube. 2021. URL: https://youtu.be/1mCTAhqEL_0.

[24] ESCAPE NOW. *Escape Now to The United Arab Emirates — Trailer — VR 360° Trailer.* Youtube. 2022. URL: https://youtu.be/7jW5iLTlITw.

[25] ESCAPE NOW. *Exploring the Wonders of Oman: A Journey Through a Land of Deserts, Oases, and Culture.* Youtube. 2022. URL: https://youtu.be/nmjQ3sWMIN0.

[26] Rennan Raffaele, Breno Carvalho, and Frutuoso Silva. "Evaluation of immersive user interfaces in virtual reality first person games". In: *24º Encontro Português de Computação Gráfica e Interação* (2017), pp. 123–126.

[27] Sylvia Rothe, Kim Tran, and Heinrich Hußmann. "Dynamic subtitles in cinematic virtual reality". In: *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video.* 2018, pp. 209–214.

[28] Saeed Safikhani, Michael Holly, and Johanna Pirker. "Work-in-Progress—Conceptual Framework for User Interface in Virtual Reality". In: *2020 6th International Conference of the Immersive Learning Research Network (iLRN).* IEEE. 2020, pp. 332–335.

[29] Rustam Shadiev, Liuxin Yang, and Yueh Min Huang. "A review of research on 360-degree video and its applications to education". In: *Journal of Research on Technology in Education* 54.5 (2022), pp. 784–799.

[30] Gray Line Sightseeing. $Grayline_CostaRica_SanJoséCityTour_360$. Vimeo. 2019. URL: https://vimeo.com/367808473.

[31] *Skybox VR Video Player.* URL: https://skybox.xyz/.

[32] Anthony Steed et al. "Directions for 3D User Interface Research from Consumer VR Games". In: *arXiv preprint arXiv:2106.12633* (2021).

[33] *Tilt Brush Instruction Guide.* URL: https://vrautism.wixsite.com/edtc/instruction-for-construction-use.

[34] *Top 40 best custom beat saber songs (all free mods).* URL: https://www.fandomspot.com/beat-saber-best-custom-songs/.

[35] Alissa Vermast and Wolfgang Hürst. "Introducing 3D Thumbnails to Access 360-Degree Videos in Virtual Reality". In: *IEEE Transactions on Visualization and Computer Graphics* 29.5 (2023), pp. 2547–2556.

[36] Maria Grazia Violante, Enrico Vezzetti, and Pietro Piazzolla. "Interactive virtual technologies in engineering education: Why not 360° videos?" In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* 13.2 (2019), pp. 729–742.

[37] *Vision - Oculus developer.* URL: https://developer.oculus.com/resources/bp-vision/.

[38] AirPano VR. *360 video, Kitties in VR. Cute inhabitants of cat cafe in Moscow. 4K video.* Youtube. 2018. URL: https://youtu.be/eWONfcEL2Dk.

[39] AirPano VR. *Manpupuner rock formations, Russia. 16K 360 aerial video.* Youtube. 2023. URL: https://youtu.be/bs-LgFR64w0.

[40] Vravel. $GRAYLINE_MEXICO_360PUERTOVALLARTACITYTOUR|FULL360$. Vimeo. 2019. URL: https://vimeo.com/349298488.

[41] WURtube. *360 Campus Tour — Wageningen University Research.* Youtube. 2017. URL: https://youtu.be/-dDOjL-UjzQ.

[42] *YouTube.* URL: www.youtube.com.