Precision Scrubbing: Improving Control Over Granularity Using 3D Interfaces for 360° VR Video Players

Master thesis

Jorian Berkhout (6240178)

Supervisor: Wolfgang Hürst
Second supervisor: Remco Veltkamp

Department of Information and Computing Sciences
Utrecht University

January 2, 2023

Precision Scrubbing: Improving Control Over Granularity Using 3D Interfaces for 360° VR Video Players

Jorian Berkhout
j.l.berkhout@students.uu.nl
Utrecht University
The Netherlands

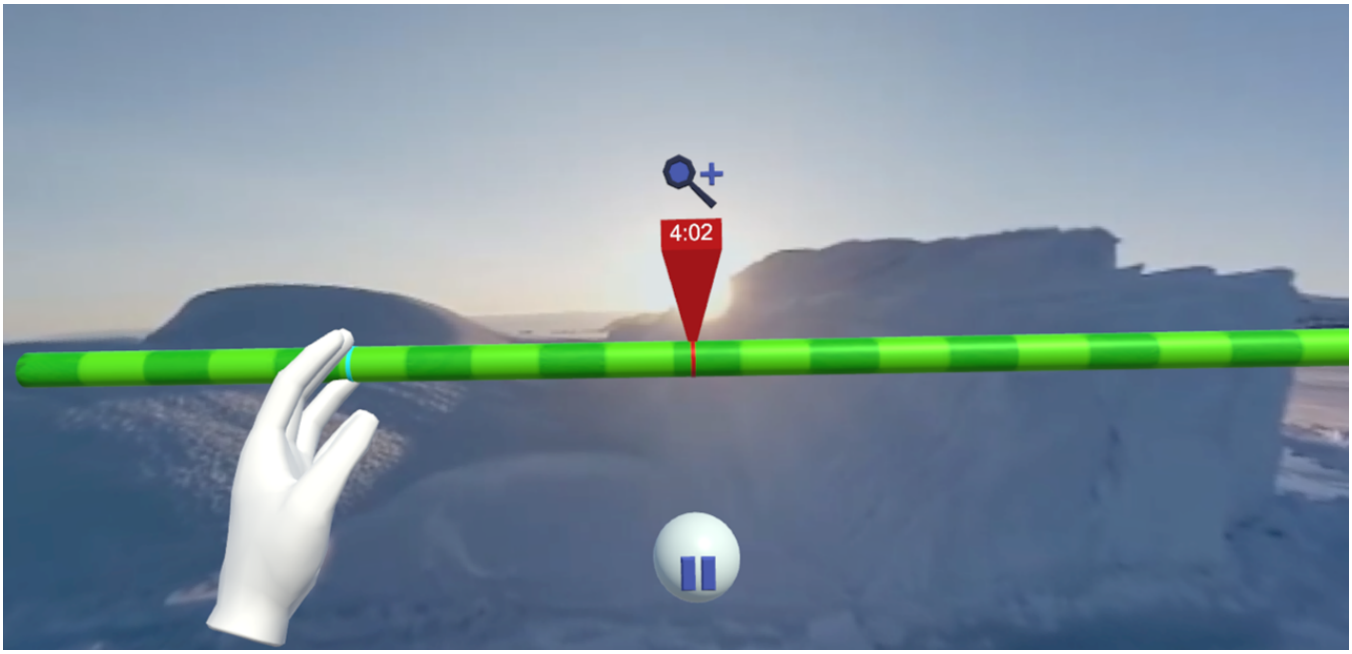

Figure 1: The stretchable timeline while it is being scaled in VR

ABSTRACT

Current 360° video players feature simple 2D interfaces that resemble their desktop and mobile interfaces. In this research two novel 3D interfaces designs are presented and compared to a state-of-the-art baseline interface, in terms of accuracy, efficiency and usability. A within-subjects study has been conducted with in person user testing. No significant difference in any of the metrics was found between the best 3D interface and the state-of-the-art interface. However, participants rated the 3D interface as significantly more fun. It is recommended that the concept of 3D interfaces is explored further by means of a developmental study, as the similar score in metrics seems to indicate 3D interfaces can compete with their 2D counterparts.

Supervisor: Dr. Wolfgang Hürst

Second supervisor: Dr. Remco Veltkamp

January 2nd, 2023, Utrecht University

1 INTRODUCTION

Since the recording of the first films in the 19th century, many different devices have facilitated video playback. Apart from advances in video quality, new devices constantly refined the viewer's control over playback. Where VHS (video home systems) were limited by a physical strip of tape, DVD players could skip to specific moments using scene selection without having to fast forward the entire video. Digital video players commonly use a timeline to navigate a video, allowing users to jump to any moment at will. Touchscreens augmented timeline-based interfaces, facilitating even more direct and precise manipulation of a video timeline.

Parallel with advancements in playback devices, video formats have also evolved over time. From the introduction of colour to constantly increasing image quality. A relatively new format is that of the 360° video, where recording and playback is omnidirectional. Virtual reality (VR) is exceptionally well-suited for viewing these

videos, as it allows the viewer to look around simply by turning their head. But how does a viewer control video playback?

The most common VR 360° video player interfaces used in VR applications resemble their desktop counterpart: a 2D interface with buttons and a slider [19]. Although most users will feel familiar with such an interface, porting it to VR results in a number of issues.

The first issue is the method of interaction, ray selection. Best compared with a laser pointer in real life, this technique draws a line from the user's hand to the point of interaction on a surface in VR. In this case, it mimics the cursor used on desktop devices. Although great for most VR interactions, it is ill-suited for precision tasks because it effectively functions as a lever. A small (accidental) movement of the hand results in a significantly larger movement on the surface pointed at. This is quite undesirable especially when scrubbing through a video, where a subtle movement may result in a jump of several seconds or even minutes.

The second issue is the interface size and placement. When viewing (360°) videos on a desktop or mobile device there is a fixed, rectangular viewport. The edges of this viewport are natural boundaries for the timeline. In a VR setting, there is no edge to function as natural bounds, as the content is presented in a sphere all around the user.

In addition to no clear interface boundaries, the length of videos varies. In a desktop situation it makes sense to compress the timeline to fit inside the viewport. If a video is very short, high levels of precision can be achieved easily. However, the longer the video, the more 'time' is compacted to the same timeline length. Navigating to the start of the second minute of a video is therefore much easier when viewing a three-minute video compared to a movie lasting two hours. It is not uncommon that it is impossible altogether to reach a specific point through scrubbing only. The different levels of granularity that apply to the same timeline depending on video length are beyond the control of the user. The variable video length can especially result in problems in combination with the imprecision of the interaction method mentioned previously. On a longer video, one might accidentally skip whole minutes when attempting to jump a couple of seconds forward.

By leveraging the freedom virtual space offers, a novel interface design could provide control over the level of granularity. By carefully considering size, placement, shape and interaction method of the interface it could enhance the VR 360° video viewing experience.

The research goal of this research is to determine the benefit of using the 3D space available in VR to construct an interface which provides the user with more control over granularity. To do so this research introduces two novel VR 360° video player interfaces, specifically designed for a 3D VR space. Both interfaces attempt to use the 3D space to provide the user with more control over granularity. The potential of the novel designs is tested by answering the following research questions:

(1) How accurate are users when using a 3D interface compared to a state of the art implementation?
(2) How efficient are users when using a 3D interface compared to a state of the art implementation?
(3) How do users perceive the usability of a 3D interface compared to a state of the art implementation?

First of all, the hypothesis that corresponds to research question one is that by utilising the full 3D space available in VR users can be given greater control over the level of granularity of their scrubbing (for example by introducing additional interactions such as modifying the size of the timeline). This is expected to raise accuracy beyond that of the state of the art interface.

Secondly, research question ensues to the hypothesis that by directly relating timeline interactions to positions in 3D space, efficiency will be higher when using a 3D interface compared to its 2D counterpart. In addition, easier access to higher levels of precision is also expected to increase efficiency.

Thirdly, the hypothesis related to the third research question is that users will consider a 3D interface to have better usability, as such an interface is more better suitable to use in a VR environment compared to a virtual 2D screen.

In order to delve deeper into the research questions and to test the hypothesis, this paper starts off with a review of relevant literature (section 2), followed by the presentation of the novel interface designs and implementation (section 3). After the methodology section describing the evaluation strategy (section 4) the results are presented (section 5). Finally, the conclusion is presented (section 6), followed by the discussion (section 7) including recommendations for future work (section 7.3).

2 RELATED WORK

Currently used VR video players such as *Oculus Video* [19] feature familiar 2D interfaces, looking very much like desktop or television interfaces. Most exploration regarding VR video viewing focuses on social aspects such as viewing videos with friends [12, 28]. Though there is little research specifically on 3D interfaces, there are a number of relevant studies outlined in this section. The topics discussed are measuring usability, cybersickness, interaction, and timelines.

2.1 Measuring usability

Usability is most commonly measured using the well-established Systems Usability Scale (SUS-test) [4]. This scale can be applied to virtually any system, including VR applications, and provides a general indication of the usability of any given system. The biggest advantage of the SUS-test is the simplicity and length of the questionnaire, so participants can quickly evaluate multiple systems. Important when attempting to gauge usability is to limit interaction with the participant as much as possible [3]. Directions from outside the system being tested (e.g. comments by the researcher) can potentially break immersion for the participant. Especially in VR it can be weird to speak with someone in the same room without being able to see them. Therefore interaction should be limited to necessary directions.

2.2 Cybersickness

Another potential immersion breaking event is the occurrence of cybersickness. Used as an umbrella term of a plethora of symptoms, it is experienced by the majority of users at some point when using a VR system [25, 26]. The most common symptoms include dizziness, nausea and headaches. Cybersickness can occur in all kinds of situations, for example when playing video games or when

watching a movie, but is most common in VR settings. The topic has been studied extensively [26, 27], as well as ways to reduce its effect [11, 25].

Rebenitsch [25] published an article which contains several design practices for reducing the risk of cybersickness occurring. For example, regular interaction with a VR environment is known to reduce the effects and susceptibility. A direct manipulation 3D interface would facilitate such interactions. Another important factor is the time spent in VR in one session. The longer the session lasts, the higher the chance of symptoms occurring. It is therefore important to keep evaluation sessions as brief, and facilitating frequent breaks to minimize the risk of cybersickness occurring during the evaluation, possibly influencing results. Another aspect that can help reduce cybersickness is to have users take a seat whenever feasible when using VR. Sitting mainly helps to reduce chances of dizziness, lowering the chance of discomfort.

Another common cause of cybersickness is the quickly changing of scenes [16]. In VR the scene is all around the user, changing the entire environment with the snap of a finger can be quite discombobulating. In a 360° video player that allows the manipulation of time this is quite an often occurring event. Therefore care must be taken to implement a system in such a way it does not trigger any discomfort.

2.3 Interaction

Moving on to the way users can interact with VR systems. Jacob et al [15] state that novel interfaces can benefit from inspiration from the real-world. For example, many smartphone interfaces simulate inertia, which helps users better understand what is happening. Mimicking real world interactions in interfaces makes it easier for users to operate them. If the interaction is more natural, it requires less time to learn how to operate it. Jacob et al identified a number of themes that can aid designing these reality-based interactions. These themes are *naive physics, body awareness and skills, environment awareness and skills* and *social awareness and skills*. The first three themes are relevant for 360° video players and should be considered when designing such a system.

Every human has a basic understanding of the physical world, including gravity and the persistence of objects. This is referred to as naive physics. In VR this can be taken to the next level by having an interface that for all intents and purposes obeys the laws of physics as we know them from the real world. In this case it can be as simple as when the user grabs something and moves their hand, the attached object will move as well. As a result, the interface is expected to be much more natural [15].

Body awareness includes proprioception (being aware of the relative position of limbs), reach and movement coordination. Currently, most VR interfaces are designed to use ray interaction instead of directly using controllers to interact. When using a method such as ray select, some of the benefits of proprioception might be lost, as opposed a direct interaction interface, where a user is expected to move their arms around more.

Environment awareness is related to the perceived physical presence in an environment. Objects observed by humans function as landmarks and orientation points. For example, the horizon tells us something about the angle we are facing, while shadow helps determine the distance of objects related to ourselves and to each other. Knowing exactly how and where our body is compared to the scene helps interacting with items in that scene [15]. Therefore, using 3D objects as an interface should in theory increase environment awareness in users.

Petry et al [22] proposed a clear distinction in interaction between time navigation and spatial navigation within a 360° video player. Gaze direction is used solely for panning the video, while a simple set of gestures is used for temporal navigation. Different types of navigation should not conflict in terms of controls. For spatial navigation in 360°, using the headset to track head rotation is the most natural. Therefore it makes sense to avoid gaze interaction for the video player controls, instead focusing on direct interactions.

2.4 Timelines

The effects of the shape of a timeline visualisation were studied by Di Bartolomeo et al [10]. They compared task execution time and accuracy on linear, circular and spiral timelines. The results showed that participants were quicker at performing the tasks on linear timelines. No significant differences in accuracy were found. According to the researchers, the user's familiarity with linear timelines was a contributing factor to its superiority. Although circular timelines were outclassed in terms of performance and readability, they suggested to use circular or spiral visualisations when it makes sense in that specific context. As this research focuses on navigating time, often represented by a clock, it is interesting to see if a circular visualisation could yield benefits.

A research by Higuch et al introduced the concept of an elastic timeline [13]. Elastic means that certain parts of the timeline, those that contain interesting events, are stretched. This allows the user to more easily locate and navigate to interesting highlights, but also to scrub within an interesting moment with increased precision. A visual analysis of the (often lengthy) first-person video was required to create a set of segments of potentially interesting moments. The video would then play at an increased playback rate until such an interesting moment was encountered (for example a conversation with another human). Then, the playback rate would drop to normal speed, until the end of the fragment. The idea of increasing and reducing granularity on certain conditions certainly has merit.

The large data requirement of streaming 360° videos, especially in VR, is a major challenge [7, 34]. In 2D video streaming interfaces, bookmarks displayed on timelines allow pre-rendering of parts of the video the user might skip to during video streaming [6]. In addition, bookmarks have been proven to reduce search times within videos [36]. A 3D timeline could incorporate bookmarks more easily than its 2D counterpart, as there is more room for interactions without conflicting with the video controls.

3 PROPOSED TIMELINE DESIGNS AND IMPLEMENTATION

In order to test the hypotheses defined in the introduction (see section 1), this research proposes two novel designs (the clock interface and the stretchable interface), as well as an implementation based on existing interfaces (the state-of-the-art interface). Both

novel timelines are designed with a different solution to the granularity problem of the state-of-the-art interface. Where the clock interface focuses on fixing the granularity to different known levels, the stretchable interface provides full control over granularity to the user. The state-of-the-art timeline functions as a baseline to compare the other interfaces to. A brief description of reasoning and components for each interface is given below.

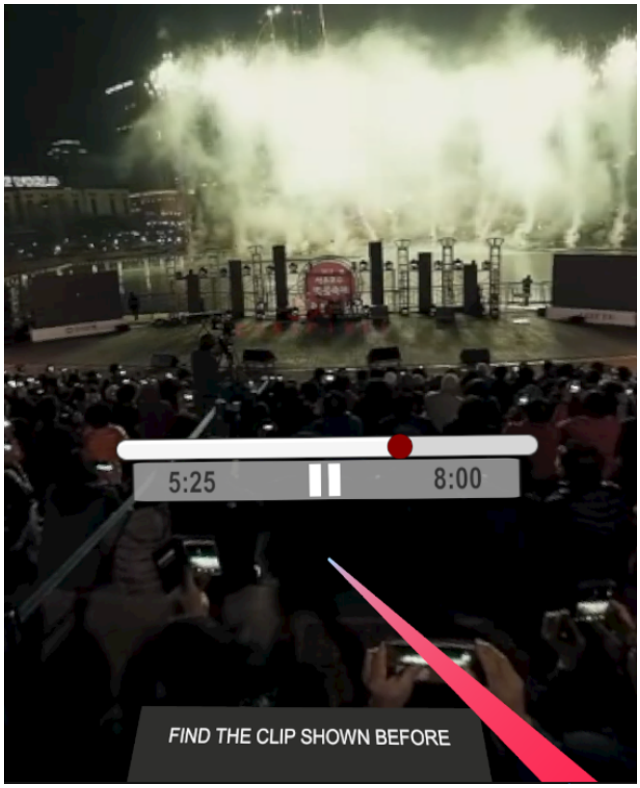

Figure 2: The state-of-the-art interface as seen in VR

3.1 State of the art interface

The flat timeline is the most common interface for VR video players, and is therefore referred to as the state-of-the-art interface. This interface will be used as a baseline to compare the novel interfaces to. It contains most features expected from a VR video player interface.

It is very similar to desktop video players, it features a timeline with a knob indicating the current position in the video (figure 2). This knob can be manipulated with the controller via ray selection (figure 3). Moving the knob along the timeline manipulates the current video time. When the user interacts with the timeline, an equirectangular preview thumbnail appears.

The current video time, a play/pause button and the full video duration can always be found beneath the timeline (figure 2).

3.2 Clock interface

The idea behind the clock design is to create an interface that has fixed, known levels of granularity. The level of granularity in the state-of-the-art interface depends on video length and therefore

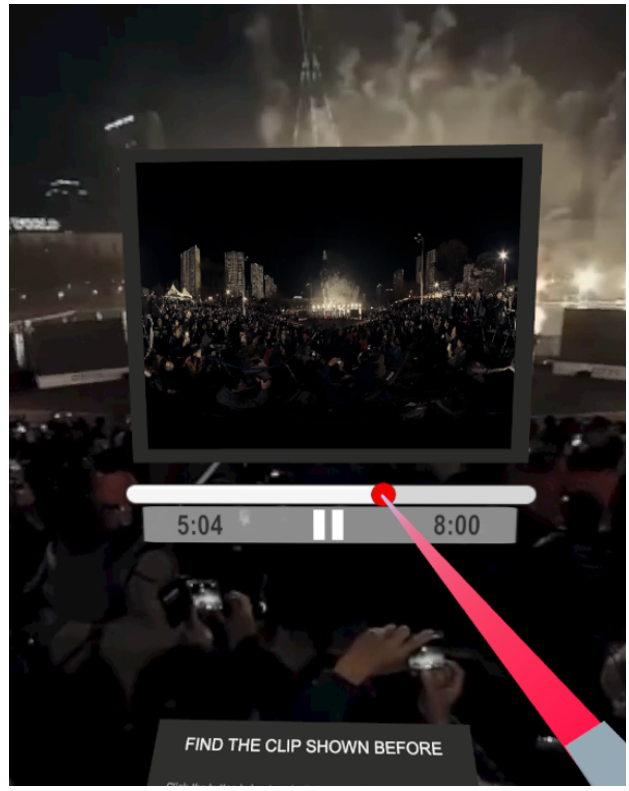

Figure 3: The state-of-the-art interface while the user is scrubbing the timeline

Figure 4: The clock interface as seen in VR

varies. A clock inherently has three levels of granularity represented by the different hands, that are familiar to most people. Although this is not full control over granularity, for general purposes it is assumed that the second hand of a clock provides sufficient precision.

The clock interface consists of three hands (representing seconds, minutes and hours) which can be manipulated directly (figure 4). Using a controller, the user can grab one of the hands and move it back or forth in a circular motion. Doing so moves the video to the corresponding time (figure 5). The other hands move along so the clock always shows the correct video time with all hands. The user

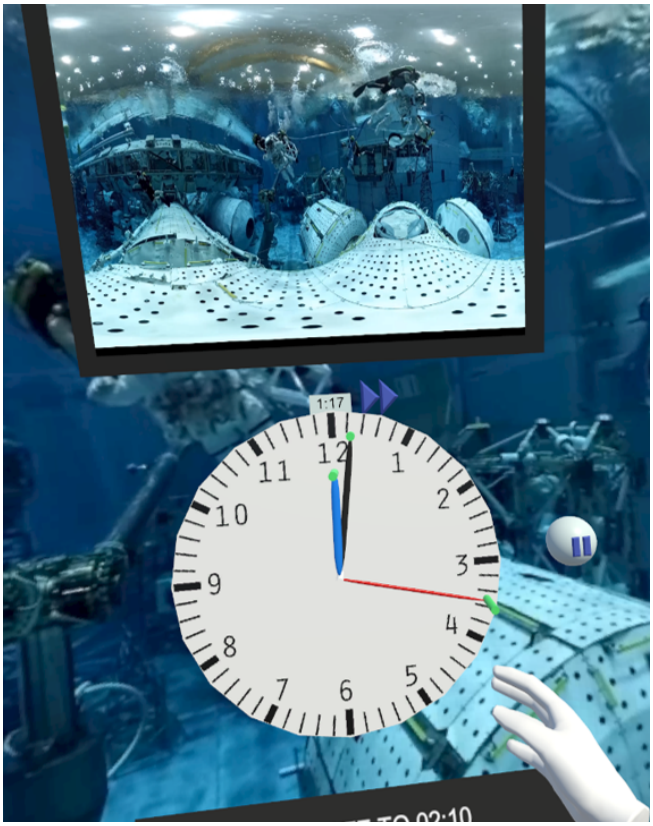

Figure 5: The clock interface while the user is scrubbing through the video

has access to three different levels of scrubbing granularity (represented by the three clock hands). An equirectangular projection is visible as a thumbnail when one of the hands is held (figure 5).

The current video time is displayed above the clock in digital format. The digital format can be useful to quickly determine the precise timestamp, as one generally is not used to reading the second hand of a clock. The video can be paused by interacting with a sphere on the right of the clock (figure 4).

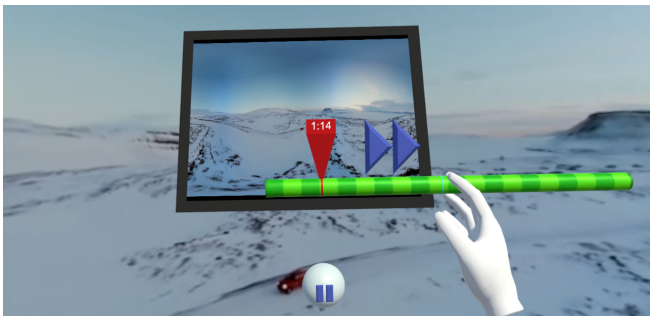

Figure 6: The stretchable interface as seen in VR while the user is scrubbing the timeline

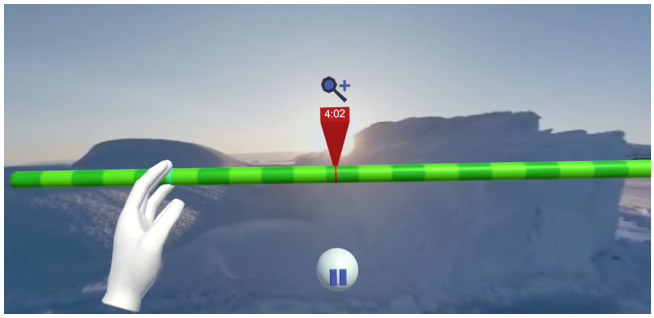

Figure 7: The stretchable interface while the user is scaling the timeline

3.3 Stretchable interface

The stretchable timeline main focus is on the interaction method, while being inspired by the state-of-the-art timeline. Interactions that mimic real-world actions are more natural, and therefore preferred over more abstract, derivative ones [15]. Allowing users to directly interact with a timeline by grabbing it is expected to lead to higher levels of control. In addition, the control over the length and therefore scale of the timeline is granted to the user. The user can therefore adapt the level of granularity to each individual situation at will.

The stretchable interface is a timeline represented by an elongated cylinder. Above the cylinder is an arrow indicating the current point of time. The user can grab the timeline with their right hand and move it past this arrow, which is static, to change the video time (figure 6). When scrubbing, an equirectangular thumbnail appears above the timeline. Figure 8 shows the headset and controllers while scrubbing the timeline.

The user can also increase or reduce the size of the timeline by grabbing it with their left hand (figure 7). By moving the left hand along the timeline, the latter stretches or compresses based on user input. By scaling the user effectively increases or reduces the level of granularity of the timeline.

Above the centre arrow is a display with the current video time. By showing the video time users can infer the required distance to move to a specific point. Below the timeline cylinder is an interactable sphere that can be used to (un)pause the video (figure 7).

4 METHODOLOGY

In order to answer the research questions a within-subjects experiment is done. It is expected to be easier for participants to rate and rank interfaces if they have seen multiple interfaces. This is especially true when including the state-of-the-art interface which gives participants a familiar point of comparison. In addition, due to the expected difference in levels of VR experience, the participant having to be on-site, and the relatively large overhead to get a participant set up in VR, a within-subjects with longer experiment sessions is preferable over a between-subjects version with short sessions.

Figure 8: The HTC VIVE PRO headset while interacting with the stretchable interface

4.1 Procedure

Participants start by reading the information sheet (see appendix 22) and signing the consent form (see appendix 23). In addition the symptoms of cybersickness are repeated and expanded on verbally. Participants are told to close their eyes and carefully remove the headset should they feel uncomfortable.

Before heading into VR, the two task types are explained (see 4.2.3). Next, the participant is seated on a rotating office chair and is handed the HTC VIVE headset and controllers. After entering VR and adjusting the headset if needed, the participant is presented with a start screen and can proceed with the first part of the study.

The study is divided into three segments, where each segment uses a different interface. Each segment starts with a tutorial where the participant can play around with the interface. All possible actions are listed in VR and participants are encouraged to try them out. Meanwhile, the researcher monitors the participant's VR view and, if needed, calls attention to any overlooked instructions.

Once the participant feels they have understood the interface, they move on to the tasks. Each segments contains two of each type of tasks, for a total of four. Each task uses a different video, meaning that after performing four tasks for all interfaces no duplicate videos will have been encountered.

During the first task type, the timestamp task, the participant is shown an objective timestamp (for example 1m30) and are prompted to navigate the video to the specified time as quickly as possible. It is left to the participant to judge how precise their answer should be. After arriving at this point, the participant presses a button to complete the task. Automatic completion was considered, but the risks of accidental completions and artificial delays was deemed to large. This task is chosen as it represents the use case of a user knowing that a specific event will happen at a certain point in time. For example, when skipping an advertisement or finding their favourite moment in a video. By recording the answer provided and the time it took to answer, both accuracy and efficiency can be measured.

For the second task type, the clip search task, participants are first shown a clip of ten seconds, featuring an easily recognizable scene. After viewing the clip exactly once, participants are then presented with the video the clip was taken from, and are prompted to locate the fragment. Once they think they have found the clip they press the hand in button. It is clearly explained that there is no requirement to find the exact start or end of the clip, but that any moment that is contained in both the clip and the video is a correct answer. In addition it is stated beforehand that should the participant forget the clip they are looking for or cannot find the clip they can simply guess an answer and continue. This task represents the real scenario in which a user is looking for a known moment they may have seen before or otherwise know what it will look like. For example, a user wants to find their favourite scene from a movie, or is watching a news report and wants to skip to the weather. By comparing the provided answers to the actual clip interval and the speed with which the answer was given, the effect of the interface on search task accuracy and speed can be determined.

After the four tasks are complete, the participant will be prompted to take of the headset to fill out a questionnaire (see appendix 26). This questionnaire will be answered once for each interface. It contains the SUS-questionnaire [4] and inquires about possible symptoms of cybersickness. In addition, it contains a blank field for any comments or thoughts they want to share about the previous interface. Once the questionnaire is done, the participant can either opt for a small break or continue with the next segment.

After all interfaces have been seen, a final questionnaire is presented 27. This one asks the participant to make three top three's for which interfaces they consider most efficient, most fun and overall regard as the best. After that some demographic questions are presented inquiring about age group, gender and experience with VR.

At this point the experiment is complete, and participants are thanked and presented with a bag of sweets or chocolate. The researcher then extracts the data and resets the setup for the next participant.

4.2 Materials

4.2.1 Videos. The 360° videos used are either public domain or usable under one of the creative commons licenses. Videos used, licenses and attributions to their creators can be found in appendix 8. The videos are selected based on length, content type (no offensive or anxiety inducing contents) and variation (not all videos should have similar subjects. The videos are cut to one of three sizes, 2m00, 8m00 or 45m00, based on total length.

The selected video are varied in terms of movement type (static, slow moving, fast moving camera), environment (city, air, indoors, outdoors) as well as theme (sports, educational, events). Videos are also selected for having a high variation between different parts, so that search tasks are easier to complete. For example, videos with multiple shots from different environments and places instead of one shot of the same room for the full duration.

After being clipped, the videos are converted to the same format and resolution (2560:1280) using FFMPEG [31]. The resolution used is rather low, however this reduces the performance impact within the unity project, leading to better (faster) interface performance, especially with regard to the thumbnail rendering.

4.2.2 Counterbalancing. Each video is used for both types of tasks (timestamp and clip finding) effectively resulting in 24 different task/video combinations. These combinations are equally distributed along the interfaces based on participant number. This is expected to minimize the difference in task performance caused by specific videos. The order in which participants see the interfaces is distributed in similar fashion. This helps reduce the learning effect for those unfamiliar with such applications, expectations after seeing a specific interface first, and other related biases.

4.2.3 Tasks. For each video a moment is selected that features a couple of unique, easily distinguishable events or objects. A ten second clip is cut around these moments. A different point in the video is selected as objective for the timestamp task. For both tasks the objectives timestamps are balanced so that the average is around 50% of the video, including targets near the start and end of the videos. No participants views the same video twice, so there is no risk of remembering information from a previous task that could influence results.

4.2.4 Test setup. The headset used in this research is the HTC VIVE pro [8]. It features a head mounted display, two controllers and two tracking base stations. The tracking station are set up to cover a small area around the chair of the participant. Like the interface (see section 3, the application used for testing was developed in Unity Game Engine [30]. Regarding cybersickness, most instant scene transitions, which are known to potentially cause symptoms (see section 2.2), a number of between-scene screens are present, consisting of a blank screen with a 'continue' button. In addition, the amount of rotating required of participants is limited as much as possible.

4.2.5 Data collection. The following data will be collected for each participant:

- Completion speed for each task
- Timestamp provided as answer by the participant for each task
- For each interface:
  - A SUS-questionnaire (see appendix 26)
  - Cybersickness symptoms questionnaire (see appendix 26)
- Ranking of the interfaces in three categories (efficiency, fun and total)
- Age group
- Gender
- Level of experience with VR applications

The test application gathers data while the participant is performing the tasks in VR. The questionnaires are executed outside of VR, as this is less cumbersome for the participant and it doubles as a small break to help reduce chances of cybersickness occurring. The survey software Qualtrics [24] is used for this.

4.3 Sampling and recruitment

The target population for this study is a group of potential early adaptors: tech-savvy people who are likely to either own or purchase a VR headset and would therefore most likely watch 360° VR videos. In addition to this group being accessible when running experiments on university grounds, it is also likely they have sufficient experience with desktop video players that feature similar interfaces to the commonly found VR ones. The age requirement of 18-35 year old was therefore chosen. It is expected that older people are less likely to (frequently) use VR and are less likely to watch a large number of on-demand video, and therefore are not part of the desired demographic.

Recruitment is done via announcements posted in several digital channels of the Utrecht University department of Information and Computing Sciences. This announcement contains a link to a survey where after confirming their age, a number of time slots could be selected for availability. The researcher then contacts the participant via e-mail, which includes a description of how to find the room in which the experiments takes place (see appendix 24). The experiments will be conducted over a period of three weeks. The desired sample size lies between 24 and 30 participants.

5 RESULTS

5.1 Demographics

In total, 23 people participated in the study (see appendix 25). Most participants reported to be at least somewhat familiar with VR (see table 1). In addition, three quarters of the participants was below 26 years old (see figure 9), fitting the early adapter demographic well.

Never used VR before	**5**
Used VR once or twice	**14**
Occasionally uses VR	**4**
Regularly uses VR	**0**

Table 1: The reported level of experience with VR applications by the participants.

5.2 Data Processing

Some anomalies were found in the data recorded during the experiments. For the timestamp tasks, two entries were excluded from the calculations as in both cases a different participant had missed the mark by approximately 60 seconds. Both occurred while using the clock interface, presumably as a result of an interface bug. Although a direct result of the improper functioning of one of the interfaces, including these values would greatly skew the data as the rest of the values average at about 1.1 seconds.

The clip search task entries are divided into three categories: *correct* (provided answer falls within the interval of the clip), *close*

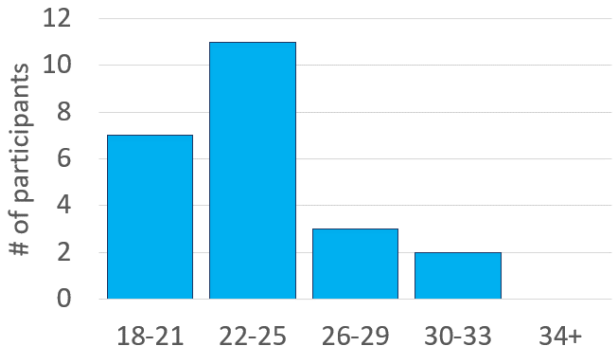

Figure 9: The age groups of the participants, with most participants being between 22 and 25 years of age. Participants were asked to provide their age to confirm that they were within the target demographic, and give a general idea of the distriubution of age.

(provided answer is within 20 seconds from the clip interval) and *incorrect* (provided answer is more than 20 seconds from the clip interval). The division was made to separate the participants who completed a task with a slightly inaccurate answer from those who either forgot the clip or incorrectly identified another section as the clip. The criteria for the *close* category was made based upon the video selection, which occasionally included a similar scene leading up to the start of a clip or continuing after the clip was finished.

5.3 Timestamp task

5.3.1 Speed. The speed with which participants completed the timestamp tasks is not normally distributed. Multiple groups with a categorical predictor value are to be compared. Therefore a one-way ANOVA test was performed which showed a significant difference between the three designs ($F(2,135) = 9.85$ $p < .001$), see figure 10. Post-hoc analysis using the Tukey HSD test revealed a significance between the state-of-the-art and clock interfaces ($Q = 4.52$ $p = .005$) and between the clock and stretchable interfaces ($Q = 6.03$ $p < .001$). No significant difference exists between the state-of-the-art and stretchable interfaces ($Q = 1.51$ $p = .537$).

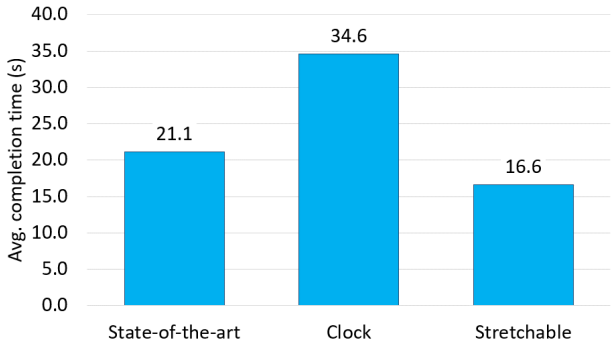

Figure 10: Average speed of timestamp task completion for each interface in seconds (s).

5.3.2 Accuracy. Using a one-way ANOVA test, no significant difference in accuracy ($F(2,133) = 0.04$ $p = .96$) was found for the timestamp tasks. On average participants deviated 1.1 seconds from the target time.

5.4 Clip search task

5.4.1 Speed. The completion time for the clip search task is also not normally distributed. A one-way ANOVA on all answers (including those marked *close* and *incorrect*) revealed no significant difference between the three interfaces ($F(2,138) = 3.96$ $p = .077$). A one-way ANOVA on the task completion time for answers marked *correct* or *close* did reveal a significant relation between interface used and time taken to complete the task ($F(2,114) = 3.96$ $p = .022$). Follow-up with a post-hoc Tukey HSD test confirmed significant difference between the state-of-the-art and clock interfaces ($Q = 3.56$ $p = .035$), as well as between the clock and stretchable interfaces ($Q = 3.42$ $p = .045$). No significant difference was found between the state-of-the-art and stretchable interfaces ($Q = 0.13$ $p = .995$).

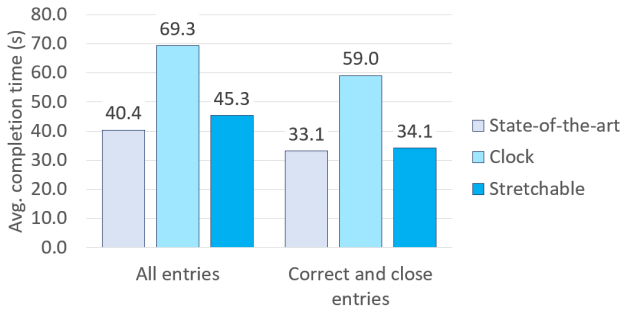

Figure 11: Average completion time for the clip find tasks. On the left all entries are included, on the right only entries marked *correct* or *close* were considered.

5.4.2 Accuracy. The relation between interface used and correct, close and incorrect answers (see table 2 and figure 12) was examined using a chi squared test of independence (as both the predictor and outcome variables are categorical). No significant relation was found (X^2 (4, $N = 139$) = 2.85, $p = .58$). Another chi squared test was performed where the *close* category was added to the *incorrect* category. This test found no significant relation between type of interface used and correctness of the answers (X^2 (2, $N = 139$) = 2.73, $p = .25$).

	State-of-the-art	Clock	Stretchable
Correct	24	25	31
Close	14	13	9
Incorrect	8	9	6

Table 2: Answers given in the clip search task (*correct*; answer within the clip interval, *close*; answer within 20 seconds of the clip interval, *incorrect*; further away).

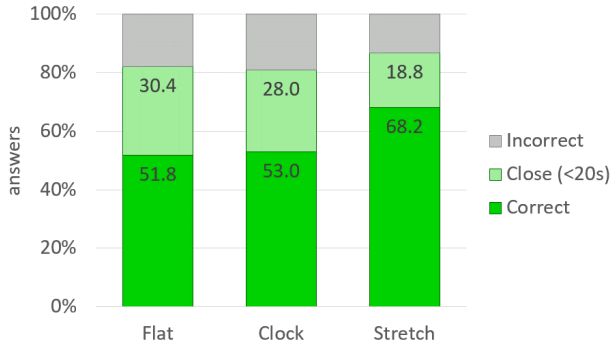

Figure 12: Correctness of answers given during the clip find task. Answers are marked as *correct* when they are within the clip interval, marked *close* if they are less than 20 seconds removed from the clip interval, and marked *incorrect* otherwise.

5.5 SUS-scores

Using a one-way ANOVA a significant difference between all interfaces was found ($F(2,66) = 37.63$ $p < .001$). Using the Tukey HSD as post-hoc test revealed a significant difference between the state-of-the-art and clock interface ($Q = 10.45$ $p < .001$) and clock and stretchable interfaces ($Q = 10.71$ $p < .001$). As with the speed for the timestamp tasks, no significant difference was found between the state-of-the-art and stretchable interfaces ($Q = 0.16$ $p = .993$). Instead, they scored very similar ratings, respectively 82.1 and 82.6 (see figure 13).

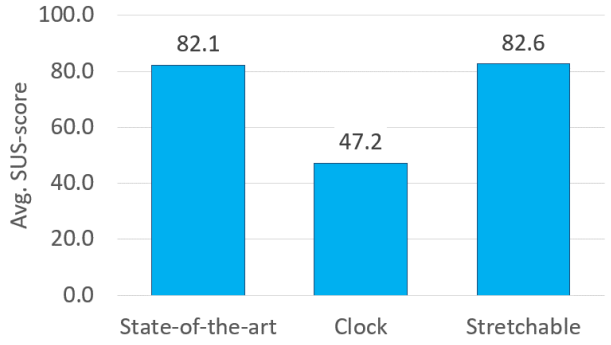

Figure 13: Average SUS-score for each of the interfaces

5.6 Cybersickness

Participants indicated in the questionnaire that they experienced little to no symptoms of cybersickness. The numbers are too small to draw any real conclusion, but most discomfort was experienced when using the state-of-the-art and clock interfaces. However, those participants who experienced symptoms stated it was likely unrelated to the interface and instead occurred because of the content or movement in a video.

5.7 Participant rankings

For each type of ranking made by participants a one-way ANOVA was computed. If a significant difference was found it was followed up with a post-hoc Tukey HSD test.

5.7.1 Efficiency rating. The stretchable interface was most often rated first, closely followed by the state-of-the-art interface (see figure 14). All but one participant rated the clock interface third in terms of efficiency.

There was a significant difference in rating regarding efficiency ($F(2,66) = 78.19$ $p < .001$). The post-hoc test once more revealed a significant difference between the state-of-the-art and the clock interface ($Q = 13.75$ $p < .001$) as well as between the clock and stretchable interfaces ($Q = 16.50$ $p < .001$). No significant difference between the state-of-the-art and stretchable interfaces was found ($Q = 2.75$ $p = .134$).

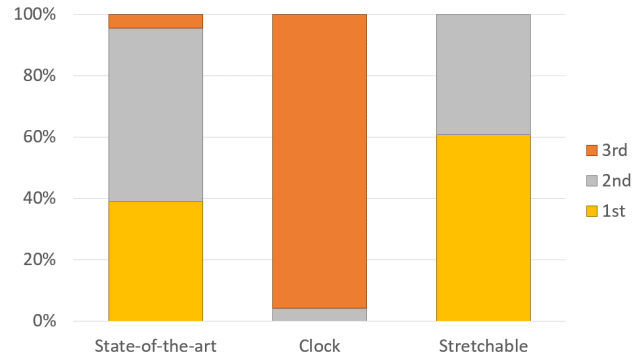

Figure 14: Ranking of the interfaces based on efficiency by the participants. The stretchable interface was ranked #1 most often.

5.7.2 Fun rating. Seven out of twenty-three participants ranked the stretchable interface as most fun (see figure 15). The other interfaces received quite similar scores.

The ratings for most fun interface also showed a significant difference ($F(2,66) = 17.08$ $p < .001$). The post-hoc test showed no significant difference between the stretchable interface with both the state-of-the-art interface ($Q = 7.31$ $p < .001$) and the clock interface ($Q = 7.00$ $p < .001$). The stretchable interface was thus ranked as significantly more fun than the others. No significant difference was found between the state-of-the-art and clock interfaces ($Q = 0.30$ $p = .974$).

5.7.3 Overall rating. In terms of first places, the state-of-the-art and stretchable interfaces both scored the exact same number of first places as with the efficiency rating (nine and fourteen respectively, see figure 16). Another similarity is that the stretchable interface was always awarded first or second and never third place.

Similar to the other comparisons, a significant difference was found between the interfaces ($F(2,66) = 48.96$ $p < .001$). With further analysis, a significant difference was found between the state-of-the-art and clock interfaces ($Q = 10.23$ $p < .001$) and between the clock and stretchable interfaces ($Q = 13.38$ $p < .001$). Once more,

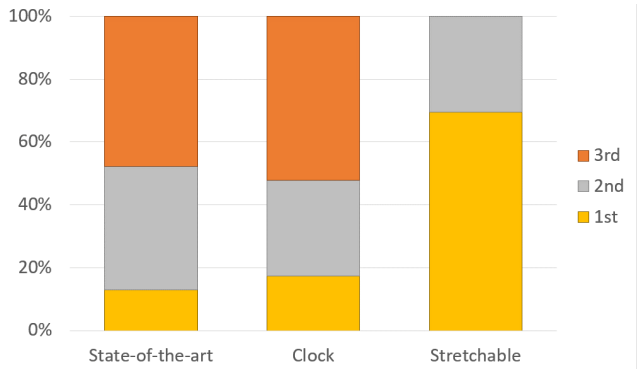

Figure 15: Participant's ratings of the interfaces when asked what the most fun interface was. The stretchable interface was elected as the most fun by more than 66% of the participants.

there was no significant difference between the state-of-the-art and stretchable interfaces (Q = 3.15 p = .074).

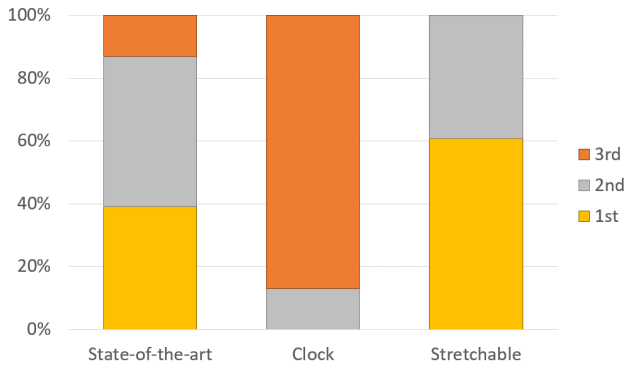

Figure 16: Ratings assigned by participants when asked to rank the interfaces.

6 CONCLUSION

This research did not find conclusive proof that a 3D interface and its increase in control over granularity leads to increased levels of accuracy or efficiency compared to a regular 2D counterpart. Although participants using the stretchable interface were on average slightly faster to reach a certain timestamp compared to the state-of-the-art interface, there was no significant difference.

What can be concluded is that the (used implementation for the) clock interface had several technical issues making it more difficult to operate. On all metrics except fun it scored significantly worse than the other interfaces, and even there it shared last place. The stretchable interface however scored very similar to the state-of-the-art interface on most metrics. Because no significant differences between the best performing interfaces was found, the first and second hypotheses expecting an increase in accuracy and efficiency when using 3D interfaces have to be rejected. However, while the hypotheses have to be rejected, it should be noted that the performance and accuracy of the stretchable 3D interface appears to be on a similar level to its 2D counterpart.

The third hypothesis, stating that an interface designed specifically for VR leads to a higher usability score is rejected as well. Although a significant decrease in usability score was found between the clock interface and the others, no significant difference in scores for the state-of-the-art and stretchable interface was found. It is not unexpected considering the SUS-scores only differ 0.5 points (on a scale of 0-100). The height of the scores does however indicate a high level of usability. Therefore regarding the third research question, a similar conclusion can be drawn to the former two: a 3D interface designed for VR does neither over- nor under-perform compared to the state-of-the-art version.

In conclusion, no decisive benefit was found for using the spacious VR setting to create a 3D interface with more control over granularity, although one of the 3D implementation was considered to be more fun to use compared to its 2D counterpart. It does appear to be the case that users can reach the same level of accuracy and efficiency with a 3D interface compared to a state-of-the-art 2D counterpart.

7 DISCUSSION

7.1 Limitations

7.1.1 Interface design. Starting with the issues of the clock interface. During testing a bug was discovered where the clock interface would, at seemingly random moments, suddenly double scrubbing distance. This of course had a major impact on the interface's measured speed, as often scrolling required additional steps. It also contributed to frustration when using that interface.

Frustration was often already present due to the high level of occlusion experienced with the clock interface. A lot of participants had trouble grabbing the correct hands. Some accidentally grabbed one of the hands when attempting to pause the video, resulting in a jump in time. One participant mentioned *"It was a bit difficult to operate the different clock hands and didn't feel as easy to get to the point where I wanted to be"*.

Finally, it was observed by the researcher that some participants did not operate the clock as expected. Multiple times, when prompted to scrub to for example 2m14, participants moved the minute hand towards the '2' mark on the clock. This of course results in a jump of ten minutes instead of two. It is unknown what caused this behaviour, though it is possible the participants that did this are unfamiliar with analogue clocks, or did not recognize the interface as being similar to an analogue clock.

Participants also reported that the clock was big, and obstructed their view of the video too much: *"The clock is too big, maybe can make it transparent"*. Another participant mentioned that they did not feel as immersed compared to the other interfaces.

Regarding the stretchable interface, a participant commented *"It was difficult to see how long the videos were, and if I wanted to go to 2 minutes it took me some time to figure out where in the bar that would be"*. In addition, some mentioned they considered the sphere used to pause the video not very intuitive. Participants also often attempted to grab the red arrow indicating the current timestamp instead of the timeline. This sometimes led to confusion, which was

usually resolved quickly when the timeline moved instead of the arrow.

A number of participants commented that being able to 'zoom in/out' on the timeline was quite useful. For example one said *"This interface was best at precisely being able to select a time frame"*. In addition, some participants mentioned that they felt more in control while using the stretchable interface. Most comments on this interface were positive, which is a reason to explore this type of interface further.

7.1.2 Sample size. Another limitation of this research is the small number of participants. Some of the metrics measured resulted in small difference between the best 3D interface and the state-of-the-art one. With a larger sample size, it is easier to detect a significant relation if one is indeed present. Apart from digital recruitment, guerrilla sampling was attempted by approaching potential participants killing time between lectures or meetings. However, due to the relatively remote location of the lab convincing people to join was largely unsuccessful. A potential solution would be to hand out (digital) flyers instead, giving time to determine if they are interested in participating. It also gives people a chance to pick a moment convenient for them.

7.2 Miscellaneous observations

As stated in the results section (see 5) little to no cybersickness was experienced by participants. Nobody reported severe discomfort, only slight. This could have been due to a number of factors: participants being seated for the entire experiment, videos featuring mostly stationary or slow moving camera work, or the frequent breaks. Whatever the case, only in fourteen of the total 69 participant/interface combinations mild symptoms were experienced. This is lower than expected, which was a pleasant surprise.

A number of participants was observed to not use the thumbnails at all during the clip search task. Instead, they looked around a lot, then used the timeline to adjust the video, then look around again. Some participants only appeared to use the thumbnails for certain interfaces, presumably due to the size and position of both the interface and thumbnails. In addition, some participants stopped using the thumbnail with a new interface, even though they had used it previously.

Several participants presumably found the use of different interaction methods within the same study confusing. They tried to use the ray to interact with the 3D interface elements. However, most realised their mistake pretty quickly and adapted. A number of participants verbally expressed excitement when they realised they could directly grab the 3D interface elements.

7.3 Future work

Because of the positive reception, as well as the very similar scores of the state-of-the-art interface and its stretchable 3D counterpart, further research in this direction is warranted. It is recommended that a developmental study explores the needs and desires of users in an iterative process. A new iteration of interface specifically suited for VR video applications that is easier to use and provides more control over granularity.

By doing a developmental study, users can be included in the design process. This can be especially beneficial for the smaller parts of the design that are easier overlooked by a small number of designers. In addition design decisions can immediately be tested leading more quickly to a solid design.

For future studies evaluating multiple interfaces it might also be interesting to collect additional data. For example, recording all actions of the user (the amount of times interacted with each part of the interface, the amount of time spent searching in large steps versus precision scrubbing). It was considered for this study, but was deemed out of the scope. For future work it can definitely provide valuable insight.

A greater variation of tasks can provide more information on different cons and pros of interfaces. For example, having to navigate to a couple of separate points in succession in the same video. It can also prove interesting to evaluate an interface as a part of a complete video playing and browsing application. As video players often incorporate options for switching to different videos, it might be interesting to see if a 3D interface would still hold up with these added elements.

In addition, if similar tasks will be used in future research it is recommended to include a 'I don't remember' button, so these entries can be separated from the successfully completed entries if desired. Furthermore, an automated system for task completion could be considered. Although most of the time participants remembered quickly to press the button, some occasionally forgot. In this study, switching between two different interaction methods (direct interaction and ray interaction) in order to press a button might have affected completion time somewhat.

ACKNOWLEDGMENTS

I would like to thank my supervisor Wolfgang Hürst for their patience, as well as insight during creative sessions. I would also like to extend my thanks to all family and friends for the support during this project.

REFERENCES

[1] James Allan, Greg Lowney, Kim Patch, and Jeanne Spellman. 2015. User Agent Accessibility Guidelines (UAAG) 2.0. https://www.w3.org/TR/UAAG20/ Last accessed: 22-02-2022.

[2] Mathias Almquist, Viktor Almquist, Vengatanathan Krishnamoorthi, Niklas Carlsson, and Derek Eager. 2018. The Prefetch Aggressiveness Tradeoff in 360° Video Streaming. In *Proceedings of the 9th ACM Multimedia Systems Conference* (Amsterdam, Netherlands) *(MMSys '18)*. Association for Computing Machinery, New York, NY, USA, 258–269. https://doi.org/10.1145/3204949.3204970

[3] Doug A. Bowman. 2002. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators & Virtual Environments* 11 (2002), 404–424.

[4] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).

[5] Arti Burton. 2020. Blockmesh Tips From David Shaver. https://80.lv/articles/blockmesh-tips-from-david-shaver/

[6] Axel Carlier, Vincent Charvillat, and Wei Tsang Ooi. 2015. A Video Timeline with Bookmarks and Prefetch State for Faster Video Browsing. In *Proceedings of the 23rd ACM International Conference on Multimedia* (Brisbane, Australia) *(MM '15)*. Association for Computing Machinery, New York, NY, USA, 967–970. https://doi.org/10.1145/2733373.2806376

[7] Federico Chiariotti. 2021. A survey on 360-degree video: Coding, quality of experience and streaming. *Computer Communications* 177 (2021), 133–155.

[8] HTC Corperation. 2022. HTC VIVE pro. https://www.vive.com/eu/product/vive-pro/

[9] Philip Day. 2016. http://www.whirligig.xyz/

[10] Sara Di Bartolomeo, Aditeya Pandey, Aristotelis Leventidis, David Saffo, Uzma Haque Syeda, Elin Carstensdottir, Magy Seif El-Nasr, Michelle A. Borkin, and Cody Dunne. 2020. *Evaluating the Effect of Timeline Shape on Visualization*

Task Performance. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376237

[11] Yasin Farmani and Robert J. Teather. 2018. Viewpoint Snapping to Reduce Cybersickness in Virtual Reality. In *Proceedings of the 44th Graphics Interface Conference* (Toronto, Canada) *(GI '18)*. Canadian Human-Computer Communications Society, Waterloo, CAN, 168–175. https://doi.org/10.20380/GI2018.23

[12] Simon N.B. Gunkel, Martin Prins, Hans Stokking, and Omar Niamut. 2017. Social VR Platform: Building 360-Degree Shared VR Spaces. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video* (Hilversum, The Netherlands) *(TVX '17 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 83–84. https://doi.org/10.1145/3084289.3089914

[13] Keita Higuch, Ryo Yonetani, and Yoichi Sato. 2017. EgoScanning: Quickly Scanning First-Person Videos with Egocentric Elastic Timelines. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6536–6546. https://doi.org/10.1145/3025453.3025821

[14] Chris J. Hughes and Mario Montagud. 2020. Accessibility in 360° video players. *Multimedia Tools and Applications* 80, 20 (Oct 2020), 30993–31020. https://doi.org/10.1007/s11042-020-10088-0

[15] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-Based Interaction: A Framework for Post-WIMP Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 201–210. https://doi.org/10.1145/1357054.1357089

[16] Eugenia M Kolasinski. 1995. *Simulator sickness in virtual environments*. Vol. 1027. US Army Research Institute for the Behavioral and Social Sciences.

[17] Lourdes Moreno, María González-García, Paloma Martínez, and Yolanda González. 2017. Checklist for Accessible Media Player Evaluation. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 367–368. https://doi.org/10.1145/3132525.3134791

[18] Oculus. 2014. Oculus VR Best Practices Guide. http://brianschrank.com/vrgames/resources/OculusBestPractices.pdf

[19] Oculus. 2016. Oculus video on oculus rift. https://www.oculus.com/experiences/rift/926562347437041/

[20] Toni Pakkanen, Jaakko Hakulinen, Tero Jokela, Ismo Rakkolainen, Jari Kangas, Petri Piippo, R. Raisamo, and Marja Salmimaa. 2017. Interaction with WebVR 360° video player: Comparing three interaction paradigms. *2017 IEEE Virtual Reality (VR)* (2017), 279–280.

[21] William Pellas and Sandra Thenstedt. 2020. *The Importance of Colour Guided Navigation*. Ph.D. Dissertation. Department of Natural Sciences, Technology and Environmental Studies.

[22] Benjamin Petry and Jochen Huber. 2015. Towards Effective Interaction with Omnidirectional Videos Using Immersive Virtual Reality Headsets. In *Proceedings of the 6th Augmented Human International Conference* (Singapore, Singapore) *(AH '15)*. Association for Computing Machinery, New York, NY, USA, 217–218. https://doi.org/10.1145/2735711.2735785

[23] Brian Michael Poblete, Emir Christopher Mendoza, Julian Paolo De Castro, Jordan Aiko Deja, and Giselle Nodalo. 2019. A Research through Design (Rtd) Approach in the Design of a 360-Video Platform Interface. In *Proceedings of the 5th International ACM In-Cooperation HCI and UX Conference* (Jakarta, Surabaya, Bali, Indonesia) *(CHIuXiD'19)*. Association for Computing Machinery, New York, NY, USA, 166–171. https://doi.org/10.1145/3328243.3328265

[24] QualtricsXM. [n.d.]. Qualrtics Experience Management. https://www.qualtrics.com/

[25] Lisa Rebenitsch. 2015. Managing Cybersickness in Virtual Reality. *XRDS* 22, 1 (nov 2015), 46–51. https://doi.org/10.1145/2810054

[26] Lisa Rebenitsch and Charles Owen. 2016. Review on Cybersickness in Applications and Visual Displays. *Virtual Real.* 20, 2 (jun 2016), 101–125. https://doi.org/10.1007/s10055-016-0285-9

[27] Lisa Rebenitsch and Charles Owen. 2021. Estimating cybersickness from virtual reality applications. *Virtual Reality* 25, 1 (2021), 165–174.

[28] Kyle Tanchua. 2021. What does the metaverse look like in the real world? get a peek in this concept video. https://www.tech360.tv/metaverse-real-world-concept-video

[29] Christophe Tauziet. 2016. Designing for hands in VR. https://medium.com/designatmeta/designing-for-hands-in-vr-61e6815add99

[30] Unity Technologies. [n.d.]. Unity Game Engine. https://unity.com/

[31] Suramya Tomar. 2006. Converting video formats with FFmpeg. *Linux Journal* 2006 (2006), 10.

[32] W3C. 2018. Web content accessibility guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/ Last accessed: 22-02-2022.

[33] XRInteractionToolkit. [n.d.]. XR Interaction Toolkit. https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@0.10/manual/index.html

[34] Abid Yaqoob, Ting Bi, and Gabriel-Miro Muntean. 2020. A survey on adaptive 360 video streaming: solutions, challenges and opportunities. *IEEE Communications Surveys & Tutorials* 22, 4 (2020), 2801–2838.

[35] Leo Zeches. 2022. Rotation Methods for 360° Videos in Virtual Reality - A Comparative Study (master thesis). *Utrecht University* (2022).

[36] Yang Zhao, Ye Tian, and Yong Liu. 2015. Extracting viewer interests for automated bookmarking in video-on-demand services. *Frontiers of Computer Science* 9, 3 (2015), 415–430.

8 APPENDICES

8.1 Videos

The following videos were used in the experiments of this study. All videos were cut to one of the three lengths used (2m30, 8m00 or 45m00). Links to the creative commons licenses are below.

Creative Commons Attribution-Share Alike 4.0 International (https://creativecommons.org/licenses/by-sa/4.0/deed.en) Creative Commons Attribution 3.0 Unported (https://creativecommons.org/licenses/by/3.0/deed.en) Creative Commons CC0 1.0 Universal Public Domain Dedication (https://creativecommons.org/publicdomain/zero/1.0/deed.en)

- *The Call of Science* by NASA Jet Propulsion Laboratory (Public domain) https://commons.wikimedia.org/wiki/File:Earth_360_Video_The_Call_of_Science.webm
- *From Dover to Dunkirk: A 360° Spitfire Experience* by World of Warplanes (Creative Commons Attribution 3.0 Unported, edited) https://commons.wikimedia.org/wiki/File:FromDover_to_Dunkirk_A_360_Spitfire-Experience.webm
- *NASA VR/360 Astronaut Training: Space Walk* by NASA (Public domain) https://commons.wikimedia.org/wiki/File:NASA_VR-360_Astronaut_Training-_Space_Walk.webm
- *Hundra knektars marsch på Forum Vulgaris* by Jan Ainali (Creative Commons Attribution-Share Alike 4.0 International, changed made) https://commons.wikimedia.org/wiki/File:Hundra_knektars_marsch_p%C3%A5_Forum_Vulgaris.webm
- *NYC in 360 - Surviving COVID* by Joseph A. Eulo (Creative Commons Attribution 3.0 Unported, changed made) https://commons.wikimedia.org/wiki/File:NYC_in_360_-_Surviving_COVID.webm
- *360VR Lotte Tower Grand Opening Fireworks (South korea)* by Kim Jaesung (Creative Commons Attribution 3.0 Unported, edited) https://commons.wikimedia.org/wiki/File:360VR_Lotte_Tower_Grand_Opening_Fireworks(South_korea).webm
- *The Swellies Menai Straits - Yacht Testa Rossa* by Eddy Jackson (Creative Commons Attribution 3.0 Unported, edited) https://commons.wikimedia.org/wiki/File:360_VR_Video_-_The_Swellies_Menai_Straits_-_Yacht_Testa_Rossa.webm
- *PARA 360 Tolmin SokoleONE - UP Kangri* by Sokole ONE (Creative Commons Attribution 3.0 Unported, edited) https://commons.wikimedia.org/wiki/File:2021_06_12_PARA_360_Tolmin_SokoleONE_-_UP_Kangri.webm
- *Anıtkabir (Atatürk) 360 Derece Video Panorama Gezinti* by 360 TR (Creative Commons Attribution 3.0 Unported, edited) https://commons.wikimedia.org/wiki/File:An%C4%B1tkabir_(_Atat%C3%BCrk_)_360_Derece_Video_Panorama_Gezinti_-_360_Degree_Video.webm
- *3D Video of a short flight in Newport and Laguna by D Ramey Logan in a Cessna* by Don Ramey Logan (Creative Commons CC0 1.0 Universal Public Domain Dedication, edited)https://commons.wikimedia.org/wiki/File:3D_Video_of_a_short_flight_in_Newport_and_Laguna_by_D_Ramey_Logan.webm
- *Wind Tunnel Test of NASA's Most Powerful Rocket (360° Animation)* by NASA and USGov (Public domain)https://commons.wikimedia.org/wiki/File:Wind_Tunnel_Test_of_NASA%E2%80%99s_Most_Powerful_Rocket_(360%C2%B0_Animation).webm
- *Fly Above Alaskan Glaciers in 360* by NASA, Goddard (Public domain) https://commons.wikimedia.org/wiki/File:Fly_Above_Alaskan_Glaciers_in_360_Y2eysSmn9VU.webm
- *360 Video of LCS-15 Christening and Launch* by U.S. Navy (Creative Commons Attribution 3.0 Unported, edited) https://commons.wikimedia.org/wiki/File:360_Video_of_LCS-15_Christening_and_Launch.webm
- Creative Commons Attribution-Share Alike 4.0 International (https://creativecommons.org/licenses/by-sa/4.0/deed.en)
- Creative Commons Attribution 3.0 Unported (https://creativecommons.org/licenses/by/3.0/deed.en)
- Creative Commons CC0 1.0 Universal Public Domain Dedication (https://creativecommons.org/publicdomain/zero/1.0/deed.en)

9 IMPLEMENTATION

The three interfaces were implemented in the Unity game engine [30]. The XR Interaction Toolkit [33] handles all VR interactions. The state-of-the-art interface used an implementation by [35]. The other interfaces were implemented for this research. Below is a more in depth description of the components and features of each interface.

9.1 State of the Art

The state of the art interface is a custom implementation of the common interface found in (360°) VR video players. It features a two-dimensional panel that responds to ray-cast style interaction (see figure 17). The panel contains a timeline and a number of buttons. The buttons available are a play/pause button and skip forward and backward buttons. The timeline is very similar to normal video player interfaces. Moving the handle of the timeline will move the video to that position in time. During scrolling an equirectangular thumbnail appears above the timeline which can be used for searching.

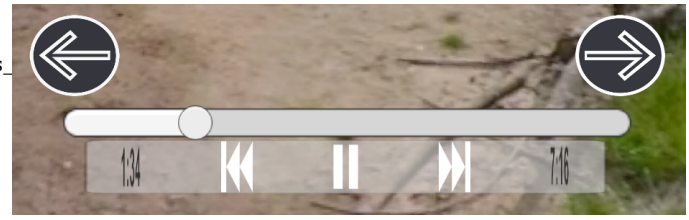

Figure 17: State-of-the-art interface as seen in VR

9.2 Clock Interface

The clock interface has all basic analogue clock components. It features an hourplate, an hour, a minute and a second hand respectively (see figure 18). In addition, each hand features a handle at the end to minimize occlusion. Each hand can be interacted with by directly grabbing the corresponding handle with the controller.

Manipulating one hand automatically moves the others, like a real mechanical clock.

On top of the clock is a small display which features the current video time in digital notation. Located to the right of the clock is a small interactable pause/play sphere. When moving the minute or hour hand, a partial disk appears to indicate the largest possible value for that hand. For example, when the minute hand is selected in an half-hour long video, a disk covering the clock from 6-12 hours would appear to indicate the video is 30 minutes long. Just like the state-of-the-art implementation, an equirectangular thumbnail appears above the interface during interaction.

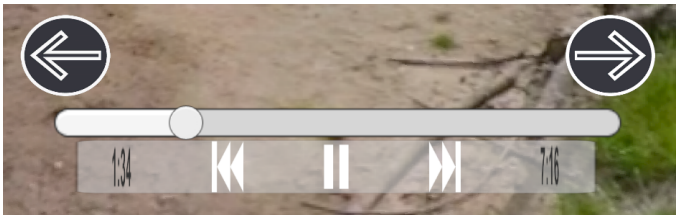

Figure 18: Clock interface as seen in VR

9.3 Stretchable Interface

The final interface is centered around a time indicator is the cylindrical timeline (see image 19).

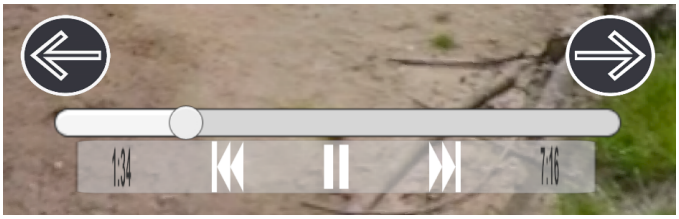

Figure 19: Stretchable interface as seen in VR

10 LITERATURE STUDY

The social side of 360° videos in VR are being explored, such as watching videos in VR together [12, 28]. In addition, very serious applications like use in law enforcement or detective work are being studied as well [23]. However, interfaces for casual viewing of 360° videos have not been studied extensively, and still closely resemble their desktop counterpart. The most popular 360 media players of this moment (for example *Oculus Video* [19] or *Whirligig* [9]) feature a 2D interface. The aim of this study is to look into components of a 3D interface for a 360° video player specifically designed for VR. This study examines research on concepts relating 360° videos, video players in general, VR interface design and evaluating interfaces in VR.

10.1 360° Videos

Almquist et al [2] defined four categories of 360° videos based on user behaviour when watching the video. Videos that have a clear main point of interest at a static position throughout the video are classified as *static focus* videos. In contrast, videos that contain a clear main point of focus that moves around the screen are categorized as *moving focus* videos. Videos that have no specific point of interest, or that are equally interesting in each direction are known as *exploration* videos. The final type of video is *rides*, in which the camera usually moves along a track like a roller coaster. In these videos the user is often expected to look forward most of the time.

This classification of different types of videos have consequences for both design and the evaluation of a 360° video interface. Regarding design, finding a specific event through scrolling in a static focus video is very similar to the process on a 2D screen. However, in a moving focus video, the point of interest may move to a different orientation. In that case the user should be able to find that point of interest while scrolling, without knowing its position beforehand. Therefore some way to adapt the viewing direction when scrolling should be implemented in a video player.

For evaluation of an interface, it is important to use a diverse set of videos from all categories. If a player is tested with only one type of video results might favour a design decision which only benefits that type of video, while simultaneously being detrimental for other types.

Petry et al [22] proposed a clear distinction in interaction between time navigation and spatial navigation within a 360° video player. Gaze direction is used solely for panning the video, while a simple set of gestures is used for temporal navigation. It is vital that the different types of navigation do not conflict with each other to avoid confusion.

10.2 Timelines

The effects of the shape of a timeline visualisation were studied by Di Bartolomeo et al [10]. They compared task execution time and accuracy on linear, circular and spiral timelines. The results showed that participants were quicker at performing the tasks on linear timelines. No significant differences in accuracy were found. According to the researchers, the user's familiarity with linear timelines was a contributing factor to its superiority. In any case, a linear shaped timeline seems to be the safest bet in a classic environment.

A research by Higuch et al introduced the concept of an elastic timeline [13]. A visual analysis of the (often lengthy) first-person video resulted in a set of segments of potentially interesting moments. The video would then play at an increased playback rate until such an interesting moment was encountered (for example a conversation with another human). Then, the playback rate would drop to normal speed, until the end of the fragment. The result is a timeline where certain intervals are stretched as they contain interesting events. A major advantage of an elastic timeline is the option to increase the granularity of a specific segment, without altering the scale of the entire timeline. This allows for more precise scrolling within an interval

A common issue with time manipulation in VR video players is the effects of time jumps. Whenever an arbitrary moment is selected to skip to, the user's entire world flashes to that new position in

time. This is known to cause nausea very quickly, and is one of the symptoms of cybersickness (see section 10.6.1 [16].

The large data requirement of streaming 360° videos, especially in VR, is a major challenge [7, 34]. In 2D video streaming interfaces, bookmarks displayed on timelines allow pre-rendering of parts of the video the user might skip to during video streaming [6]. In addition, bookmarks have been proven to reduce search times within videos [36]. The latter is most interesting for interface design, as this literature study does not cover the technical challenges.

In VR it is desirable to minimize the amount of times the user has to jump to a different moment in time due to cybersickness (see section 10.6.1). Bookmarks can help improve navigation, and should be considered when designing a video player interface.

10.3 Interactions in VR

New interaction methods are constantly developed as the technology they are designed for evolves. For example, smartphones use interfaces that simulate physical properties such as gravity and inertia. Jacob et al [15] stated that this is a step in the process of moving away from the early text-based interfaces, and moving toward interfaces that are real-world inspired. By mimicking real world interactions in interfaces, it is easier for users to operate them. If the interaction is more natural, it requires less time to learn to perform it. Jacob et al identified a number of themes that can aid designing these reality-based interactions. These themes are *naive physics*, *body awareness and skills*, *environment awareness and skills* and *social awareness and skills*. The first three themes are relevant for 360° video players and should be considered when designing such a system.

Every human has a basic understanding of the physical world, including gravity and the persistence of objects. This is referred to as naive physics. Smartphone interfaces often use basic physical phenomenon for small components (such as simulated inertia when swiping across a document). In VR this can be taken to the next level by having an interface that for all intents and purposes obeys the laws of physics as we know them from the real world. As a result, the interface is expected to be much more natural [15].

Body awareness includes proprioception (being aware of the relative position of limbs), reach and movement coordination. Currently, most VR interfaces are designed to be interacted with one hand at the time. As humans are pretty good at coordinating both hands simultaneously (for example, humans are able to operate a mouse and keyboard at the same time), this possibility should not be forgotten in regards to interface interactions.

Environment awareness is related to the perceived physical presence in an environment. Objects observed by humans function as landmarks and orientation points. For example, the horizon tells us something about the angle we are facing, while shadow helps determine the distance of objects related to ourselves and to each other. Knowing exactly how and where our body is compared to the scene helps interacting with items in that scene [15]. It is therefore important to provide sufficient objects in a scene so that a user perceives themselves as physically present. In the context of a 360° video player this is challenging, as the interface designer has little to no control over the videos that will be played.

10.3.1 Controller Input. Many different ways of interacting with VR UI's (user interfaces) have been researched. A couple of these methods are widely used when operating video players in VR. The first logical step coming from a two dimensional video player interfaces is the *remote control* interaction. Within the video player a 2D area is added with virtual buttons (for example a pause/play button) and other elements such as a timeline. These buttons can be selected using the buttons on the controllers, like a classic television remote [20].

Similarly, this type of interface can be operated using the controller as a pointer device, simulating a computer mouse. The controller shoots a ray, which can be aimed at intractable elements. Pressing the interact button on the controller activates the currently selected element. This interaction style usually allows the user to directly manipulate a timeline (as is common in classic 2D video players).

10.3.2 Gestures. Alternatively, gestures can be used to control the video playback. The user uses their controller to perform gestures within the language defined for the video player. For example, if the user moves their controller away from their chest in a straight line, the video pauses. Any number of gestures can be implemented as long as they do not conflict.

When using gestures it is important to make a selection that is easy to learn. Complex gestures both difficult to perform and remember. In addition, gestures should not be cumbersome when performed repeatedly. Continuously waving your arms at full length in order to scroll through a video can become exhausting quickly [29].

Ideally, gestures are also natural. In the context of gestures in VR this usually involves mimicking an interaction from the real world. As discussed in the *interaction* section (10.3) having gestures reflect a real world counterpart helps with making them feel natural. For example, a VR user is presented with a globe, and is tasked to rotate it to show their favourite country. A natural gesture to rotate the globe would be to move your hand alongside the globe while touching it, just like you would in real life [29].

A disadvantage of gestures is that interactions often take longer to complete compared to other methods. They are also experienced as more frustrating and demanding [20].

10.3.3 Gaze. An interaction that is used is based on gaze. A very common interaction in VR is of course looking around. However, gaze can also be used to interact with a UI. For example, buttons can be selected by looking directly at them. Then, buttons can be activated by keeping your gaze fixed at that button for a certain period of time. A major advantage of this method is that it is completely hands-free. However, this method is often perceived as less accurate by users [20].

10.4 Design Considerations

Moving interfaces, for example those attached to a body part, can be difficult to interact with. They are best suited for quick, imprecise actions [29]. Therefore they might be useful for quickly pausing a video in a VR player. It seems inadvisable to design an entire interface around it.

As is the case for all interfaces, interactable elements should communicate both that they can be interacted with. In 2D interfaces, buttons are often clearly marked with an outline, or react when selected by changing colour or brightness. In VR, for example, a certain colour can be used to mark anything that can be directly interacted with [21]. This is a common practice in game design [5].

In addition, elements should also indicate what (type of) interaction is expected to operate them (see section 10.3)[15]. As an example, when a lever is present in a VR setting, the handle can be bright red to mark it as interactable, while its shape and position communicate how it can be moved.

10.5 Accessibility

When designing any kind of interface it is important to take into account the diverse user base. This applies especially to multimedia interfaces such as video players, as they are gateways to vast amounts of digital media. If a user, due to disabilities or other reasons, is unable to operate the interface acting as gateway, they are effectively locked out this whole collection of media. Guidelines exist for accessible web design [32], as well as more specific recommendations for 'user agents', which include media players [1].

In 2017 Moreno at [17] published a concise checklist of accessibility guidelines specifically on the domain of video players. This list contains a selection of checks based on the guidelines mentioned before ([1, 32]) that can be used to evaluate video player designs.

Hughes and Montagud [14] conclude that most available 360° video players currently provide little support for accessibility. They identified four key features that promote broad access. These features include support for transcripts, subtitles, sign language and audio description tracks. They plead for these features to be implemented in such a manner they can be toggled on or of as desired. Right now, these features are often not implemented within the player, but instead provided by means of an alternative video file.

When creating prototypes accessibility should be kept in mind. If accessibility features are not added right away, designs should leave sufficient room to implement them in a later stage. When the design process reach its later stages it is good practice to use the materials available to evaluate the accessibility of the design.

10.6 Interface Evaluation

10.6.1 Well-being. The well-being of both testers and users is extremely important. Cybersickness is a common occurrence in VR application users, as it is at some point experienced by the majority of users [25, 26]. Cybersickness can occur in all kinds of situations, for example when playing video games or when watching a movie. So although it is not limited to VR applications, it poses a risk especially in VR if the user is not in a seated position. Cybersickness has been studied extensively [26, 27], as well as ways to reduce its effect [11, 25].

Rebenitsch [25] published an article which contained several good design practices regarding cybersickness. Regular interaction with a VR environment is known to reduce the effects and susceptibility. An important factor is the time spent in VR in one session. The longer the session, the higher the chance of symptoms. It is therefore important to keep evaluation sessions as brief as feasible, and no longer than 15 minutes without breaks, to minimize the risk of cybersickness occurring as a result of the evaluation. Another aspect that can help reduce cybersickness that is mentioned is to have users take a seat. Sitting helps to avoid dizziness, which in turn lowers the chances of cybersickness.

Another common cause of cybersickness is the quickly changing of the scenes [16]. In VR the scene is all around the user, changing the entire environment with the snap of a finger can be quite discombobulating. In a 360° video player that allows the manipulation of time this is quite an often occurring event. It is therefore wise to include some sort of fade-out/fade-in system when switching scenes or skipping in time.

When testing it is extremely important to monitor the participant for any symptoms, and introduce a break or abort the study when necessary.

10.6.2 Usability. Usability is most commonly measured using the well-established SUS-test [4]. Although the possibility to be applied to practically any digital system is one of its strong points, it is good practice to also look at domain specific issues. Bowman et [3] identified a number of such issues as well as a number of considerations related to evaluating the usability of virtual environments. In a desktop environment, input methods are often limited to one or two simultaneous actions (key presses and mouse movements). In VR the user input often consists of more synchronized actions and elements. For example, holding down a button while performing a gesture with a controller, starting from a specific location in the virtual space.

Therefore, when performing a qualitative study on a virtual space, Bowman et al [3] suggest to either employ multiple observers or record the session. Another recommendation is to keep interaction with the user limited, as immersion is an important factor for most virtual environments. Interacting with the user has a chance to break this immersion, potentially influencing the results of the study.

A ORIGINAL DESIGN CONCEPT

A.1 Stretchable interface Components

The stretchable video player interface consists of two main components: a stretchable timeline and a preview thumbnail (see figure 20). The timeline is a 3D bar that is present in front of the user, a bit below the equator. At the center of the interface there is a knob (like the slider in state of the art interfaces, only it does not slide in this interface), which represent the current position in the video time-wise. Above the knob is an equirectangular thumbnail representation of the current time. As opposed to the state of the art, the timeline moves along the knob instead of the other way around. In addition, the timeline can be scaled by the user to increase or reduce (local) granularity.

A.2 Clock-based interface Components

The clock-based interface instead removes the concept of a timeline, and maps the duration of the video to a 3D clock (see figure 21). A clock comes with three levels of granularity represented by the different hands. This interface uses these different levels represented by the three hands of the clock. The hands are distinct not only

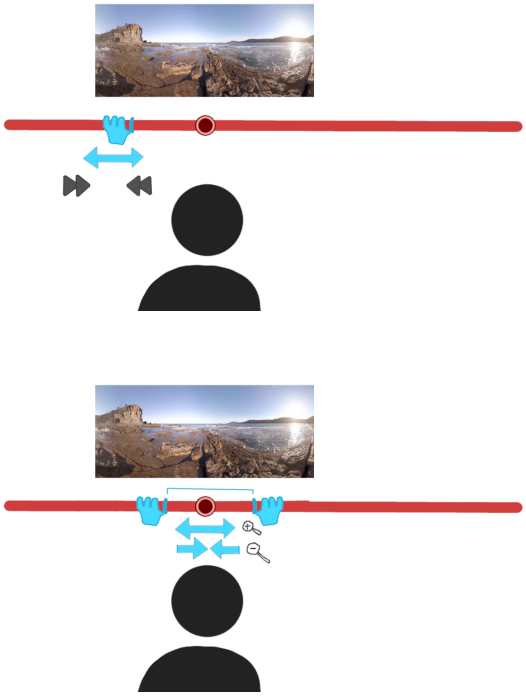

Figure 20: A sketch of the stretchable interface. Blue hands represent the virtual hands (controllers) of the user. The diagrams show the two basic interactions. The user is able to manipulate advance and revert the video by using one handed gestures, and scale the timeline by using two handed gestures.

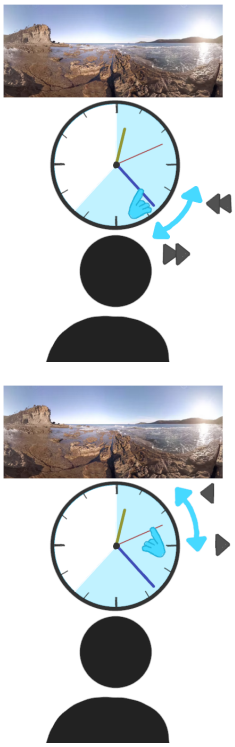

Figure 21: A sketch of the clock-based interface. Blue hands represent the virtual hands (controllers) of the user. The highlighted area represents the area and level (hours/minutes/seconds) of the area that can be navigated. Using the different hands different scrolling speeds can be attained.

in size, but also in colour. The hands can be grabbed by the user and dragged in a circular motion to manipulate time. Like a normal clock, each combination of positions for the hands represents a moment in the video. Depending on the length of the video the handles cannot be dragged further than their maximum position. For example, for a video of less than 60 seconds, the interactable part of the clock is highlighted with a colour similar to that of the seconds hand.

B SCRUBBING

Normal, 2D video players commonly have a timeline with a knob that moves along the timeline. This knob represents the current position in time related to the total duration of the video. This knob can be dragged along the timeline to move the video to a chosen point in time. A major issue with this system is the problem of granularity: the timeline is always the same length, namely the width of the video player. Respectively the left and right side of the window in which the video is played are the timeline's anchor points. As a result, the same distance on the timeline represents different distances based on the video's total length. In short videos, it is easier to reach the desired spot. In longer videos, this is much more challenging, as a slight nudge of the knob jumps the video forward much further.

For 360° videos in VR, the length of the timeline is not limited by the edges of the video, as there are no edges. Therefore, there is no clear logical starting or ending place for a timeline. It is however very important that a VR timeline has a suitable length. If the timeline is long, interacting with it will require additional head movement from the user. As a best-practice for interface design, it is recommended that the interface sits in the middle 1/3rds of the display to avoid the need for head movement [18]. Additionally, gestures required to operate the timeline will require more movement, resulting in faster exhaustion of the user. If a timeline is small, the granularity becomes so low it becomes more difficult to operate the timeline accurately. Most state of the art interfaces employing the pointer interaction technique, where a ray is cast from the controller towards the interface. This functions similar to a computer mouse. Depending on the virtual distance to the interface, rotating the controller a little bit has a much larger effect on the interface. Especially in combination with a smaller timeline this can make accurate timeline scrolling extremely challenging.

In VR there is a much higher degree of freedom in terms of interactions. In addition, there is a royal amount of space available

around the user in VR compared to a desktop or mobile interface. The granularity issues arise from having a confined space in which to present the video in, as well as the limited possible interactions on desktop and mobile devices. The proposed interfaces aim to capitalize the abundance of space and possibilities of new interactions provided by the VR setting to combat these issues.

B.1 Stretchable interface

The stretchable timeline design addresses the granularity issue by changing a fundamental part of the interface. In state of the art interfaces the timeline itself is fixed, and the knob (the slider) can be moved along the timeline. With the stretchable timeline, the knob instead is fixed below the center of the user's gaze. The timeline moves through the knob as the video plays. As a result, the effects of having a long timeline are reduced. The user can still focus on the center of the screen to navigate the video. Of course, the user might occasionally glance at the ends of the video to get their bearings, but it is expected to be less frequent.

The user can grab the timeline with one controller and move it through the knob to advance or reverse the video. This gesture can be performed anywhere on the timeline. This solves the issue of large timelines requiring uncomfortable gestures to operate it. The user can simply perform the gesture close to the knob.

As a way to improve control and reduce the issue of low granularity with longer videos, the timeline can be stretched. The user can grab the handles at both ends of the timeline with the grip buttons, and move their hands closer together or further away. These gestures respectively shorten and lengthen the timeline. This does not affect video duration. The position of the knob on the timeline will be locked during scaling operations to avoid accidentally moving the temporal position of the video. As soon as the grip button is released the timeline's new length is locked and is open for scrubbing again.

A major advantage of this approach is the complete control over the granularity. As a result, moving through the video can be adapted to the situation. If a user wants to skip through a video rapidly, they can shrink the timeline. If they are interested in several specific scenes, they can reach these scenes with great precision by stretching the timeline. When scaling the timeline it is possible to run into the issues described above. For example, when the timeline has been made quite large it might require either a very big gesture to reduce it to a smaller size, or even multiple normal gestures. In addition, having a 3D timeline means that it should be within the user's reach in order for them to be able to interact with it. An interface with buttons that have to be selected with a ray cast from the controllers are more flexible in terms of environment. They can still be easily operated when sitting at a desk. For a 3D interface, if the desk (or other physical object) intersects with the timeline in VR, it becomes difficult or impossible to properly operate.

In addition to scaling the timeline as a whole, the user can also stretch it locally. This can be done by grabbing the timeline at any position between the begin and the end. The rest of the interaction is the same as the normal scaling interaction. When the user approaches the timeline with their controllers, a phantom handle will appear. On pressing the grip button, the handle will materialize and the user will be able to stretch the part between the two handles freely. Once the grip buttons are released, the handles will stay in place and the timeline cements again. To return the timeline to its normal state, the user simply has to grab the handles that remained visible, and move their hands together. This will reset the timeline to its shape prior to the local stretch.

This can be done as many times as desired, and multiple local stretches can be created. Video playback will continue normally when passing across a local stretch, although the speed at which the timeline moves is temporarily increased. Scrubbing speed will be lower when passing over a local stretch, as a result of the longer distance to cover.

Local stretches are especially useful in longer videos when a user wants to reach a specific spot. For example: the user is watching a movie and wants to skip to a scene which they know to be located somewhere around the middle of the movie. They can then simply stretch the area of the timeline where they expect the scene to start, and scrub through the stretched part with great precision.

A beneficial side effect of local stretches is that they could also function as temporary bookmarks or points of recognition. By stretching the video, a position is marked with the handles. This can helpful when navigating the video. A potential issue could be that the timeline becomes to cluttered with local stretches to the point where it becomes difficult to reset the timeline to its original state. Additionally, multiple local stretches of different durations can be stretched to be similar in length. This does not mean they have the same level of granularity, though it could be interpreted as such. This could be solved by colour-coding (or otherwise visually marking) the stretched parts based on local granularity.

B.2 Clock-based interface

The clock interface takes a different approach to solve the granularity and timeline size issues. As opposed to the state of the art, a clock has a fixed level of granularity. Each degree travelled by for example the minutes hand has the same meaning every time. The three hands of the clock can be manipulated directly by the user. This allows for temporal navigation with three different levels of granularity. Instead of representing the beginning and end of the navigatable time with the endpoints of a timeline, the clock highlights the navigatable area on its most relevant level. This means that if a video is shorter than 60 seconds, the area highlighted will be coloured similar to the seconds hand. Similarly, if a video is between 60 minutes and 60 seconds, the relevant area in minutes will be marked. This does introduce a weakness into the system for videos of certain length; the length of videos that are just over a 60 second/minute threshold are less clearly marked. This is the result of having the same distance always represent the same amount of time. In addition it is no longer possible to immediately skip to a specific point in the video with a simple interaction. In a state of the art interface this is often possible by selecting a specific point on the timeline directly.

B.3 Appearance

To make sure the interfaces are easy to use, interactable elements will communicate their importance to the user (see section 10.4). For example, a handle will be easily recognizable as such. In addition, interactable elements will have a bright colour to indicate their

importance, as well as provide sufficient contrast with the video in the background. Whenever the user approaches an interactable element with a controller, their virtual hand will snap to that element and indicate that an action can be performed here. The snapping position of the hands are placed in such a manner that they obstruct the view as little as possible. In addition, it will be clear at all times which temporal position is currently selected. For example, when manipulating the clock-based interface, the handles end points are never obstructed by the virtual hand. For the stretchable interface this means the potential handles that appear on the timeline have a clear edge which is used as the point of interaction. This is done by putting the point of interaction at the inner edge of the hand, next to the thumb. As the user will generally be looking at their hands from the inside, this is the most sensible spot. This is communicated to the user by having a thin circle around the timeline attached to the virtual hand.

B.4 Preview miniature

In both interface designs, a miniature is placed above the timeline or clock. Similar to state of the art video players (especially 2D players) this miniature serves as a preview of the video at the specific time currently selected. This miniature is updated in real time, as opposed to the full 360° video, which is only updated after the interaction (user dragging the timeline or manipulating the hands of the clock) is finished. Updating the full video screen while scrolling is expected to induce cybersickness, and is therefore done sparingly. As soon as the user starts manipulating the timeline, the video blurs slightly, and once the interaction is finished fades to black before fading into the new location in time.

The miniature will be in the form of an equirectangular projection, similar to the smal rectangular minature used in desktop or mobiel video players. The user will experience some distortion as a sphere (which is essentially what a 360° video is) does not map perfectly to an equirectangular projection. An alternative would be to use a miniature, spherical version of the actual video, which potentially reduces distortion. However, that is outside of the scope of this study.

For the stretchable interface, the miniature is visible above the knob, but only if the user is actively manipulating the timeline to avoid obstructing relevant parts of the video. Similarly, in the clock-based interface the miniature will also be hidden when the hands are not being manipulated.

Figure 22: The information sheet participants signed before taking part in the study.

Information sheet

Title: 3D Interfaces for 360° VR Video Players
Date and location: __-11-2022, Utrecht

Goal of the research

Thank you for expressing interest in participating in this study on virtual reality (VR) video player interfaces for 360° videos.

360° videos are videos recorded in all directions. You probably know this from the Google Maps Street view function on a desktop PC or your phone. But these 360° videos can also be viewed in VR, where the video is then projected all around you, and you can look around by moving your head or rotating yourself.

The purpose of this research is to evaluate different ways to navigate and scroll through these types of videos in VR.

Procedure

In this study you will be putting on a VR headset and use two VR controllers. During the experiment you are presented three different interface designs. The following process will be repeated three times, once for each of the designs.

You will first receive an introduction on how to operate the interface. This allows you to get acquainted with the interface and play around with it. Next, you will be asked to perform four tasks with that interface. Then, you can take the headset off and answer a brief questionnaire on your experience with that interface.

After doing this for all three interfaces, you are asked to fill out a final questionnaire comparing them with each other. This entire process is expected to last about half an hour.

Researcher

This research is conducted by Jorian Berkhout, a student at Utrecht University, for a Human Computer Interaction (HCI) master thesis.

If you would like to contact the researcher for concerns, questions, or comments you can do so by sending an e-mail to j.l.berkhout@students.uu.nl.

For any (other) concerns regarding the study or the researcher that you do not wish to discuss with the researcher themselves, please contact the supervisor of this thesis project, Wolfgang Hürst huerst@uu.nl.

Potential risks & important information

- Using a VR headset can result in cybersickness. Symptoms of cybersickness include nausea, headaches, dizziness, or other physical discomfort. These symptoms often rapidly decline once the headset is taken off. If you experience one or more of these symptoms, or any other form of discomfort during the experiment, please inform the researcher immediately. You can take a break or withdraw your participation without having to provide a reason or facing any negative consequence.

- The headset and controllers are cleaned after each participant, as proper hygiene is even more important than normal due to the COVID-19 pandemic. You are free to clean the equipment again yourself if desired. Simply ask the researcher for the cleaning supplies.
- There are no judicial or economical risks to participating in this study. You do not have to answer questions that you do not want to answer. Your participation is completely voluntary, and you are allowed to quit at any moment you like. If you choose to quit the study, your results and information will be deleted and not used for the study.
- The videos used in this study were selected with great care to be suitable for everyone. If any material makes you uncomfortable, please indicate to the researcher that you do not want to use this video. The researcher will then provide an alternative video.

Confidentiality of data processing

Your privacy is and will be protected according to Utrecht University and GDPR rules and guidelines. No confidential information or personal data will be disclosed or publicized in any way that will be traceable to you.

- You will not be asked to provide any information that is not relevant to this research.

- The results of the experiment will only locally be saved in documents on a password-protected computer that is only accessible to the researchers and supervisors involved in this study. Once the study is completed the data will be transferred to secure university servers.

- If you want to gain insights and see exactly what data we have saved about you, you can request this by sending an email to the researcher (see *researcher*). Requesting your data will not be possible anymore after all the data has been anonymized, as the research will have no way to know which data belongs to you. The original data will be destroyed after the data has been anonymized.

- For details of our legal basis for using personal data, the rights you have over your data, and the contact details of our Data Protection Officer for any data protection queries, please see our privacy information at https://www.uu.nl/en/organisation/privacy-statement-utrecht-university.

Compensation

As reward for participating in the study participants are offered a snack and coffee or tea. No other (monetary) compensation is provided apart from the (not legally binding) eternal gratitude of the researcher.

Figure 23: The consent form participants signed before taking part in the study.

Consent form

Statement of consent:
The purpose of this declaration is to establish the terms of my participation in this study. By signing, I consent that I am properly informed about this study, the way the data is collected, stored, and processed, and any foreseeable potential risks that are attached to my participation.

Please tick all the boxes and sign below. To participate you must agree to all statements.

☐ I agree to participate in the research on 360° video interfaces VR (as specified in the information sheet).

☐ I understand what this study is about and have been provided sufficient information. I am aware that I can ask any questions regarding the study at any point and have had sufficient opportunity to do so.

☐ I consent to providing information relevant to the research.

☐ I consent to my data being anonymized and stored safely according to Utrecht University and GDPR regulations (please refer to the privacy statement of Utrecht University here: https://www.uu.nl/en/organisation/privacy-statement-utrecht-university).

☐ I understand that using a VR headset can result in cybersickness, including symptoms such as nausea, headaches, or dizziness.

☐ I am aware that I can take a break at any point during the study, if any of the symptoms occur, or for any other reason.

☐ I agree to indicate when I experience physical or mental discomfort because of the study (direct or indirect) so the researcher can pause or abort the experiment.

☐ I am aware that I can withdraw from the research with no consequence at any time without having to provide a reason.

☐ I have read and understood the information sheet and the informed consent form. All my questions have been answered satisfactory and I agree to participate voluntarily.

Signed

_____	_____
Researcher: Jorian Berkhout	Participant: _____
__-11-2022, Utrecht	__-11-2022, Utrecht

Figure 24: The e-mail send to participants to confirm their appointments.

Hi **<Participant>**,

Thank you for participating in my experiment!
If that suits you, I'd like to schedule you for **<Day> <Date> <Month> at <Time>.**

The experiment takes place in the Human Centered Computing Lab (Caroline Bleeker building 1.09). Below are instructions on how to get there from the BBG building.
If you cannot find the way, please contact me via WhatsApp (+[redacted]).

No need to reply to this email if this time works for you! If you would like to reschedule or cancel, just send me an email!

See you there!
Jorian Berkhout

How to get to the Human Centered Computing Lab:

1. Go to the first floor of the Buys Ballot Building (BBG)
2. Walk to the west end of the building (see image 1)
3. Go through the double doors into the connecting bridge between BBG and the Bleeker building (see image 2)
4. Continue walking until you see a blue door on your left (see image 3)
5. Enter this door, and you should find yourself on an elevated catwalk above a room with a lot of machines
6. Keep walking on the catwalk until and check the doors on the left until you arrive at room 109. Either wave at me through the window or knock on the door so I can let you in!

As mentioned above, please contact me via WhatsApp if you have any trouble finding the lab!

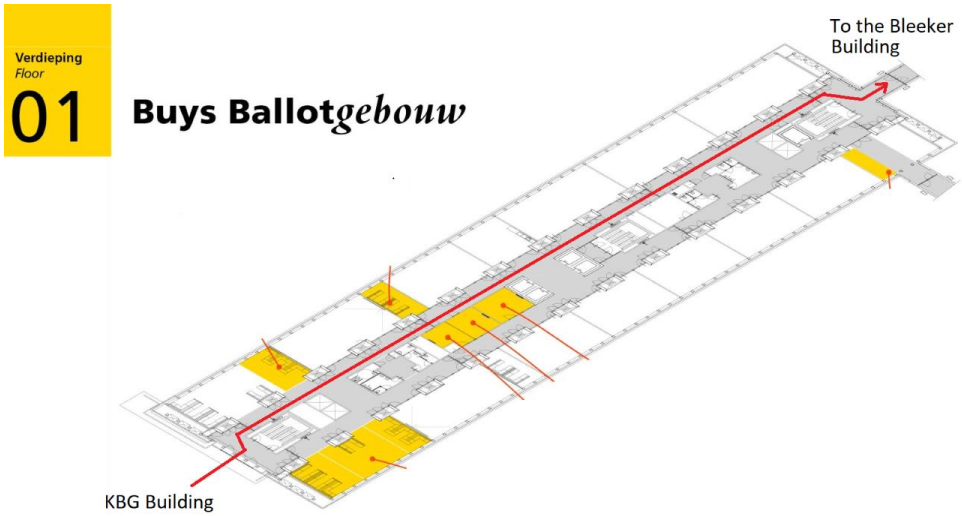

image 1

image 2

image 3

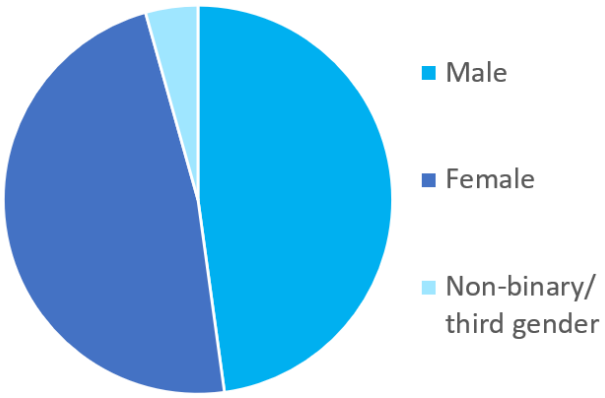

Figure 25: Gender of participants (23 total, 11 male, 11 female, 1 non-binary/third gender).

Figure 26: SUS-questionnaire exported from Qualtrics.

Utrecht University

User info

FILLED IN BY RESEARCHER:
What is your participant number?

FILLED IN BY RESEARCHER:
What interface did you just use?

- ◯ Flat Timeline (2D)
- ◯ Clock Interface (3D)
- ◯ Stretchable Timeline (3D)

Cybersickness

Did you experience any of the following symptoms at any point during the study?

	No	A little	Yes
Nausea	○	○	○
Headache	○	○	○
Dizzyness	○	○	○

Did you experience any other feelings of discomfort? If so, which?

SUS-Questionnaire

Please answer the following questions about the interface you just used in VR. You will be asked to answer these questions after each interface.

Please indicate how much you agree with each of the following statements

	Strongly Disagree	Somewhat Disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
1. I think that I would like to use this system frequently.	○	○	○	○	○
2. I found the system unnecessarily complex.	○	○	○	○	○
3. I thought the system was easy to use.	○	○	○	○	○
4. I think that I would need the support of a technical person to be able to use this system.	○	○	○	○	○
5. I found the various functions in this system were well integrated.	○	○	○	○	○

Please indicate how much you agree with each of the following statements

	Strongly Disagree	Somewhat Disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
6. I thought there was too much inconsistency in this system.	○	○	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	○	○
8. I found the system very cumbersome to use.	○	○	○	○	○
9. I felt very confident using the system.	○	○	○	○	○
10. I needed to learn a lot of things before I could get going with this system.	○	○	○	○	○

Do you have any comments on this interface you would like to share?

Powered by Qualtrics

Figure 27: Final questionnaire exported from Qualtrics.

Utrecht University

Block 2

FILLED IN BY RESEARCHER: What is your participant number?

Rating

Thank you for participating in the experiment. Please answer the following questions about the different interfaces!
For the questions below, please drag the different timelines to create your ranking.

Please rank the interfaces according to how **efficient** they felt to use

Clock Interface

Standard Timeline

Stretchable Timeline

Please rank the timelines according to how **fun to use** you thought they were

Standard Timeline

Stretchable Timeline

Clock Interface

Please rank the timelines a final time on which you **overall think is best!**

Clock Interface

Stretchable Timeline

Standard Timeline

Demographics

Finally, please answer some questions about you!

How would you describe your gender?

- ◯ Male
- ◯ Female
- ◯ Non-binary / third gender
- ◯ Prefer not to say

What age are you?

- ◯ 18-21
- ◯ 22-25
- ◯ 26-29
- ◯ 30-33
- ◯ 34+
- ◯ Prefer not to say

Which statement describes your experience with VR (prior to this study) best?

- ◯ I had never used a VR headset before
- ◯ I used a VR headset once or a couple of times
- ◯ I use a VR headset occasionally (e.g., about once a month)
- ◯ I use a VR headset frequently (e.g., about every week)