A Collaborative Previsualization Tool for Filmmaking in Virtual Reality

DUI ARDAL
ABSTRACT

Previsualization is a process within pre-production of filmmaking where filmmakers can visually plan specific scenes with camera works, lighting, character movements, etc. We developed and assess a prototype for previsualization in virtual reality for collaborative purposes where multiple filmmakers can be present in a virtual environment to share a creative work experience, remotely. The costs of computer graphics-based effects are substantial within film production, using previsualization, these scenes can be planned in detail to reduce the amount of work put on effects in the later production phase. By performing a within-group study on 20 filmmakers, our findings show that the use of virtual reality for distributed, collaborative previsualization processes is useful for real-life pre-production purposes. These results provide insights on how to best design collaborative, virtual reality-applications used for remote work, and highlights the pitfalls of certain design choices.
SAMMANFATTNING

A Collaborative Previsualization Tool for Filmmaking in Virtual Reality

Dui Ardal
Royal Institute of Technology
Stockholm, Sweden
dui@kth.se

ABSTRACT
Previsualization is a process within pre-production of filmmaking where filmmakers can visually plan specific scenes with camera works, lighting, character movements, etc. We developed and assess a prototype for previsualization in virtual reality for collaborative purposes where multiple filmmakers can be present in a virtual environment to share a creative work experience, remotely. The costs of computer graphics-based effects are substantial within film production, using previsualization, these scenes can be planned in detail to reduce the amount of work put on effects in the later production phase. By performing a within-group study on 20 filmmakers, our findings show that the use of virtual reality for distributed, collaborative previsualization processes is useful for real-life pre-production purposes. These results provide insights on how to best design collaborative, virtual reality-applications used for remote work, and highlights the pitfalls of certain design choices.

CCS CONCEPTS
• Human-centered computing → Interface design prototyping; Collaborative content creation.

KEYWORDS
Virtual reality (VR); collaborative virtual space; film-production; distributed workspaces; previsualization

1 INTRODUCTION
The use of computer graphics-based (CG) technologies for providing visual effects within filmmaking is a well-established practice. The costs of these technologies are substantial, therefore it is important to plan shots that are intense with visual effects in detail [16]. With the recent advancements in virtual reality (VR) technologies, the film industry sees a shift where planning for different scenes and takes can be done in a more immersive and tangible way. VR enables a user to be present in virtual environments, resembling specific locations and thus helping filmmakers to find new ways for them to work creatively. In this thesis, we study the use of VR for previsualization (or previs) in a collaborative setting. Previsualization is a collaborative process of planning scenes and shots within pre-production stages of filmmaking. Traditionally, this process has been performed with drawings, concept images, sketches, etc. [17], and it is not until recently that previs has been performed with 3D animation tools and software for modelling scenes and test different camera works, lighting, etc. This project looks into this process by putting filmmakers into an actual film scene with the use of VR technologies.

Pre-production is a phase that is focused around planning to optimize the later production phase. In films with complicated shots or with large amounts of visual effects, this planning can be harder to visualize and the costs for producing visual effects and CG-based effects is considerable. 3D animation tools such as Maya1 or Blender2 can provide creators with tools for animating characters, environment and cameras, and test different lighting, etc., hence providing resources for executing the planning process. However, these tools do not include the interactive layer that is an integrated part of game engines.

Game engines have been used for providing real-time experiences for other scenarios than pure gaming, by both including animation and 3D modelling from traditional 3D animation tools and interactive practices from gaming creation. These tools can also be an integrated part of previsualization, where filmmakers can immerse in film scenes and test different camera works, character movements, etc. in real-time [23]. On the other hand, these 3D animation and game engines requires a high level of skill sets for modelling, animating and scripting. Hence, filmmakers might not have the required skill sets to perform previs [22]. VR has the affordances of providing immersive experiences of being in virtual environments. Similarly, it can mimic real-life situations in various scenarios which induces interaction principles that people are accustomed to, e.g. picking up objects using hands or walking around in a physical space that is reflected in a virtual environment.

By making use of these affordances, a tool was developed for providing an immersive and collaborative previs environment where filmmakers can create, discuss and validate different takes, shots and entire scenes. A novel interface

1https://www.autodesk.com/products/maya/overview
2https://www.blender.org/
where the intended users do not need advanced computer literacy to understand the functionality. The tool was built with Monocular®, a company working with previs for filmmakers. They provided a foundation for the proposed tool that was further developed for the purpose of this thesis. The main software used was Unity3D game engine⁴, as a means to realize the interactive tool for real-time previsualization in VR.

Working remotely from different geographical locations is increasing in the film industry and with the advent of extended reality (XR) tools, collaboration in remote places can be achieved with immersion and sense of being present with virtual environments and colleagues. By developing a multi-user and collaborative tool for previs in virtual reality, this project provides insights on how to best interact with other users in VR and be able to conduct previsualization of film scenes as efficiently as possible. To realize this, a within-group study was conducted, where the participants (professional filmmakers, primarily from the Stockholm Academy of Dramatic Arts) performed previs sessions in both solo and collaborative setups.

Communication and co-presence are fundamental for providing an effective experience for collaborating. Additionally, providing a proper foundation for grounding is important to enhance a collaborative environment. As VR technologies are still at an early stage when it comes to representing other users and their communicative efforts, we hope to contribute to this research with this project. Through the literature review, collaborative VR and working environments for the field of filmmaking and its use-cases were found not be thoroughly represented. Therefore, the main research question has been defined as:

**How to best design collaborative tools for interactive distributed previsualization within film production in virtual environments?** By answering this question, this thesis can provide knowledge for how filmmakers can make use of virtual technologies for enhancing their creative process, while at the same time providing the possibility to work remotely.

What makes the film industry an interesting topic for collaboration in virtual environments is because of the clear distinction between roles and separate responsibilities. Also, the process is collaborative in nature and requires strong communication skills.

## 2 RELATED WORK

### Communication

In face-to-face (FTF) interaction, communication can be achieved in a delay-free and multi-functional way [7]. Because of this, FTF communication has been more efficient than computer-mediated communication (CMC) for conveying messages. Furthermore, by possessing more visual cues, a shared context and a natural way of grounding communication, collaborative aspects of FTF communication have also proven to work more effectively [4, 11, 12]. In [7], Clark and Brennan claim that grounding, i.e. finding common ground, is important for how communication works effectively. Grounding can take different forms based on medium and purpose, e.g. a telephone call requires a different grounding in contrast to FTF conversation and therefore the purpose of the communication directly affects the medium as well. A telephone call is not ideal for communicating visual aspects. Thus, the costs of constraints can affect formulation, production, and understanding, to mention a few that Clark and Brennan declares. Furthermore, communication is a collective activity where participants in a communicative context need to coordinate their tasks for finding proper grounding and thus a foundation for communication [7].

### Collaboration

Collaboration calls for similar grounding as for general communication in order to work effectively. The importance of shared visual space, context, and speech is evident in earlier research [11, 13, 14], where the context and understanding of how to collaborate and what to collaborate on can be made clear. In [11], Fussel, et al., accounts for four types of visual information that is important for grounding in collaborative contexts: participants’ heads and faces, participants’ bodies and actions, shared task objects, and shared work context. Similarly, having a shared visual space in collaborative work environments makes participants’ more likely to let their actions speak for themselves instead of having to explain verbally at the same time [14], something that provides a more natural communication, as described in [7]. Fussel, et al., also emphasizes that remote collaboration can be supported by providing grounding and facilitating visual information [11]. A further approach of understanding how communication is constituted within collaborative environments is from [5], where the authors claim that the need for shared mental models is important for how effective collaboration can be achieved, which is also described in [29]. Furthermore, specialized roles in collaboration can be a resource, as well as the knowledge-in-common of participants [5, 18].

### Collaboration through technology

Technological solutions for remote collaboration have taken many different forms: letters, telephone, video-conference, etc. Similarly, CMC for collaboration has seen many different solutions. For example, video conferencing systems with participants visible on a screen for creating co-presence [19], head-mounted displays for sharing visual space [12, 18], and
virtual software [6, 27] are some of the different applications that has showed potential in creating efficient collaboration. In [18], a "helper" and "worker" scenario was tested for remote collaboration with the use of tablets and head-mounted displays (HMD). It was made clear that there are advantages to both these setups, where using HMD resulted in faster task completion. In a similar research topic, Bateman et al. measured how people can collaborate in distributed locations via the use of mobile phones and stationary computers. The results showed that the mobility of these devices and network connections that can send and receive data in real-time provides functional collaborative environments. The authors also claims that the lack of developed technology can hinder good collaboration. Poor field-of-vision in interfaces and lack of asymmetry in controls are two examples of how the design of technical solutions could have been standing in the way of usable interfaces for interactive collaboration [3]. This is something that VR technologies could possibly be better suited for, as the immersive qualities of virtual headsets and its controls are big enough to facilitate collaborative environments.

In recent years, the rise of XR for collaborative purposes has increased with both mixed-and virtual reality (MR/VR) applications for collaborating in remote locations now being possible due to increase in network speeds and bandwidth. In addition, the advancements of XR technologies has led to decreased prices where more consumers can have access to VR technologies at home [10, 20, 25].

Virtual reality

Attaining grounding in remote communication is an important aspect of achieving co-presence in a virtual 3D environment [14, 28]. With social VR as a relatively new concept within social media, VR is being seen as a way of potentially achieving co-presence with other people [25]. However, there are problems remaining with VR for fully achieving co-presence. In [1], Anthes et al., raises a number of issues on the state of VR technologies. It is clear that representation of users still is a problem as a result of lack of information on facial expressions, eye-gazing, and body movement. However, recent advancements is attempting to tackle these problems with the use of eye-tracking inside VR headsets, gloves that can represent fingers, etc. Thus, user representation is likely to improve in the future [20, 26].

Within multi-user VR services, Elvezio, et al., proposed two scenarios: symmetric and asymmetric. Symmetric use-cases for multi-user VR highlights users who work together on the same terms, while asymmetric use-cases emphasizes how certain users can be specialized in tasks, or how some users are more experienced than other users, e.g. guides for completing tasks [9]. In regard to collaboration in VR, finding grounding can be accomplished by the use of avatars and biometric signals, as described in [21]. The affordances of elements in a VR environment needs to be carefully decided upon, when designing for VR. In [8], Ellis makes clear that affordances in VR should mimic the real world as much as possible, as the affordances of real artifacts are expected to have the same affordances in the virtual environment. Furthermore, in the virtual environment, extended affordances can be developed, as the virtual environment can in its nature facilitate this [8].

Previsualization

Previsualization is an integrated part of the pre-production stage within filmmaking. It is a process where filmmakers are able to "see" scenes in a film that are otherwise difficult to visualize. Different areas of use-cases are e.g. scenes with visual effects, scenes dependent on green screen, or where real actors are to interact with virtual characters. Previs tools in VR have been attempted before. In [22], the authors developed a prototype used for previs in VR that showed great promise in using VR technologies for certain fields that can make use of technology without the need for computer literacy [22]. Similarly, a VR prototype for previs is presented in [28]. The experiments showed that making use of a traditional video-game controller could however be less intuitive than the 3D gesture-based controllers used in HTC Vive® and Oculus Rift® [28]. Lastly, what is lacking in the related literature is to explore multi-user and collaborative aspects of previs solutions for VR, which is what this thesis will be further investigating.

3 PREVIS VIRTUAL REALITY INTERFACE

The tool used during this project was developed in Unity3D game engine and was made for the HTC Vive VR system. The hardware included a head-mounted display (HMD), two hand-controllers and sensors. The use of this system allows for six degrees-of-freedom (DOF) where users can move freely within a certain area, defined by the sensors. Networking connectivity for multi-user purposes was achieved with the use of Photon Networking7, allowing users to interact and make use of voice chat. The tool is designed for filmmakers, where different roles plays an important part in how pre-production is being executed [16]. The different roles typically common in pre-production varies between different films. A common scenario typically includes a director, which is the creative lead of a film; a photographer (or cinematographer), which is in charge of the visual image of the film; and a scenographer, which is mainly focused on the environment, objects

---

7https://www.photonengine.com/
and mood of a film [17]. However, for the evaluation of collaboration between filmmakers, the focus is limited to the collaboration between the director and the photographer.

Scenario
The scenario is based upon an actual animated short-movie that is in the process of being made. By letting that material be used for this experiment, a clear scenario with both a virtual environment and script for a final scene existed and tasks could be derived from those assets. The story takes place in a small house on a remote island where a solitary man plays the main character.

User interface
The actual user interface allows users to see their own controllers in the virtual environment. On the left hand for each user, a menu is floating above (see Figure 1). The right-hand controller does not have any additional menu graphics added to it. Each controller have four types of interaction elements: two grip buttons, one trigger button, one application menu button, and a touchpad. The grip button is used for pulling oneself through the virtual environment; the trigger button is used for interaction with objects and handling virtual elements; the application menu button is interactable on the right-hand controller, where it toggles recordings; and the touchpad toggles a laser pointer that can be used both for referencing (e.g. showing other users where to look) or choosing interactable elements on the left-hand menu. To choose an element on the left-hand menu, the user points with the laser pointer and presses the trigger button. By pressing the grip buttons on both controllers at the same time, the user can scale the virtual environment and get a miniature setting where the users can puppeteer the characters and cameras. Likewise, the user can maximize the environment for more precision.

Spawning a camera is done with the left-hand menu. When a user clicks the “Add new camera”-button, a camera is spawned in front of the user. The camera has a screen attached to the top of it, providing visual feedback of what the camera “sees”. Grabbing the camera opens up a new menu on the hand that grabbed it. This menu includes the option to change lenses, ranging from 12 mm to 180 mm. A character can also be spawned. Much like with the camera, when a user clicks on the button to add the character (“Add new character”), the character is spawned right in front of the user. The user can grab the character by the feet, which also opens up a new menu on the hand that grabbed the character. This menu includes a button for snapping the character to be on the ground at all times. Therefore, when a character is grabbed and the button for snapping it to the ground is clicked, a user can hold the character like a puppet. When a user goes back and forth with the hand that has grabbed the character, automatic walk cycles are activated which are tuned to the pace of the character (still, walking, and running). Thus, the characters does not require full body motion capture.

Figure 1: Menu on left-hand controller.

The right part of the left-hand menu is the part for interacting with recordings. There is a rewind-button, play/pause button, and a record button. Underneath these buttons is a timeline that shows where in the recorded take the scene is. This timeline can be interacted with via a slider. Grabbing the slider while going left/right will make the recorded animiations of both character and camera go back and forth in its animated timeline. Furthermore, on the sidebar menu, the button “Go to cinema room” teleports the user to another room in the virtual environment.

On the right controller, clicking the application menu button will toggle recording of all movements of character and cameras. By pressing the record button, moving the character, and then clicking the record button again is how the user can record different animation events.

Lastly, the cinema room is an empty room with a big screen, where the user can see all the cameras he/she has spawned, providing a fixed visual feedback of the cameras, compared to how the camera in the scene is animated therefore the visual feedback is also moving along with the camera. The cinema room also has the interactable timeline, play-button and rewind timeline-button. The cinema room was created to allow participants joint visualization of the recorded cameras, providing a room for discussion.

Roles
The first implementation from Monocular was designed for one person. The implementation for this thesis was extended for multi-user purposes. This collaborative design is made for two roles: a director and a photographer. The visual feedback of other participants were represented with a circular head and two hands, and were updated in real-time.

Photographer. The photographer has the same interface as shown in Figure 1. This role is the one that has control over the timeline and recording of the character and cameras.
Furthermore, this user can grab the character and cameras, and go to the cinema room.

**Director.** This role has a more limited interface where the entire right part of the left-hand menu (seen in Figure 1) is removed, as is the record-button on the right-hand controller. The main reason for this was to coerce the roles into a collaborative mode where communication needs to be performed to complete the task. The director can grab the character and cameras and animate these as soon as the photographer toggles the record mode. The director can also go the cinema room.

4 **METHOD**

To measure the effects of collaborative work in the virtual environment, experiments on the intended user segment was performed for evaluating the prototype. The evaluation was based on a within-subject study in which participants conducted a previs session in two different settings: one in a solo condition, and the other in a collaborative condition. The reason for this was partly due to the amount of participants and to minimize individual differences. By also letting participants try out each of the settings in different orders, we could reduce the effects and bias from having performed the study in a specific order.

**Procedure**

Before each previs session, the participants received training in a supervised setting. The training was divided into two parts: the first part included a showcase of the tool, its controls and functions. Afterwards, each participant was given a station with computer, VR-headset and controllers for practicing 10-15 minutes alone in the tool. The participants could then ask questions and receive recommendations on how to use the tool. Afterwards, they read a short script explaining the scene that they should produce in the previs interface. This was done alone before the solo setup and together with another participant before the collaborative setup. After reading the script, the participants wrote down a short statement about their idea on how they would plan and shoot the scene, to be able to compare with their final result afterwards. The participants were then given 15 minutes for the current previs-session. After this, the groups switched setup, read the script for the second task and then performed the second previs-session. After each session, the participants filled out questionnaires. The setting for this study can be seen in Table 1.

The main task for the participants was to make a single-shot scene that corresponds to the script that the participants read. No movie-editing was therefore involved for this experiment. The single-shot was supposed to be between 20-30 seconds long and the participants could use how many camera takes as they pleased, they had to choose one take as their final end-resulting clip, however.

<table>
<thead>
<tr>
<th>Table 1: Experimental setup</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1</strong></td>
</tr>
<tr>
<td><strong>Training</strong></td>
</tr>
<tr>
<td>Solo task 1</td>
</tr>
<tr>
<td>Collab task 2</td>
</tr>
</tbody>
</table>

**Participants**

For the experiments, 20 participants were recruited. The participants were all specialists. One person had been active in the film industry for 1-2 years, one person for 2-3 years, four persons for 3-5 years, seven persons for 5-10 years and seven people for more than 10 years. The ages differed between 18-59 years, with 10 female, 8 male, and 2 who preferred not to disclose.

**Measures**

After each experiment, the participants filled out a questionnaire asking them about presence in the virtual environment, the co-presence of other participants (if collaborative condition), their experience of performance and behaviour, and their feeling of their part in the process and the resulting clip.

The questions on presence and co-presence were collected from [15] where Gerhard et al. have extended questions from [31] on involvement and immersion with questions regarding multi-user virtual environments. For the purpose of this thesis, parts of the questions from Gerhard et al. questionnaire has been extracted. Questions on presence and co-presence are important for this project as they can describe how the participants felt about their task in relation to their experienced presence of being in a virtual environment as well as other participants. Immersion has to do with experiencing that a user is present in the VE and is highly connected to the form of the system, presence is similar to this, yet is more of a 'response' to the VE system that is achieved at a certain level of immersion. The involvement is more related to the content of the VE, where a user is interested and can lose track of time [30].

After all experiments, one questionnaire included questions about the participants’ personality, experience in VR, game controllers and their working experience. The questions regarding personality traits was collected from The Big Five personality traits [2]. These factors could be of importance to how the participants experienced their own task performance and collaboration and was thus part of the evaluation. In the question about each participant’s experience in using VR we expected to find differences between participants with
experience compared to those with little or none experience. Likewise, with the question concerning the participant’s experience in using controllers typically used in games, we expected to find differences between experience and results.

5 RESULTS

In this section, we analyze the results of our three questionnaires. Firstly, we analyze whether the participants’ personality traits might have effected the responses, using The Big Five personality traits [2]. Secondly, we measure the differences between the two different experiment conditions by analyzing the responses from the questionnaires answered directly after each experiment. The within-subject setting demanded that each participant performed each previs condition, hence these could be compared. Lastly, we measure the participants’ opinions about the two conditions and present both statistical analysis and their subjective opinions from the qualitative parts of the assessment.

The independent variables are coded into the setups that the participants performed the experiments on: solo and collaborative. Furthermore, the order in which the participants performed both setups showed no effect on the statistical results. Thus, it is not included as an independent variable. However, in the qualitative section, the order was shown to have some effect: several participants mentioned that they could handle the tool better in the second experiment. This will be further discussed in section 6.

Figure 2: Two participants working in the tool.

Big Five Personality Traits

To first determine which variables correlated and therefore understanding the data, bivariate correlations were performed between the questionnaire responses and the personality traits. The correlations showed significance between the five personality traits and the questionnaire responses. It showed that the correlation between Extraversion and Were you involved in the experimental task to the extent that you lost track of time? was significant (Pearson’s correlation: 0.372, p < 0.05). Likewise, Extraversion together with Which setup did you prefer? and Extraversion together with

| Table 2: Significant correlations, personality traits and questionnaire results. * p < 0.05, ** p < 0.01 |
|-----------------------------------------------|----------------|-----------------|
| How compelling was your sense of being present in a virtual world? | Pearson Corr. | Sig. | Extraversion | Conscientiousness | Openness to exp. |
| -0.331 | 0.442 | 0.018 ** |
| How difficult was it to perform the task? | Pearson Corr. | Sig. | Extraversion | Conscientiousness | Openness to exp. |
| -0.353 | 0.473 | 0.004 ** |
| How would you rate your performance in this task? | Pearson Corr. | Sig. | Extraversion | Conscientiousness | Openness to exp. |
| 0.221 | 0.473 | 0.009 ** |
| Were you involved in the experimental task to the extent that you lost track of time? | Pearson Corr. | Sig. | Extraversion | Conscientiousness | Openness to exp. |
| 0.042 | 0.473 | 0.005 ** |
| I was aware of the actions of other participants | Pearson Corr. | Sig. | Extraversion | Conscientiousness | Openness to exp. |
| 0.599 | 0.473 | 0.005 ** |
| Which setup was more enjoyable? | Pearson Corr. | Sig. | Extraversion | Conscientiousness | Openness to exp. |
| 0.293 | 0.473 | 0.005 ** |
| -0.128 | 0.473 | 0.009 ** |

Within-Subject Analysis

By performing experiments as within-group, using a repeated measures ANOVA showed two significant results (see Table 3). In the question How well could you concentrate on the assigned task rather than on the mechanisms used to Which setup was more enjoyable? showed a significant result (Pearson’s correlation: 0.442, p < 0.01, Pearson’s correlation: 0.473, p < 0.01). Furthermore, Openness to new experiences correlated on a significant level with How compelling was your sense of being present in a virtual world? (Pearson’s correlation: 0.331, p < 0.05). How difficult was it to perform the task? (Pearson’s correlation: 0.353, p < 0.05), and I was aware of the actions of other participants (Pearson’s correlation: 0.599, p < 0.01).
perform these?, solo setup received a higher rating than collaborative \((F(1, 17.875) = 5.979, p < 0.05)\). In the same way, the solo setup received a higher rating in How would you rate your performance in this task? \((F(1, 19) = 6.066, p < 0.05)\).

The repeated measures ANOVA showed results that were not hypothesized. The initial hypothesis was that the filmmakers would prefer the collaborative setup, as it is more relatable to a real-life scenario where filmmakers collaborate in the production of a film.

The mean values for each question from the questionnaires directly after each experiment can be seen in Figure 3.

The mean values for each questionnaire question in both setups.

### Table 3: Significant repeated measures ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>How well could you concentrate on the assigned task rather than on the mechanism used to perform these? * p &lt; 0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>1</td>
<td>5.625</td>
<td>5.979</td>
<td>0.024*</td>
</tr>
<tr>
<td>Error</td>
<td>19</td>
<td>0.941</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you rate your performance in the task?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>1</td>
<td>3.025</td>
<td>6.066</td>
<td>0.024*</td>
</tr>
<tr>
<td>Error</td>
<td>19</td>
<td>0.499</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Comparison

Chi-squared tests were performed to measure the relation between solo and collaborative setup. One of these shows that it was significantly more participants who answered "Collaborative" on the question of Which setup do you think is most useful in an actual pre-production? Why? 17 participants answered the question (3 people answered "Both" and were thus not included in the comparison), where 14 answered collaborative setup and three participants answered solo setup.

\[ \chi^2(1, N = 17) = 14.235, p < 0.01 \]

There were no significant results in terms of Which setup was more enjoyable? Why? or Which setup did you prefer? Why? which will be further discussed in section 6. A second chi-squared test showed that participants also felt like the role they had during the collaborative setup, to a significant amount.

\[ \chi^2(1, N = 20) = 5.000, p < 0.01 \]

### Qualitative Comparisons

The final questionnaire that the participants filled out after the entire experiment was finished included questions of a qualitative character. These answers partly provided feedback on the design of the interface where the participants found difficulties, such as how controlling elements in the interface. Three questions in the final questionnaire included parts that encouraged the participants to motivate their choice of either solo or collaborative setup. These answers were analyzed and coded. After this, a thematic analysis divided and connected codes into themes. The themes that were eligible are those with codes occurring two times or more. The nature of these questions can be divided into two categories: comparison between the two conditions, and
design of the interface. The parts regarding conditions can be seen in Table 4, whereas the results regarding design will be further discussed in the discussion section. The identified difficulties in the interface design was to a majority about the controllers and usability issues with HTC Vive.

"One of the bigger challenges were that the picture from the camera got lost inside walls and characters. I wanted it to stay visible the whole time and not move around."

"Move the actor in a constant speed"

Other difficulties included the controllers and learning them correctly:

"Learning to handle the tools"

"In the beginning learning all functions with the hand controls."

Table 4: Thematic analysis of qualitative results.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Focus</td>
<td>3</td>
</tr>
<tr>
<td>Collaborative Fun</td>
<td>4</td>
</tr>
<tr>
<td>Solo Focus</td>
<td>3</td>
</tr>
<tr>
<td>Which setup do you think is most useful in an actual pre-production? Why?</td>
<td></td>
</tr>
<tr>
<td>Collaborative Communication/work together</td>
<td>7</td>
</tr>
<tr>
<td>Testing</td>
<td>2</td>
</tr>
<tr>
<td>Real-life scenario</td>
<td>2</td>
</tr>
<tr>
<td>Different skills</td>
<td>2</td>
</tr>
<tr>
<td>Tangible</td>
<td>2</td>
</tr>
<tr>
<td>Which setup did you prefer? Why?</td>
<td></td>
</tr>
<tr>
<td>Collaborative Creativity</td>
<td>2</td>
</tr>
<tr>
<td>Solo Safe/ no pressure</td>
<td>3</td>
</tr>
<tr>
<td>Easier</td>
<td>2</td>
</tr>
</tbody>
</table>

The themes found in the analysis (see Table 4) reflects the occurrences of how many times a theme was mentioned. In the first part, on the question of which setup was more enjoyable, participants mentioned that the focus was superior in the specific setting. These could therefore have been mentioned in both setups, as different participants mentioned focus in different setups. Furthermore, the answers revealed how the communication and working together with someone was important, and that the use of different skill sets reminded of a real-life scenario where filmmakers would collaborate, as described by the following quotes:

"[…] you cannot manage all tasks during a real shoot either and when you already start to collaborate in the pre-viz it will be easier to do the real shoot in the end"

As mentioned in the previous subsection, there was a significant result in how the participants felt like the role they played during the experiments. This was evident in both the quantitative and the qualitative part.

"I could comment more, like a director would"

"I felt like a director because the dialogue with the photographer was just like on the set"

6 DISCUSSION

As hypothesized, the participants ruled the collaborative setup to be more useful in a real-life pre-production scenario. This was due to the nature of collaborative work that was present in the virtual environment tested in this project. The participants mentioned the use of different skills and the relation to a real-life scenario where working together is a standard part of filmmaking as reasons for the usefulness of the tool. However, what was also hypothesized was that the collaborative condition would be more enjoyable and preferable (Which setup was more enjoyable? Why?, Which setup did you prefer? Why?) due to the same reasons as the usefulness. We did not find support for this, as no significant differences were found, and the participants found it slightly more enjoyable in the solo condition, and marginally preferred this condition. The reasons for this can be found in the relationships between the questionnaire questions. In the repeated measures ANOVA (Table 3), we found that the concentration level and how each participant rated their performance showed significantly higher results for the solo condition. This is also something that was not hypothesized yet showed interesting trends. The level of performance is something that can be explained by the comments from the participants:

"[T]he one where I was alone was a bit more enjoyable since I could focus solely on how I wanted the scene and how I could perform. But the interaction of being two people in the same VR-scene collaborating was a new and very fun experience which I think, given more time would give a better end result."

"Since I did the single (solo) part last I preferred that one. I felt most secure then."

Interface

It is evident from the results that feeling in control of the tool and its components had impacts on the participants’ experience. The interface design and the collaborative aspects of working in this particular virtual environment are
Thus connected. The participants thought the tool would be useful in a real-life scenario, to a significant level, whereas a small majority of the participants enjoyed and preferred the solo condition. This is somewhat contradictory, as one could assume that professional filmmakers would enjoy a situation similar to a real-life scenario more. It can be argued that the reason for why they enjoyed and preferred the solo condition to a slightly higher extent could have been due to several effects: how well the participants knew each other and if they had previously worked together in a similar situation, which condition the participant started out with, and how they compared themselves with other participants during the collaborative condition. Having someone else present in the creative work could have made some of the participants less likely to rate their performance high, as they may feel judged by the other participant. On the other hand, this could also be related to the way the interface design was built, where the role of director did not have the same controls as the photographer and was thus more limited in the functionality that could be accessed. The reason behind this was due to the hypothesis that an asymmetric system where different roles have different specialties should be amplified to force the users to collaborate, as is described in [9] and [24]. It can be argued that this fact could induce the participants to feel more secure and free in the solo setup, thus rating that one higher in regard to their own performance and preference. Also, some directors complained about not having a monitor and tried to stand behind the photographer to see the image. This caused limitations the collaborative work. Other participants teleported to the cinema room so they could see the image in real-time. However, this may have caused a lack of FTF communication with the photographer.

In both the experimental conditions, the question How compelling was your sense of being present in a virtual world? scored 5.9 on the 7-point Likert scale, as can be seen in Table 3. This highlights how the design of the interface was successful in the participants’ sense of presence in the virtual environment. The affordances of VR should mimic the real world as much as possible, as mentioned in [8]. Our results clearly show that this was the case due to the above-mentioned results together with the results of how the participants were involved to the extent that they lost track of time to a high extent (solo: 5.4, collaborative: 5.1). Furthermore, after the collaborative setup, the questionnaire statement I was aware of the actions of the other participants received a result of 5.95, further highlighting that the interface did in fact reflect a real world scenario where filmmakers collaborate with other colleagues. Furthermore, the representation of other participants with avatars and audio feedback of voice made the communicative aspect of the collaboration successful. A majority of the participants could find grounding in their communication early on in the collaborative experimental setup, where most participants initiated discussions to make creative decisions immediately.

The amount of training that the participants got before the experiments could have been longer to obtain the full potential of the participants’ work. According to one participant, having about an hour of training would probably make the interface more usable. The training session lasted for 10–15 min during the experiments, and the participants did learn the controls and functionality in different places. Thus, if a participant started out in the collaborative setup with the role of director, the training would include functionality that the experiment did not. These functionalities would come back to the participant during his/her solo session. However, this opinion was not shared by most participants and can therefore not be assumed to be a significant cause.

“The second one (solo setup) because I had got better grip on the technique the second time.”

“[...]the last one (solo setup), when I had more experience”

Moreover, the above mentioned quotes point out that the order did in fact have some effect on their perceived performance and experience. Several participants mentioned that the collaborative condition was easier to perform in if the solo condition had been performed first.

Lastly, the main difficulties mentioned from the qualitative results points to the controls and usability issues with the HTC Vive, highlighting that further improvements to the interface should focus on usability issues rather than on the communicative and collaborative aspects of the interface design.

**Personality Traits**

Another possible explanation to why the collaborative setup was not significant in regard to enjoyment and which setup the participants preferred had to do with personality traits: the higher the score within extraversion, the more enjoyable the collaborative setup scored. In addition, the collaborative setup was more preferred by people rating high in extraversion. The significance of extraverted people enjoying and preferring a collaborative setup was not surprising, as extraverted people can be assumed to be more interested in being with other people. Also, before the experiments this was an expected result and the main reason for why the personality questionnaire was included after all the experiments, this result highlighted the hypothesized correlation.

How often the participants interact with VR and the level of difficulty did not show any significance, which was in high contrast to the hypothesis. The same results can be found in how comfortable the participants felt with controls typically used in video games and the level of difficulty. As discussed earlier, the interface did receive high ratings in
Insights on the Design

Focusing on the main research question (How to best design collaborative tools for interactive distributed previsualization within film production in virtual environments?) the current design had some flaws that can show pitfalls for future work. First and foremost, the decision to divide the user interface between the two roles had benefits for the collaborative aspects, as participants felt as their role, to a significant level. However, it also showed some negative effect as some participants mentioned that they felt limited when working as the role of director. The effect of order and the different personality traits did also show some effects as a few participants mentioned that they preferred the second task because they had more experience in the tool. Therefore, pin-pointing a universal design for this kind of collaborative interface is complex.

Insights. The results showed findings within the design of the interface and its components. In this sub-section, the findings will be summarized into design insights of the interface design. First of all, based on the related research, it was evident that communication and grounding is crucial for collaborative environments. These include visual and audio feedback of other users present in the virtual environment, e.g. avatars as representation of other users. Furthermore, the use-case of the VR application plays an important role. In this project, the use-case was clearly defined as a previsualization tool primarily made for filmmakers. Thus, the intended users had knowledge-in-common that helped them understand the task and make it clear for them how to work creatively in the tool. Therefore, collaborative VR applications should have clearly defined users where the virtual environment reflects the knowledge-in-common of the users. To provide a collaborative context, the need to coordinate the tasks becomes more simple when the users have this shared knowledge that is present among people in a specific discipline or working context.

Another important aspect is the shared visual information. As described in the related literature, having shared visual information is crucial for providing grounding which furthermore provides a foundation for communication and collaboration.

7 CONCLUSIONS

VR applications are becoming more advanced and are finding their way into more peoples’ lives. The use-cases of VR applications are expanding and filmmakers can benefit from these technologies. Previsualization is an important part of pre-production that can help with the planning of scenes and takes in a cost-friendly way. We developed a VR tool for previs made for filmmakers who don’t necessarily have the required skills to perform previs in a traditional way. This thesis have reported on an experimental analysis where 20 participants, specialized within filmmaking, performed experiments. These experiments were assessed to analyze results and define insights on how to best design collaborative applications for filmmakers in VR. The subjects in the experimental design tried out both a solo setup and a collaborative setup where they could see each other represented as avatars, and have audio feedback for conversation and discussions. Our results indicate that the tool is useful for real-life previs, where filmmakers can discuss, edit, and validate different camera works and character movements. The participants in this study provided positive feedback in terms of sense of presence, immersion, involvement in the task and the collaborative aspects of working together with another coworker in a virtual environment.

The insights obtained for collaborative VR applications within filmmaking show that making use of the knowledge-in-common among a specific set of users and enhancing this shared knowledge is important for providing a tool that can be perceived as serious and useful for real-life scenarios. Making use of the shared visual context for providing proper grounding and understanding of the working environment is furthermore important for how the filmmakers can collaborate in a fulfilling way.

Future Research

This study presents experiments with 20 participants. For future research, further experiments with a higher amount of participants with this specialization should be performed. A larger number of participants would provide more data for analysis which can be generalized to a higher extent than what was done in this project. Also, more iterations on the the design should be performed to further prove the advantages of symmetrical versus asymmetrical design of collaborative VR applications.

8 ACKNOWLEDGEMENTS

I would like to especially thank my supervisor André Pereira and Simon Alexandersson (from Monocular) for supporting me through the entire process. Also, a big thanks to Mirko Lempert (Monocular) for helping me with the project.

REFERENCES
