LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Department “Institut für Informatik”
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Heinrich Hußmann

Bachelorarbeit

Creating a Shared Virtual Environment Using a Head-Mounted Display and a Smartphone

Tristan Xaver Firsching
Tristan.Firsching@campus.lmu.de

Bearbeitungszeitraum:	1. 4. 2020 bis 30. 7. 2020
Betreuer:	Ceenu George
Verantw. Hochschullehrer:	Prof. Butz

Abstract

Virtual reality (VR) headsets are a powerful tool for immersing the user in a simulated environment. During the experience, however, the user is visually isolated. The headset completely shuts out the outside world, often making it a very solitary experience. When other people are in the room, they commonly struggle to communicate with the VR user. Such interactions demonstrate how important non-verbal communication is in face-to-face situations. Removing the barrier between the virtual environment and the outside world can open new possibilities for collaboration using virtual reality. This paper explores the possibility of using a smartphone to give others access to a shared virtual environment with the VR user. To assess the viability of the concept, we developed a prototype that allows a second user to connect to a VR application with a smartphone and interact with the virtual environment. We performed a small scale study that observed how collaboration changed when allowing the participants to use this method of interacting.

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, 28. Dezember 2020

. . . .

List of abstracts

1 Introduction

Head-mounted displays (HMDs) allow users to immerse themselves in interactions with virtual environments in a more intuitive way than other input methods. Previously interaction with 3D environments(VE) was limited because interfaces had to project 2D inputs into 3D spaces. An HMD together with motion tracked controllers, allows users to navigate and manipulate virtual environments by mimicking real-world actions instead of learning abstract controls. Reducing the levels of abstraction in the input methods makes virtual reality (VR) uniquely suited not just for entertainment but also for education, training or scientific applications.

One drawback with the use of HMDs is separation from the real world. Wearing a VR HMD shuts out the outside world visually, making communication with bystanders difficult. The virtual reality user cannot see other people in the room, and bystanders cannot easily understand the virtual environment he or she is interacting with. Attempts have been made to mitigate this disconnect, but the usability and interactivity is limited. Some HMDs include cameras attached to the front, the feed from which can be displayed to the VR user. Additionally, software running on a desktop PC usually mirrors the HMD view on the desktop screen, offering bystanders a better understanding of the virtual environment. The practicality of this, though, is limited because of the small field of view displayed on the monitor and bystanders needing to turn away from the user to look at monitor.

For some applications, especially entertainment, this limited interaction is not a problem, because they can be designed to not require collaboration between users. VR is however beginning to expand into fields other than entertainment. The technology is being used as training tools for specialized workers [1] or for scientific visualization [2].

Providing multiple people access to a collaborative virtual environment(CVE) can create difficulties. Using multiple headsets is possible but often impractical. Making the virtual environment accessible without the requirement of a headset would allow more participants in this type of collaboration. By making the CVE accessible on any modern mobile phone the number of users who could access the CVE can be increased greatly.

This thesis explores a possibility of creating a shared virtual environment by connecting a smartphone with the VR application.We developed a prototype that connects two applications to access this virtual environment simultaneously. This prototype can create the effect of simulating a window into the CVE with the smartphone by tracking the phone's position allowing simple interaction with someone using a VR headset. To judge the prototype's viability, we conducted a study to assess differences in performance and workload when using the system. The study required two participants to collaborate in solving a task. One participant used a VR headset and was able to manipulate the environment, while the other used a smartphone to assist them. THis study tested the following hypotheses:

- H1: Participants assisted with the smartphone application need less time to complete tasks and make fewer mistakes than participants who are assisted only by voice.
- H2: Participants using the mobile app experience a lower cognitive load when assisting another user than others only assisting by voice.
- H3: Virtual reality users also experience a lower cognitive load when being assisted using the app.

2 Related Work

Collaboration using virtual environments has been explored in various contexts. Fraser et al. identified an awareness of others as an essential feature of shared workspaces. [3]. The paper aimed to provide a descriptive theory of awareness, while focusing on collaboration within small groups. This theory is to be used as a framework for developers of programs that provide a shared workspace. Although the paper was written in 1999, before major technical advancements in virtual reality as well as interactive 3D software, it served as the basis for many concepts that would later be implemented in this space. The authors defined workspace awareness as an "up-to-the-moment understanding of another person's interaction with a shared workspace", [3, p. 3] meaning an understanding of where someone else is working, what they are doing, and what they plan to do next.

Groupware applications are programs that create a virtual shared workspace.These can be a shared text editor, an online video game or a distributed control system. The researchers focused on understanding the physical workspace where people work face to face with real world artifacts. The researchers described these concepts to give future groupware designers a guideline for applying them to virtual shared workspaces.

The major limitation of groupware workspaces as compared to face-to-face is the lack of information that can be exchanged. Subtle cues in gestures, sounds and peripheral vision are lost, forcing users to use more inefficient verbal communication. Making all that information available however is impossible, leading the Fraser et al. to focus on determining what information to gather, when and for what activities it is important, and how to best present it within a groupware system. The concept of workspace awareness was based on the understanding of situation awareness, which has its origins in military aviation. There it is described as consisting of three stages: perception, comprehension and prediction. Fraser et al. saw workspace awareness as a specialization of situation awareness where the first two stages are especially relevant. Perception of others in the shared workspace requires presenting information in a way that is easily comprehensible. Neisser [4] described the act of maintaining awareness with the perception-action cycle. This cycle describes how actors continually explore the environment and gain knowledge by perceiving changes. The workspace awareness framework adds a new link to the cycle (action) to describe how actions on the environment help people understand it.

In addition to compiling existing research on the topic, the authors conducted their own observational studies as well as developing and testing concepts practically. Fraser et al. performed informal studies were performed observing different collaborative face-to-face environments including various tabletop task or games, as well as in professional workplaces. These studies helped them understand how awareness information is gathered as well as how it is used. The testing included monitoring the study participants while they performed table top tasks like puzzles or cardboard construction as well as watching real work situations in an air traffic control center.

Workspace awareness requires different kinds of knowledge of the environment. In their framework the authors separate awareness into different elements. Instead of supporting all of these elements equally, the researchers recommend carefully selecting them depending on their relevance to the task.

Observations indicated that for face-to-face interactions, awareness information is gathered in large part by gathering visual information from another persons position, posture, or movements. This consequential communication is information that is exchanged unintentionally through the actions of a person in the environment. Movement was also an especially important component to consequential communication. People could perceive actions of others while focusing on something else, since attention is naturally drawn to movements.

Another way of gathering information in an environment is through artifacts. These are physical or conceptual objects that can be perceived and manipulated. In a face-to-face work environment artifacts could be tools, buttons or gauges or the conceptual desktop of an interface. When people use an artifact, it often produces distinct sounds that are recognizable to people. For example, a scissor makes a distinctive sound when cutting, giving feedback to the user but also communicating information to others in the room. This concept is described as feedthrough and is an important part of how people naturally understand the actions of others in their vicinity.

Intentional communication is also an important way people gather information about the environment. This mostly takes the form of verbal conversations or gestures, that are directed toward a specific person. In groups, this has the side effect of allowing others to overhear conversation to gain new information. Public conversations can allow larger groups to maintain awareness of the environment. Navy ships have open radio communication channels to take advantage of this; any crew member can hear communications without taking part. Additional information can be conveyed with "verbal shadowings", where people commentate their own actions. In small groups, these can help others passively keep track of another’s progress and tasks.

In a collaborative setting it is common for people to continually shift between individual and shared work. They will work alone until an opportunity to collaborate becomes apparent. Making the transition to shared work requires awareness of what the other person is doing. Monitoring peripheral actions is necessary to recognize these opportunities and prevent inappropriate interruptions of others.Artifacts or props can simplify communication and increase efficiency. This is obvious when reviewing transcripts of the observed collaborative activities, which become almost unintelligible without knowing the full context. The conversations are reliant on deictic references which are dependent on the context of who is talking, what they are doing and what they are looking at. References like "this", "that", "here" or "there" are crucial in face-to-face collaboration and significantly increase the efficiency of communication. However, people do not only use gestures to point out things. Often they use gestures to illustrate actions, behaviors of objects or relationships. People can trace a path with a finger, illustrate a size or distance or mimic an action, to accompany a verbal explanation.

Another part of effective collaboration is coordination. For some tasks it is necessary to coordinate the order of actions, timings as well as division of labor between people. While this can be achieved with explicit communication, awareness can make this unnecessary. When a workers are aware of the actions of others and understand what they are trying to accomplish, they can realize for themselves when to step in to help, or what actions are best to assist each other. Another key part of effective coordination is anticipation. When people maintain awareness of their environment, they can predict changes and act accordingly. Conflicts can be avoided, because future actions of others can be predicted by extrapolating their movements. After a while people recognize patterns of how others work and can anticipate their next steps. When these patterns are disrupted people can naturally realize that assistance is needed.

Duval et al. applied the concept of awareness to create a system that helps experts explore scientific data collaboratively [2]. The authors created a prototype that allows a group of experts to view three-dimensional diagrams in a CVE, while supporting communication between them by using additional systems. The researchers saw an opportunity in using virtual reality techniques to make complex scientific data more digestible to groups of experts. By giving multiple experts access to a shared virtual environment and providing additional tools for interaction, their prototype aims to improve the effectiveness of communication when analyzing higher dimensional datasets. To effectively use a shared virtual environment, the system needed to provide workspace awareness. The researchers made use of the concept of avatars to represent the users inside the VE. These were human figures of workers wearing a colored helmet representing other users. While testing participants wore helmets in the colors relating to their avatar, making each user identifiable by a color. To provide better awareness of activities, the user interface and tools, which allows a user to interact with the program, were made visible to others in the virtual environment. These avatars greatly helped users understand the other's intent by making visible what they were focusing on. The prototype also implemented a system for shared annotations. Any user could create and view these to mark a point of interest. Users could save a viewpoint as an annotation and make it visible to others. These could be selected to view a section of the data from the intended perspective. To facilitate the manipulation of a 3D environment, a raycaster was seen as especially useful. The raycaster was used mostly to convert the position of a 2D -cursor into a position in the 3D virtual environment in an easy and intuitive way. This was used as an easy way to allow users to select objects with a pointer.

Other research has focused specifically on analyzing the behavior of users when collaborating around a single display [5]. Inkpen et al. developed a system to project a display onto a table and tasked the study participants with playing a "memory-style" game. The authors recorded the participants and focused on observing their non-verbal communications.

Researchers introduced additional independent variables in the form of two different input devices the participants could use: a mouse and a stylus pen. The study found that an awareness of intent and a shared understanding could benefit communication. When participants used a mouse as the input method, the system supported awareness of intent by showing two cursors on the shared display; however, usability proved to be poor because the small cursors were hard to see. While the use of a stylus did not significantly improve performance in the game, the behavior of the participants and their responses to questionnaires suggested it felt more natural and easier to use with a desktop display than using a mouse. Inkpen et al. also suggested using a touch-sensitive display as a more intuitive input method but dismissed it because of the limitations of the technology at the time.

Mai et al. [6] attempted to improve communication with a person wearing an HMD, by making the face visible using a display attached to the headset. The prototype, however, was limited, as it was unable to track the HMD user's real facial expressions, and instead generated eye movements based on bystanders in the room. The researches used a smartphone to display a 3D model of a face that could be animated. By using the phones camera to track faces in the room, they could animate the eyes to react to bystanders. Additionally, the researchers used tracking data to create a perspective illusion, which, however, could only work correctly for one bystander. The researchers created the illusion of a transparent headset by adjusting the perspective of the 3D rendering based on the position of a bystander.

Stereolabs[8] implemented a method of combining real-world footage with the virtual environment. By attaching a 3D tracker to a camera and placing a greenscreen behind the user, they could combine the footage captured of the user with the footage captured by a virtual camera. The tracker allowed them to determine the correct perspective to choose in the VE while the greenscreen made it possible to cut the user out of the picture. Combining both video feeds creates the illusion that the user is situated inside the VE. A system like this could be useful for demonstration purposes, since it is able to effectively the actions of a VR without having to switch between first-person and third-person perspectives.

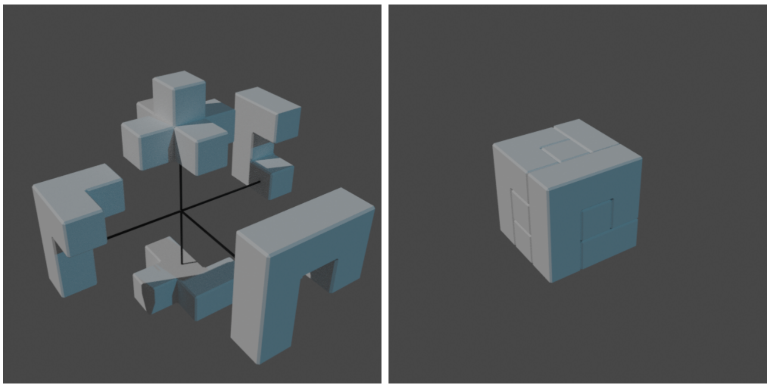

Figure 4.1: The cube puzzle, exploded and assembled view.

3 The Concept

Modern virtual reality hardware prevents awareness almost completely. As opposed to augmented or mixed reality headsets, VR inherently visually isolates the user from others in the room, making any kind of communication other than verbal impossible. Our concept aims to provide methods to improve consequential as well as non-verbal intentional communication by giving a bystander a method to interact with the virtual environment. By creating an application that works with both a smartphone and a VR headset, we can create shared virtual workspace. This allows for a shared understanding of the virtual environment, giving bystanders a better awareness of a VR users actions without the need for additional VR hardware. While the interaction capabilities using a smartphone are limited relative to the motion tracked controllers common in VR systems, the inconvenience and hardware requirements of using multiple headsets simultaneously are a significant barrier for using VR in collaborative settings.With increased awareness VR could become a more viable option for collaboration in specialized task, that benefit from the 3D capabilities it can provide.

4 The VR Application

The VR application uses the SteamVR platform and was tested with a HTC Vive pro headset and two Vive controllers. The users were placed in a virtual room similar to the real environment in which the study is performed. Users could navigate the entire room and interact with objects by picking them up and placing them down.The application is designed with a room-scale environment in mind, meaning that the user could partially navigate the room by physically moving in the real environment. In case space was limited, users can move their virtual position using the teleport system.

For the study, the program included three phases. When first starting the program, the room was empty to allow the person running the study to explain the basic navigation system. In the second phase, a single puzzle piece was placed in the room to introduce the method for picking up and laying down objects. The third phase was the main task.

4.1 The Task

The application presents the user with a task to complete that requires the user to manipulate objects in the virtual environment. At the start of the task, five differently shaped pieces appear scattered across the room. When placed together in the correct configuration, the pieces fit together in a cube shape (see figure 4.1). The users can pick up, carry, and put down any of the pieces in the room.

To make the placement easier, the pieces conform to a grid when placed. In their placed state, the pieces are not affected by gravity or collisions to prevent the user from accidentally knocking over a half-finished puzzle.

The program automatically checks whether the puzzle is completed every time a piece is placed. The program tracks the time it takes each user to complete the task and how many times objects are moved. These and other statistics are saved when the task is completed.

4.2 Interaction

Users could interact with certain objects in the environment by moving the controller near them and pressing the trigger button. When the controller is moved near an interactable object, it is highlighted with a yellow outline. When the user holds down the trigger button, the piece is attached to the virtual hand until the button is released. Puzzle pieces snap to an invisible grid when released to reduce the accuracy required from users. During snapping, the piece is moved to the nearest point on the grid while the orientation is rounded to an increment of 90 degrees for each axis.

Since collisions are not considered while a piece is being carried it could be possible to overlap parts of the pieces in space. Accidentally placing two pieces inside each other could be confusing to users. To avoid this, the program checks for overlaps after the user places a piece down. When an overlap is detected, instead of attaching to the grid, the piece's physics simulation will to be enabled. This makes the piece appear to be loose and communicate to the user that it should be placed again.

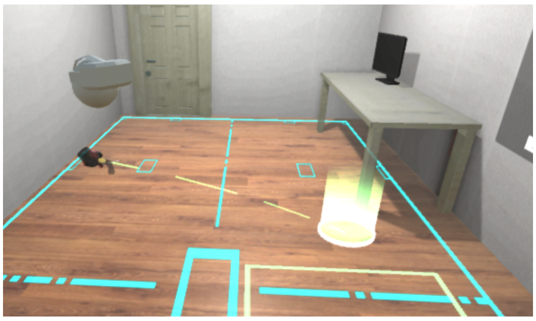

Figure 4.2: Visualization of SteamVRs teleport function

4.3 Navigation

Since HTC Vive provides a large range of motion, a large part of the room can be navigated by simply moving within the real environment. In case space is limited, the application additionally provides a teleport function that allows the user to move to any location in the virtual environment without moving in the real world. By pressing a button on any controller, the user can point to any location in the room. To indicate the target location, a raycast metaphor is used. A curved line is shown, with an indicator highlighting the target where it intersects with the environment (as seen in Figure 4.2). Valid positions are defined by a predefined plane. If the target is invalid, the target indicator turns red. The ray moves in an arc, which makes choosing more distant locations easier by limiting the reach. When the button is released, the virtual perspective is moved to the target location.

4.4 User Interface

Additional control is provided by a user interface in the virtual environment. The interface is shown as a floating panel that can display buttons and menus. To adjust its position, the user can grab a handle attached to the bottom of the panel and move it in the same way other objects can be manipulated. Buttons can be pressed by moving one controller near it until the button is highlighted. When a button is highlighted, the trigger button on the controller will activate it. Buttons allow users to repeat a task or advance the phase of the study. The same functions can be triggered with a keyboard to allow the person running the study to advance phases using a computer. When a task is completed, a questionnaire can be filled out using this panel. The results of the questionnaires and tasks are saved to a file locally.

Figure 5.1: Screenshot of the full smartphone application

Figure 5.2: Outside view

5 The Smartphone Application

The smartphone application (SA) connects to the VR application (VA) via internet to allow data transfer. With this connection, the app can display a synchronized version of the same virtual environment the VR user sees. By tracking the phone's position in space, the virtual perspective can mimic the behavior of a physical camera. This creates an effect where you can look into the virtual world the same way you would use a handheld camera. In the study, the SA user acted as an assistant to the VA user. Additionally, it is possible to create pointers in the virtual environment by tapping the desired position. The SA user also has access to a solution to the task, making communication between VA and SA users a necessity.

5.1 Navigation

The app allows users to control the perspective of the scene in two ways. One option is to use an HTC Vive Tracker attached to the phone. This allows full position and rotation tracking and allows the user to look at the virtual environment in a similar way to using a handheld camera. With this solution, the virtual environment can be navigated simply by traversing the physical space. This allows much more intuitive interaction with the environment but might be more physically demanding for users. Additionally the application provides an option to navigate the scene using touchscreen gestures. Swiping the screen rotates the virtual camera in the opposite direction. Horizontally the rotation orbits around a point two meters in front of the camera, making it easier to look at different perspectives of objects that are at the scale of the puzzle pieces. Vertically the camera will rotate around its own axis to allow users to easily look down at the floor or up into the distance. By using a pinch gesture with two fingers, the perspective can be moved forwards and backwards. While placing two fingers on the screen, the program will measure the change of distance between the touch points over time and apply it as forwards or backwards motion of the

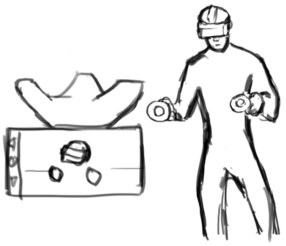

Figure 5.3: Application sketch

Figure 5.4: View when using a tracked phone

virtual camera. The user can manually switch between the motion-tracked and the touch controlled modes by pressing a button on the touch screen.

5.2 User Representation

Because the SA and VA both use the same tracking solution, the system knows the exact relationship between all the tracked objects. This makes it possible to use avatars that correspond to the real position of the users in the room. The VR user is represented as a simplified floating head wearing a headset and two hands, reflecting the positions of the Vive controllers (see figure 5.4). To the VR user, the smartphone user is also represented in the virtual environment with an avatar. The user is represented with a floating smartphone, that corresponds to the virtual camera used in the smartphone app. This allows the VR user to maintain awareness of the smartphone user in the real environment. When the SA user is not using the motion-tracking solution, the avatar will reflect the position of the virtual camera. This way awareness of intent can still be provided with the same avatar. This way awareness of intent can still be maintained even when the system can not track the position of the SA user in the room. The position and rotation of the phone avatar can communicate information and provide context for verbal communication.

5.3 The Pointer System

By tapping any point in the environment, the user can create a pointer that is also visible to the VR user (see 5.5). This allows a bystander to easily direct the VR user to a desired point in the virtual

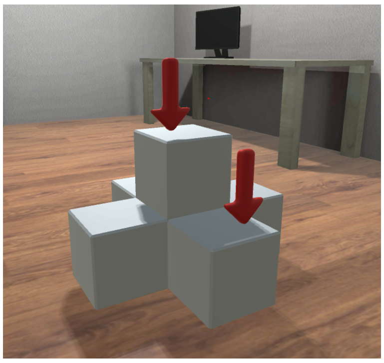

Figure 5.5: Example of the pointer system.

Figure 5.6: Screenshot of the instructions in the full smartphone application

environment. The pointer also creates a 3D sound to help the VR user to more easily notice and locate the arrow when it is not in his or her field of view.

To convert the 2D position of a tap into 3D coordinates in the virtual environment, a raycast metaphor is used. When a tap is registered, the program finds the first intersection behind the position of the tap. Non-verbal communication is severely limited for the smartphone user, since the VR user does not have visibility of the movements of others in the room. This makes this feature essential to improve awareness of intent. Because the SA user is not able to use gestures, this feature aims to implement a way of pointing at an object or in a direction.

5.4 Instructions

The smartphone app provides access to instructions with the press of a button. The instructions show an interactive solution to the puzzle. When pressing a button labeled "Help", an overlay appears that shows all the pieces in an exploded view of the correct configuration(see figure 5.6). This view can be rotated by swiping on the touchscreen with one finger. When this overlay is active, the pinch gesture animates the pieces to move between the solved state and and the exploded state (similar to 4.1). When the instructions are active the "Help" button changes to display the word "Back", allowing the user to switch back to view the virtual environment. To maintain awareness, the virtual environment is still visible in the background when accessing the instructions. The virtual environment is darkened in this state, but still allows the user to be aware of movements and actions.

5.5 The Limited Version

Specifically for use in the study, the application has a limited version that disables most of its features. This version only provides the functionality of the instructions overlay. This way participants have the exact same way of accessing the instruction between versions, giving them no advantage or disadvantage which could emerge from changing the way the information is delivered. The limited version still connects to the VA in the same way but shut out all additional information from the user to allow the program to automatically gather user statistics and share them with the main program running on the PC.

It can be activated or deactivated by the person running the study using a special command on the PC.

6 Software Design

Both the VR and smartphone application are developed using the Unity3D game engine version 2019.4 [10]. The same project can be compiled for Windows PC and Android. The VR interaction system is based on the SteamVR plugin for unity [9]. The plugin contains basic functionality for locomotion and interaction with objects in the environment.

The teleport locomotion system included in the SteamVR plugin system allows the user to move his perspective to any position in the virtual room. The teleport function was not modified other than making adjustments to the reachable area. To facilitate interaction with objects, the plugin provides a system that highlights usable objects when touched and attaches them it to the hand when a button is pressed. To make it usable for our use case, the system was customized to align the object with an invisible grid when the user releases the button. When the chosen location is valid, the object will no longer be affected by the physics simulation. This was implemented, to prevent pieces from falling over or from colliding with each other. A position is considered invalid when the object overlaps with another. To improve usability, the system ignores small overlaps and aligns the object correctly. If an overlap is detected, the object is still detached from the hand, but will be affected by physics. Since unity's inbuilt physics engine can reliably resolve overlaps between rigid bodies, the object is visibly pushed away. This way, users can better understand when a piece would collide with another piece at the position they choose. Questionnaires can be filled out within both applications. While the mobile version can use regular touch inputs, the VR version uses the interaction system provided by the SteamVR plugin.

For the purposes of the study, the program writes statistics as well as questionnaire answers to a log file. Both the PC and Android version write their own log files while synchronizing the entries over the network connection.

6.1 Data Transfer

Communication between the the smartphone and the PC is achieved with Unitys Multiplayer Networking API (HLAPI) [11]. The connection uses the client-hosted model, with the PC version acting as the server and local client. The mobile version connects to the PC as a client. Both versions display the same 3D scene which is synchronized using a combination of 'NetworkTransform' components and network messages which are provided by HLAPI.

Positions and orientations of movable objects are synchronized using the 'NetworkTransform' component, which periodically sends the position and orientation of objects and applies them to other clients. This way, the position of the headset and controllers as well as the puzzle pieces can be transmitted to the smartphone app, which displays representative versions of them in its own scene.

The same method is used to transmit the position of the virtual camera used by the mobile-application. A representation of this camera communicates to the VR user that they can be seen and what the other person is looking at. Network messages allow the transfer of custom data types, which are used for additional information transfer.

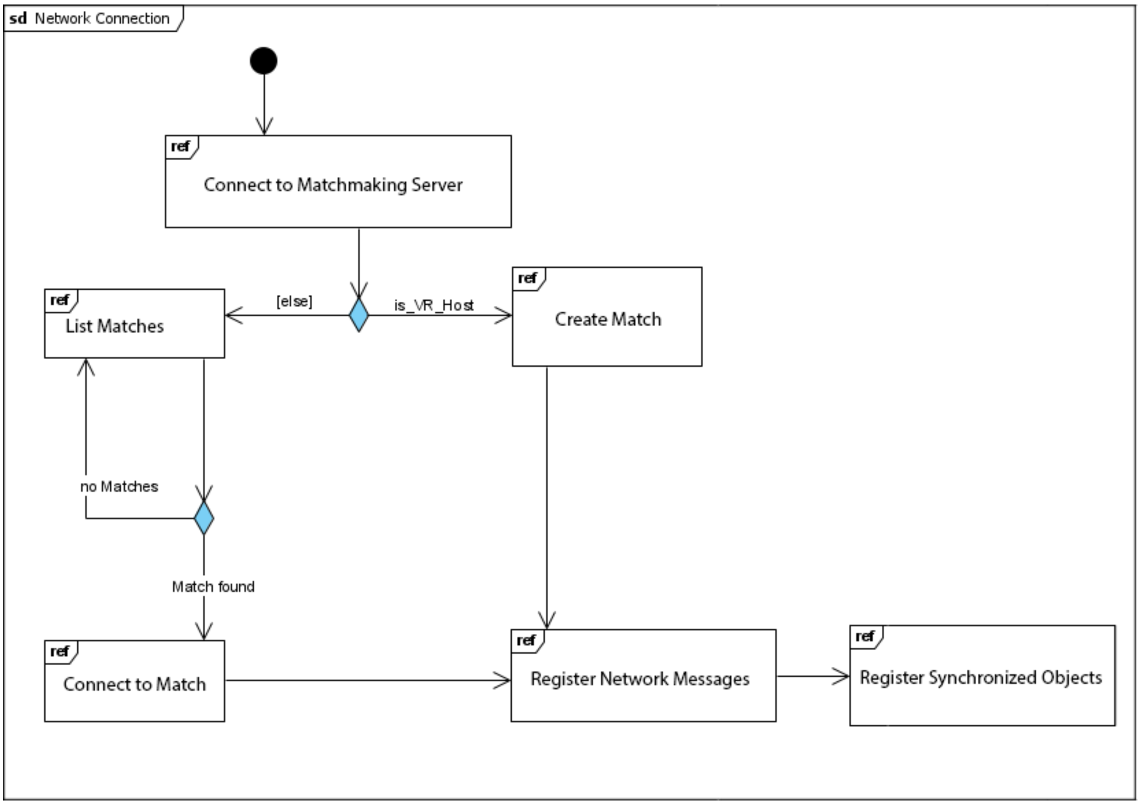

Figure 6.1: Interaction overview diagram for establishing the network connection

Figure 6.1 provides an overview of how the applications establish a connection. Unity provides a matchmaking server that allows programs to register matches using the HLAPI. These matches can be requested by any other instance of the program with the same project ID. After establishing a connections both instances register all network message types to their respective callback functions. This is necessary for the API to pass custom messages to specific functions. Synchronization occurs in parallel with every movable object periodically sending updates.

Questionnaires are built in to the application and can be filled out while using the HMD or the mobile-app. The results of each question as well as additional statistics are shared with the other app when they are saved to a log file.

6.2 Tracking

All the tracking is done via the HTC Vive system, which can track the HMD, controllers and additional trackers. These devices have sensors that detect laser signal from base-stations in the room and send the data to the PC via Bluetooth. One tracker is attached to the phone, which allows the VR program to detect it's position in the room. The system shares the information with the mobile application, allowing it to position the virtual camera based on the tracked position and rotation.

Sending the position data to the computer first and then to the phone introduces some latency, but it ensures that the relative position between both users always represents the real world relationship.

7 Study

We conducted a between-subjects laboratory study to test whether our solution helps users interact with the virtual world from the outside. The study included two roles: the VR user and the smartphone user. The VR user was given a task to complete in the virtual environment, while the assistant instructed them using the smartphone application. The assistant could use the smartphone application to access instructions on how to solve the task, which he or she needed to communicate to the VR user. The participants were split into two groups. Group A used the full version of the application to assist the VR user. group B could only use the smartphone to access the instructions. Since PC-based VR systems usually mirror the view of the VR user on the desktop, participants in group B were also allowed to watch this view from a PC monitor in the room.

The following hypotheses were tested:

- H1: Participants being assisted with the smartphone application need less time to complete tasks and make less mistakes than participants who are assisted only by voice.
- H2: Participants using the mobile-app experience a lower cognitive load when assisting another user than others only assisting by voice.
- H3: VR users also experience a lower cognitive load when being assisted using the app.

To assess differences in the users' behavior, various statistics were recorded while the participants were working on the task.

The study was performed in a laboratory setting room of similar size to the virtual environment.

7.1 Independent Variables

The study used two versions of the smartphone app as the independent variable. Version A attempts to provide a high degree of interactivity with the virtual environment, while Version B limits the interactivity to only provide instructions. Version A allows the user to view the virtual environment (VE) from any perspective, either by tracking the position of the phone in space or by navigating the environment with touchscreen inputs. The user can switch between these two modes by pressing a button. Additionally, the smartphone user can create pointers which can be seen by the VR user in the virtual environment. The application also provides instructions on how to solve the task. Version B only provides access to the instructions which can be interacted with in the same way. Information about the virtual environment can only be accessed via a monitor in the room which displays the first-person view of the VR user.

7.2 Dependent Variables

During the study we tried to gather data on how the participants performed and how they used the application. By tracking inputs directly within the applications and analyzing video footage were able to use the following dependent variables:

Performance is measured by two variables; (a) *completion_time*, the time it takes the VR user to complete the task excluding preparation time. (b) *moves*, how often the VR user places a piece in the environment after moving it.

The program also tracked touchscreen inputs by the smartphone users, using the following variables: (c)*swipe_count*, how often the mobile user uses a swipe gesture to move the virtual camera. (d) *zoom_count*, how often the mobile user uses a pinch gesture to navigate the virtual environment, (e) *tap_count*, the amount of times the user tapped on the screen to create a pointer or press a button.

erbal communication was also analyzed by counting the number of times statements of different types were made. Statements made by the participants were categorized into five groups: (f)*find_piece*, is any instruction made to identify or locate a puzzle piece. (g)*define_target*, are instructions that describe a desired location or orientation. (h)*confirm_or_deny*, are simple statements, to confirm something as correct or incorrect. (i)*question*, is any question relevant to the task. (j)*other*, are other instructions relating to the task, such as telling someone to look to the right or to step away from something.

Since participants often repeat statements or make half statements and wait for the other to react, the counting follows the following rules.

Repeated statements are counted as one as long the participants does not wait for the other to react. Common statements like "Yes, yes correct" are counted as one, to better reflect the amount of information that is exchanged. Combined statements are counted as one when the speaker does not wait for a reaction. Thus a statement like "put it here and rotate it by 90 degrees" will be counted as one, as long the speaker does not pause and wait for a reaction.

However if the the parts of a combined statement falls into separate categories it will be counted to each of them. Meaning a statement like "Yes, now put it there" is counted as one *confirm_or_deny* and one *define_target*, even if the speaker does not wait for a reaction.

Statements that do not convey any information relevant to the task like "wait a second" are ignored.

(k)*questionnaires*; To asses the cognitive load experienced by the participants answered a questionnaire based on the NASA Task Load Index (TLX) [7]. This questionnaire is deigned as a subjective workload assessment of human-machine interfaces and is commonly used in various domains relating to human factors research. The questionnaires were integrated into the respective application for both users. This allowed smart-phone users to answer questions using a touchscreen interface, while VR users could answer questions displayed on virtual screen using the tracked hand controllers.

Additionally, observations were made during testing to asses performance and other non-verbal communications as well as the comments of the participants.

7.3 The Task

To test the concept and its functionality, two users were instructed to collaborate in completing a task. To finding the optimal task, various concepts were tested to determine their difficulty and suitability before the study before settling on the final iteration. Four major properties are important to the study.

- The task should require the use and understanding of the three-dimensional space created by the virtual reality system. Since the study should reflect real-world use cases, the task should not reflect a use case that is as easily achievable without the use of virtual reality hardware. A fundamentally two-dimensional task like a tabletop game is likely hindered by the limitations of VR hardware, while not taking advantage of its strengths.
- The goal needs to be easily and quickly explainable as well as understandable. If the instructions are prone to misunderstandings performance measures can be affected negatively by creating higher variation.

- Difficulty of the task is also a major issue for creating useful performance measures. If too many participants fail to complete the task in the allotted time, their performance becomes difficult to compare. A Task that is too easy on the other hand could severely reduce the time to complete it, giving the study conductor overall less information to collect and analyze. Tweaking the difficulty requires repeated testing with different subjects, while iterating on the design.

- Since the study aims to analyze collaboration between two people, the task should drive participants to naturally try to help each other. Optimally the collaboration should be balanced, meaning both participants contribute equally.

The final task was a 3D puzzle where five pieces needed to be assembled into a cube. The VR user can pick up and place the pieces anywhere in the room while getting assistance from the user of the SA. Only the smartphone user can view the solution to the puzzle by pressing a button. The solution is shown as a 3D model that can be manipulated by using touchscreen gestures. To promote non-verbal communication between the users, the pieces are not color-coded.

7.4 Procedure

Participants were invited in pairs via e-mail as well as personal requests. Pairs were assigned to a group when both were ready, making sure both groups were balanced by alternating the between the two groups with every new pair.

At the start of the study, the goals of the research were explained to the participants. The participants were given some time to familiarize themselves with the technology and software. Since the participants often had no or very little experience with VR, they needed a short introduction to the hardware. When first putting on the headset, navigation was explained to the participants. They were told to stay in the designated area and shown how to use the teleport system. When the user was familiar the features, the program is advanced to the second phase, where a single puzzle piece appears in the room. This piece was used to teach the actions of picking up and placing it down.

While the VR user can experiment with the single piece, the other user is given the smartphone. They were taught how to navigate the environment by using the tracker or switching to gesture controls. After that the SA user was shown how to use the instructions. When ready, both participants are told to collaborate in solving the puzzle and the program was advanced to the main phase. At this point all puzzle pieces appear and the program starts recording statistics.

After completing a task the participants filled out a standardized questionnaire and are allowed to rate the application and give feedback. If the participants do not finish the task after 15 minutes, the test will be stopped and moved to the questionnaires.

The questions can be filled out either on the phone or in VR.

7.5 Limitations

Unfortunately, the number of participants that could be tested was limited because of the world-wide pandemic occurring at the time of testing. The participants were recruited via the university's mailing list for user studies as well as from people inside the university building. As a result of the recruiting methods the participants tended to be students and younger than 30 years old. Because of the limited number of participants, this research functions as an exploratory study serving as a model for a possible later study with enough participants to make more meaningful statistical statements. The requirement of testing with pairs of participants further limited the number of people that could be invited. While individuals could register to come alone, we had to organize a timeslot with another single participant to make the testing possible. To help in filling out slots, participants were asked to bring a friend if possible while, if needed additional participants were recruited from the building.Another result of the difficulty in finding participants was that the duration for testing was limited. To increase the amount of participants that would spontaneously participate, the study was kept short. About 25 minutes were planned for each pair that was tested. The main task was limited to a maximum of 15 minutes, with additional time planned to introduce participants to the system.

The nature of the subject made it difficult to eliminate noise in the results. Natural variances in the participants could have a large influence on the dependent variables. Some participants may be more talkative, which would increase measurements for verbal communication. Others might have to talk in a non-native language, which could increase the cognitive load. Familiarity with virtual reality may also affect how fast they understand how to use the system.The participants had to be instructed by the study runner on how to use the system. Variations on how things are explained could hinder some participants from understanding aspects of the system as quickly. Optimally these influences could be reduced by increasing the scale of the study or selecting participants to control these variables. Neither was possible because of the small pool of willing participants. For example, testing with only participants who had prior experience with VR could decrease some variability in performance, but putting this into practice would complicate the recruitment process and possibly reduce the amount of participants. To mitigate the variance in VR experience, this study used a training phase before starting the full task. The dependent variables were only measured during the main task, after both participants expressed, that they were ready to start.

Additionally, the prototype we tested has room for improvements in usability. Sometimes participants would not as easily understand the control or had difficulty using them. This likely directly affects performance and cognitive load measures, with participants having the additional difficulty of learning to understand the system as well as using it to solve a task. While the training phase was aimed to solve this problem, some participants were eager to start quickly while not having fully understood everything. Though everything was explained beforehand, some participants forgot or misunderstood things, leading them to asking additional questions during the full task. By improving usability and ensuring systems are more naturally understood, these irregularities could be reduced. Additionally having a way to ensure the user understands the needed inputs fully before moving on to the full task could improve the accuracy of results.

Throughout development it was unclear whether a laboratory study was possible. Conducting the study remotely was a possibility that was explored for a while, but proved to be difficult in practice. Since the prototype connects a PC with a phone over the internet, it was technically feasible to find participants that own a VR headset and let them connect with others that own an Android phone. The amount of oversight and control the study runner would have, however, would have been reduced greatly, making managing and directing participants difficult. Implementing systems to increase remote controllability of the prototype as well as ensuring compatibility with multiple types of headsets to not decrease the amount of possible participant, would increase the scope of the practical project too significantly. The overall uncertainty during development resulted in some time lost pursuing concepts that would not come to fruition. Systems were implemented, allowing participants to answer questionnaires digitally through the respective application. This allowed a study runner to gather statistics as well as questionnaire answer remotely by connecting to the PC used for the VR application. While this was not fully used, parts of it could be adapted for an in-person study.

Though the task used was chosen because it required non-verbal communication as well as making full use of the three-dimensional capabilities of VR, it did not cover all the possible use cases of the concept. Finding the best method to assess the viability of the system optimally is difficult. While task was created to demand a similar amount of mental effort from both users, in practice, the VR user would sometimes just passively follow orders. Unexpected effects like this are difficult to prevent. Optimally, the study could test a variety of tasks that provide a more balanced overview of possible use cases.

Finally, though the study is not large enough to make a conclusive statement about the viability of the concept, insights can still be gained by finding possibilities for improvements in the study conduct as well as in user experience and new possibilities for development.

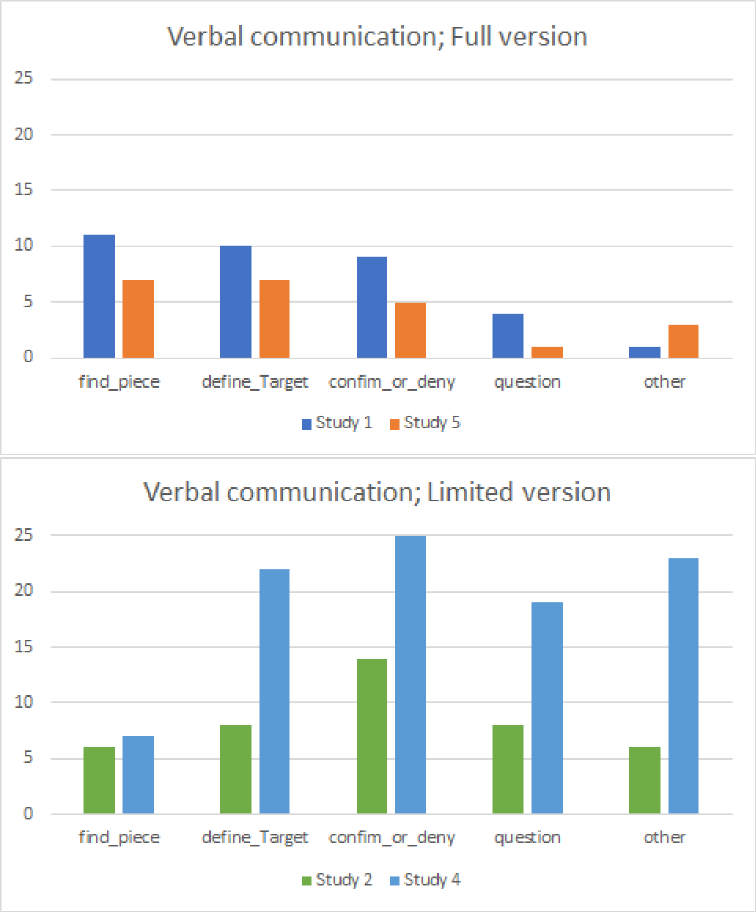

Figure 8.1: Number of verbal statements related to the task

8 Results

The study had 10 participants, six female and four male. Data from computer logs and screen and voice recordings were gathered to compare the use of the full version against the limited version. Additionally, observations regarding the behavior of the participants as well as their feedback were considered.

8.1 Verbal Communications

Verbal communication was analyzed by recording audio and tallying different types of instructions. To compare the two versions of the application, the amount of communication was compared. The results, shown in figure 7.1, clearly demonstrated that the use of the full system can reduce the amount of verbal communication required. Interestingly, not all the categories were affected equally.

Both groups used a similar number of statements to identify a specific piece (*find_piece*) in the room. This could suggest that describing the shape and location of something is not significantly hindered by not being able to point it out non-verbally. This specific task, however, only required a choice between five objects, and the choices became easier in the process of solving the puzzle since the pieces already placed were not likely to be used again. Even when a change to an already placed piece was needed, it could often easily be identified by the sequence in which pieces were placed. Participants could identify the piece by calling for the last or second last which was placed. Most participants quickly realized, that describing pieces by their shape was difficult and easy to be misunderstood. Since most of the pieces did not have a shape, that could be easily compared to a letter or object. Other than one piece that was normally described as a cross, most of the time this method of identifying a piece caused confusion, leading the participants to resort to contextual clues or describing and pointing out its location.

Showing where a piece should go ($define\_target$) appears to be more difficult in the limited version though. Participants especially had difficulties describing a desired orientation. This was observable in both groups but amplified in users of the limited version. Without the functionality of the pointer system, some participants tried to use objects in the environment as landmarks to describe directions. In those cases, it was especially helpful that the virtual room was modeled after the physical laboratory setting, even though the dimensions were not exact. Directions could be described as "towards the window" or "towards the wall".

During the study, the VR users often misunderstood the desired orientation. This commonly resulted in the VR user trying out different possibilities and asking if they were correct. Most occurrences of $question$ and $confirm\_or\_deny$, stem from this type of interaction. The higher number for the group using the limited version indicates that communicating the target was significantly more difficult for them.

For users of the limited version statements in the category of $other$ would mostly consist of instructions to look in certain direction or to get an overview. This type of coordination could be almost completely eliminated by giving SA users control of to their own perspective via the full version.

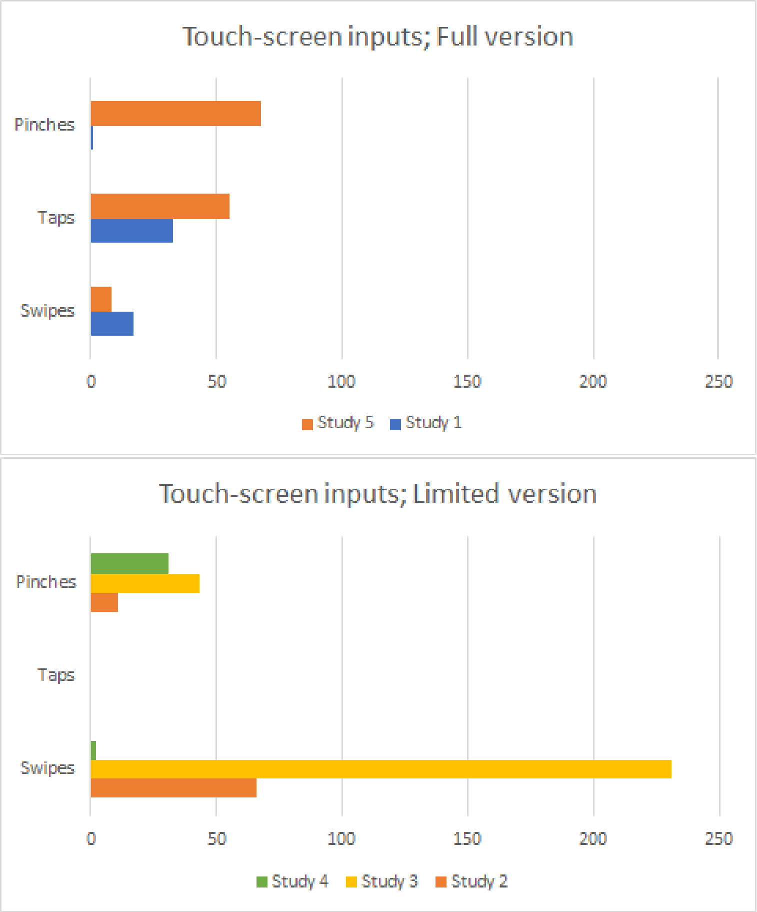

Figure 8.2: Inputs recorded during the task

8.2 Inputs

Figure 8.2 shows the inputs recorded during the task. Inputs controlling navigation, such as *pinches* and *swipes* had, in general a high variability. When interacting with the instructions, some participants preferred to keep a mostly static perspective, while others continually rotated the view. Constant small adjustments to the perspective seemed to help some participants to better understand the 3D shapes, while others preferred to keep the frame of reference the same to avoid confusing orientations.Especially when testing with the limited version, where the participants did not have to switch between instructions and the view of the virtual environment, some continually adjusted the view even when not directly looking at the screen, leading to increased swipe inputs registered in this category.

The pinch gesture allowed users to adjust the distance of the pieces in the exploded view of the instructions, allowing them to see how the cube assembled correctly as well as separate the pieces. Overall the participants this feature less frequently than others, with one participant not using it at all. However, the use of this feature did not seem to strongly affect performance in the task. While the participant of study 1 did not use the feature and performed worse than the others, the participants of study 2 performed while using it only minimally 8.1.

The users of the full version had the option to switch between viewing the instructions and the virtual environment. When viewing the virtual environment, the perspective can either be manipulated with touch screen controls or by using the motion-tracking system. Because most users preferred using the motion tracking to control their perspective of the virtual environment, the overall number of touchscreen inputs was generally lower than when using the limited version. Because participants using the limited version could not switch to a view of the virtual environment, some spent more time rotating the instructions.

Since tapping had no function in the limited version, only users of the full version used it. To support non-verbal communication, the full application implements the pointer system. In practice, the feature seemed to be of great use but some participants forgot they could use it and needed to be reminded. Users also often accidentally triggered the pointer by touching the screen when holding the phone. This did not, however, appear to cause significant confusion, because the pointer would normally be used together with verbal instruction.

The participants made good use of the pointer function, which can be seen by the amount of taps they used. Often multiple pointers were needed, to help the VR user to locate the pointer by sound or looking around. This suggests that extending the duration of the sound and the visual arrow appears could help make it more useful. Others forget they could use it when communicating and were often only reminded by accidentally triggering it when touching the edge of the screen.

Even though communications via physical gestures were limited since only the smartphone user could see the full body of the other participant, we observed that some participants tried to use the tracked smartphone to gesture to the VR user. To instruct the VR user some participants attempted to show desired movements, acting them out by moving the smartphone.

Full	**Time**	**Moves**	**Success**
Study 1	15:12 min	51	dnf
Study 5	05:11 min	19	f
Limited	**Time**	**Moves**	**Success**
Study 2	05:33 min	13	f
Study 3	12:23 min	28	f
Study 4	8:59 min	22	f

Table 8.1: Total performance measurements

8.3 Performance

Table 8.1 shows all the performance measurements. If participants did not finish the puzzle after 15 minutes, the success field is marked with "dnf"; otherwise, the success field is marked with with "f".

The upper part of the table shows the group using the full version while the second group is in the lower part. The variance in performance proved to be high, with some pairs solving the puzzle very efficiently while others struggled. The low number of participants that could be tested makes it difficult to draw conclusions, but the task seemed to be overwhelming for some participants. Overall, performance was slightly worse when using the full version of the application. Making conclusive statements with such a low number of data-points is not possible though.

These results are also very dependent on the effectiveness of communication between participants. When a pair knew each other well beforehand, they were more comfortable with speaking and understanding each other. Some participants had to use a not native language to talk to the other person, which also negatively affected performance results. Aspects like this can add additional randomness to the results.

When participants communicated effectively, they were able to finish the puzzle with minimal errors. In this case, the task could be completed in about five minutes and required approximately 15 moves. A common error during the testing was that the participants placed a piece on the wrong side of the cube, which required them to recognize the error and rearrange two pieces. Finding mistakes and correcting them can add significantly to the final time. From observations, however, most errors were not caused by miscommunications, but by the SA user giving the wrong instructions. Often, the SA user confused the orientation of a piece in the instruction, causing the VR user to misplace the piece. However, errors like this do not directly relate to the quality of communication the prototype enabled. Increasing the scale of the study would minimize unrelated effects like this. A larger sample count could provide a better picture in the differences between the tested groups.

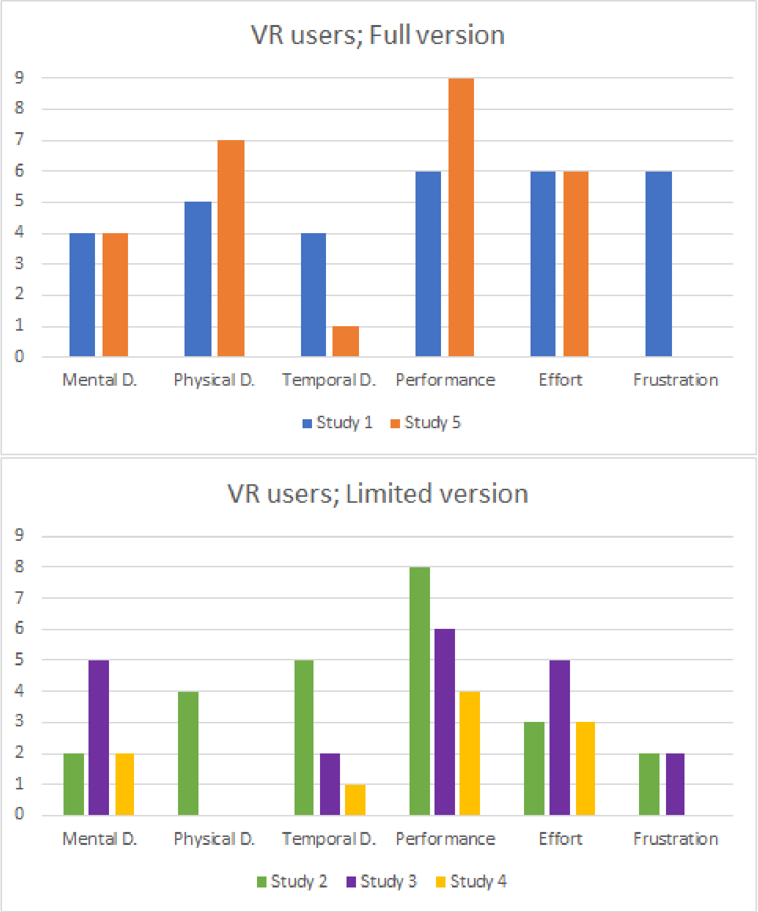

Figure 8.3: TLX questionnaire results of VR users using the full and limited versions

8.4 Questionnaire

Figure 8.3 shows all the results of the VR user's TLX questionnaires. The answers ranged from 0, least demanding, to 9, most demanding. Overall the results for VR users are similar to each other, which can be expected because the limited version mostly affects the smartphone user. The only difference is that instructions can only make use verbal communication which overall seemed to result in a slightly lower judgment of the workload. Figure 8.4 shows the results of the smartphone users. Notably the limited version seems to exhibit on average higher ratings for frustration, while mental demand is slightly lower. The lack of non-verbal communication is a probable cause for the higher rates of frustration. Some of the comments and other observations indicate that users had difficulty describing complex shapes and orientations to each other. The additional functions of the full version can likely explain the higher results for mental demand and effort.

Figure 8.5 shows the averages of all answers by user types. Showing that only VR users who were assisted with the limited version of the system experienced significantly lower workloads.

In general, the smartphone users experienced higher workloads when using the full version. It seems the additional options provided in the full system was more difficult to manage. From observations it seemed that especially switching the instructions on and off made it more difficult to compare between views.

The overall higher effort and mental demand from SA users may be attributed to the task solution being more difficult for them. In practice, SA users would talk the most while VR users only followed instructions.

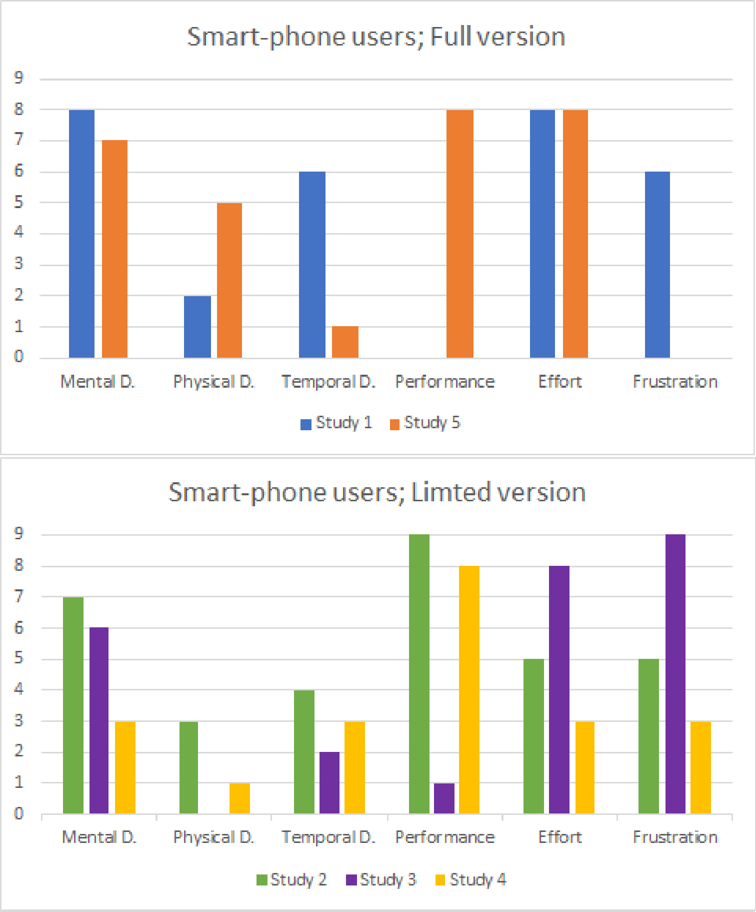

Figure 8.4: TLX questionnaire results for SA using the full and limited versions

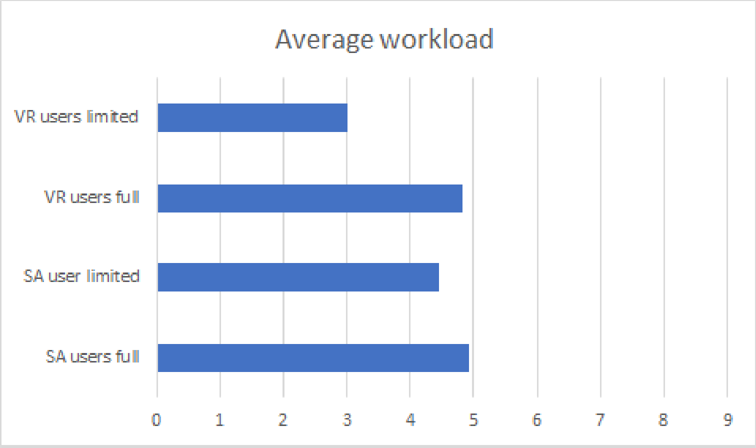

Figure 8.5: TLX questionnaire results for SA using the full and limited versions

Student's t-test analysis To determine if the significance of the difference in workload, we perform a two-sample students t-test across the tested groups. Simply comparing means does not consider variation, it is susceptible to misinterpretation due to outliers in the dataset. The Student's t-test is a more reliable way of determining the difference between two sets of data than simply comparing mean values. By also considering the spread of the standard deviation the test can provide a value for the probability p of the significance. This probability is an estimate how likely it is, that the difference is not caused by random deviations. The value has a range from 0 to 1 where 1 would indicate random data and 0 absolute certainty of significance.

VR users from group A will be compared against VR users of group B, while SA user from group A are compared against SA users from group B.

Applying the independent two-sample T-Test to compare the ratings between smartphone users in opposing groups resulted the values t(21)=0.41 and p=0.69, indicating a certainty for significance of 31 percent. This low certainty indicates, that the difference in mean is most likely a result of noise in the values. The test for VR users resulted indicates a higher significance with a value of t(22)=2.066 and p = 0.05 resulting in a certainty of 95 percent.The threshold for statistical significance is usually defined at between 0.1 and 0.01. In the case of the VR users, the result are significant, but in the direction opposite from what we expected. For VR users the cognitive load seems to be significantly reduced when using the limited version. It is difficult to definitively explain this result, since arguably, the VR user is less severely affected when using the limited version than the SA user.

One aspect we expected to increase cognitive load in the participants, was an increased amount of verbal communication. The results, however, indicate an increase in verbal communication (see figure 8.1) for users of the limited version, who also experienced a decreased cognitive load. In both groups, the VR user communicated less overall than the SA users did, when using the limited version. However, this was increased similarly for both users when using the limited version. Here, the VA users frequently asked simple questions, while gesturing or demonstrating actions. The results do indicate an increase in verbal communication from the VR users when using the limited version, but we do not observe a corresponding increase in cognitive load.

However it is important to keep in mind, that because of the low sample count of the study outside influences could have affected this outcome. Variables like the prior experience with VR in the participants, the understandability of the explanation provided by the study runner, or differences on how the participants judged their experience could have an amplified effect on the results, when only a few participants are tested.

Overall it is hard to make conclusive statements though, because of the limited amount of data.

8.5 Other Observations

During testing, some additional interesting observations could be made based on comments participants made or from analyzing the recorded videos.

In the context of the study, it was difficult for users to immediately understand all the features of the smartphone app. When everything was explained at once, sometimes participants forgot they could use a feature by the time it would be useful to them.

One feature that participants easily forgot about was the camera switch button on the smartphone application. The button would toggle between the two control methods for the camera. By default the camera was in motion-tracked mode, so the user could control his perspective by moving the phone. The option to switch to the a touch-controlled camera, was demonstrated and explained, but most users chose to keep using the default method. Later in the test, some participants attempted to rotate the perspective with a swipe gesture, but forgot how to toggle the control method, leading them to ask for help. Because of conventions in control methods for touchscreens users expected to be able to use swipe and pinch gestures, which were supported but required an additional action to activate them. This seems to have lead to additional confusion, which may have reduced the usability of the application. A possible solution for this problem may be to automatically switch to the gesture controlled when a gesture or pinch gesture is detected. The option to switch back to the motion-tracked mode could only appear when the gesture controlled mode is activated, giving a better indication of the state of the program.

Some participants found the representation of the smartphone user unintuitive. The fact that the smartphone user is represented by a floating phone needed to be explained for the participants to make use of the awareness of intent it provides. Even after understanding the concept they often did not pay much attention to it and only noticed it when it blocked their view. Making the avatar more human-like could be more intuitive even if the position might not be possible to track entirely accurate.

Even for users of the full version of the smartphone app often eventually discovered, that the monitor in the room displayed the perspective of the VR user. This second perspective would help them to understand what the other is looking at better.

Overall, the participants expressed, that they had the most difficulty explaining where and at which orientation pieces should go. While the pointer system helped in pointing out a specific piece or direction, it was less helpful in explaining an orientation. Participants sometimes used multiple pointers to draw the shape in the correct orientation, but it would often not be easily understood by the VR user.

While the task was designed to require a similar amount of effort from both types of users, ensuring this in practice was difficult. From observations during the study, in some cases the VR users were more passive. They tended to prefer following orders without giving much verbal feedback. The system gives VR users more opportunities for consequential communication as opposed to SA users. Being able to see the VR user's full body gave the SA users a higher awareness, allowing them a better understanding of what the other was doing. This kind of consequential communication might have reduced the need for intentional communication like pointing or verbal instructions. VR users could often just demonstrate the movement to place a piece and ask if it was correct.

While this effect was reduced when using the limited version, questionnaire results indicate a lower workload in this case, especially for the VR users. Even though participants were more actively and verbally communicating with each other, this did not seem to have increased their assessment for workload. One possible explanation is that VR users, when being constantly instructed, were not engaging directly with the puzzle but just following commands without thinking of a solution themselves. Some in-person observations indicated this, where at some points the VR user was just trying all possible rotations and asking the other participant to stop him when it was correct.

9 Conclusions and Future Work

To facilitate awareness between a VR user and a bystander, we developed a prototype to allow bystanders to interact with the virtual environment without requiring them to also use a headset. To asses the prototypes effectiveness in improving collaboration between two users, we conducted a study where participants were tasked to solve a simple puzzle together. While the study did not conclusively prove that the system improves performance in the chosen task, some valuable observations that can be made. Since the number of participants is low, some of the trends we observed are likely not predictive. Some of the results may still help steer the technology in the right direction or help improve the study design for future experiments.

To determine the success of the study we review the hypotheses that were created at the beginning of the study.

- H1: Participants assisted with the smartphone application need less time to complete tasks and make less mistakes than participants who are assisted only by voice.
- H2: Participants using the mobile-app experience a lower cognitive load when assisting another user than others only assisting by voice.
- H3: VR users also experience a lower cognitive load when being assisted using the app.

When reviewing the first hypothesis (H1), we can not determine an improvement in performance when participants used the full version. The data suggests the opposite. The results seemed to be largely dependent on the ability of the individuals to cooperate effectively. Some participants had a significantly easier time describing shapes and orientations or understanding the 3D geometry than others did. This variance may be mitigated by testing multiple tasks that require different skills or by expanding the study's scope to more participants.

The results for the second hypothesis (H2) are similar but less pronounced. Again the cognitive load that smartphone user experienced was slightly higher when using the full system than when using the limited system. This might be attributed to usability issues with the user interface design. Additionally, the amount of options might have been overwhelming to participants. A longer testing period per participant may give more time to become familiar with the controls.

The third hypothesis (H3) could not be proven with a significant margin. Cognitive load was slightly increased when VR users were assisted with the full system.

While the application presented the instructions identically to both groups of users, the complexity of accessing them was increased for the group testing the full version. When using the full version, users had to access the instructions as well as the virtual environment through the single screen of the smartphone. To accommodate this, the application is able to toggle between the two modes of showing instructions and showing the virtual environment. This, however, requires users to actively press a button to switch between the two modes while also being aware of the state the application. Observations seem to suggest that participants more readily kept the solution in mind when it was displayed in a separate screen. When participants of the limited smartphone application were looking up at the monitor and down a the screen to see the instructions, they seemed to be more effective than switching the view via an onscreen button. The choice to make the instructions accessible with a button press might have caused higher workloads for users of the full smartphone application. A future study could place the instructions on a separate monitor or on paper to reduce the amount of functions the participants need to learn to use effectively.

The observation, that some participants found it useful to look at the monitor even when they could use the full version of the smartphone app, also suggests that providing more perspectives could improve efficiency. Functions could be added that allow the virtual camera to follow the VR user's perspective.

There is also room to improve the usability of the pointer system. To prevent users from accidentally triggering pointers, the area of the screen that registers the touch input could be reduced in size. This could prevent touches on the edges of the screen to be triggered when holding the phone.

While VR users had no direct difference in user experience between the groups, the way they could be assisted by the smartphone user was changed. VR users could only receive outside information verbally when using the limited version, since they could not see the other person. Shapes, directions, and positions had to be described without the use of pointers or gestures. Even though our expectation was that these limitations would increase the cognitive load for VR users, the results do not show the expected improvement. The cause might be a flaw in the design of the task.

Observations showed that it was not necessary for the VR user to actively search for solutions. Some participants resorted to blindly following orders without giving much of their own inputs. When assisted only verbally, participants tended to focus only on trying to understand what the SA user was explaining without trying to find their own solution. This could have decreased the results for cognitive load for VR users using the limited version.

Feedback also suggests that using a floating phone to represent other users was unintuitive. Representing smartphone users as simple human-like figures behind the phone could make the interaction more natural. Without an additional tracking solution however, this solution is only able to approximate the position of the SA user. Optimally, finding a way to track a bystander's pose or gestures would allow this avatar to convey more information about the person. Camera-based body tracking could make it possible to insert a more detailed avatar into the virtual environment, giving VR users a better awareness of others in the room.

While adding additional trackers could make this avatar more accurate or even make natural gestures possible, the practicality may limit it's usefulness. Instead, the avatar's position could be based on the phone, which can already be tracked. Placing a simple figure behind the phone may help VR users to better understand where the other person in the room is. This solution might however have a slight mismatch with the actual position of the person, especially if they would place the phone down. We expect a small mismatch would not be a problem, and to avoid the larger mismatch, the program could be designed to detect whether the user is holding the phone by checking for movements of the tracker. Another method of tracking bystanders could be achieved by using additional tracking systems. For example, a camera attached to the HMD could be used to detect faces in the VR user's field of view. The position could be estimated by using a stereoscopic camera. By placing avatars at the positions where faces were detected, every person in the room could be represented in the virtual environment without the need of any additional tracking devices.

Using a tracker attached to the phone is the most intuitive way to use the prototype; however, this solution could be impractical for many use cases. The Vive Tracker is not included in standard sets when buying the headset and is not compatible with many other brands of HMD's which might not have an equivalent solution available. In a real-world application, users are unlikely to have access to a tracker or a means to attach it to a phone. Optimally, the phone application could determine it's own position in the room, without the help of additional hardware. Tracking based on camera and gyroscope information is a possibility that is already used in many augmented reality applications on smartphones. This could help in solving the problem; however, this method of tracking would not know the phones own position in relation to the HMD. This could result in even more confusion by causing the avatar of the VR user to appear in a position not related to the position in the real world. Maintaining the relationship between real-world and virtual objects would require both tracking systems to know the common positions of points in the room.

Overall, the prototype proved functional but we could not prove significant advantages. Some refinement and further development of the system is necessary to make it a viable option for effective collaboration.

Contents of the attached CD

- VRLink.zip - Source Project
- VRLinkBuild.zip - Runnable VR application
- VRLinkAndroid.zip - Runnable android application.
- ReadMe.txt - Instructions
- Bachelor.pdf - pdf version of the thesis

References

[1] Christian R Larsen, Jette L Soerensen: " Effect of virtual reality training on laparoscopic surgery: randomised controlled trial". Research article, 2009

[2] Thierry Duval, Cédric Fleury, Bernard Nouailhas, Laurent Aguerreche: *"Collaborative Exploration of 3D Scientific Data"*. 2008

[3] Carl Gutwin, Saul Greenberg: *"A Framework of Awareness for Small Groups in Shared-Workspace Groupware"*. University of Saskatchewan, University of Calgary ,1999

[4] U. Neisser: *"Cognition and Reality"*. 1976

[5] K.M. Inkpen, M.S. Hancock, R.L. Mandryk, S.D. Scott : *"Collaboration Around a Tabletop Display: Supporting Interpersonal Interactions"*. Dalhousie University , Simon Fraser University , 2002

[6] C.Mai, L. Rambold, M. Khamis: *TransparentHMD: revealing the HMD user's face to bystanders"*. 2017

Web-References

[7] *"NASA TLX: Task Load Index"* `https://humansystems.arc.nasa.gov/groups/TLX/`, accessed July 18, 2020.

[8] Stereolabs:*"Unity - Green Screen VR Capture"* `https://www.stereolabs.com/docs/unity/green-screen-vr/`, accessed December 19, 2020.

[9] Valve Corporation:*"SteamVR Plugin"* `https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647`, accessed December 19, 2020.

[10] Unity Technologies:*"Unity"* `https://unity.com/`, accessed December 19, 2020.

[11] Unity Technologies:*"The Multiplayer High Level API"* `https://docs.unity3d.com/Manual/UNetUsingHLAPI.html/`, accessed December 26, 2020.