DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics: Games Engineering

A Calibration Tool to Establish Interpersonal Synchrony Using a Motion-Based Multiplayer WebAR Game

Jonas Goos

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics: Games Engineering

A Calibration Tool to Establish Interpersonal Synchrony Using a Motion-Based Multiplayer WebAR Game

Kalibrierung interpersoneller Synchronität mit Hilfe eines bewegungsbasierten Multiplayer-WebAR-Spiels

Author:	Jonas Goos
Supervisor:	Prof. Gudrun Klinker, PhD.
Advisor:	Chloe Eghtebas
Submission Date:	15.09.2021

I confirm that this bachelor's thesis in informatics: games engineering is my own work and I have documented all sources and material used.

Munich, 15.09.2021 Jonas Goos

Abstract

Not all experiences with video conferencing are positive. We are trying to use the concept of interpersonal synchrony, which has many social benefits, to compensate for the negative impacts which the COVID-19 pandemic has made us deal with, such as not being able to meet in person. In order to induce synchrony between people in a video conference, we created a multiplayer WebAR game, which takes part in this setting. The game is movement based as players must shake their head and hands. A repeated measures user study with seven dyads was designed which compared a sandbox, synchronous and asynchronous mode in real-life conditions. The results are not conclusive in showing that movement synchrony was achieved. However, the experience in general had positive effects on the participants, like increased feeling of connectedness and overall good enjoyment. Technical issues such as the performance of the game played a big role on the outcome of the study. This game shows that the concept of multiplayer AR games is promising and should be explored further, though under more controlled hardware conditions.

Zusammenfassung

Die Erfahrungen mit Videokonferenzen sind nicht immer positiv. Deshalb versuchen wir diese mit Hilfe von interpersoneller Synchronität, welche viele positive soziale Effekte hat, zu verbessern. Damit könnten auch Einschränkungen der COVID-19 Pandemie, wie zum Beispiel Kontaktbeschränkungen, überwunden werden. Um Synchronität herbeizuführen, wurde ein Multiplayer-WebAR-Spiel entwickelt, welches in einem Videokonferenz-Setting stattfindet. Das Spiel, in dem man seine Hände und seinen Kopf schütteln muss, ist bewegungsbasiert. Es wurde eine Studie mit Messreihenwiederholung und sieben Spielerpaarungen erstellt, die eine Sandbox, einen synchronen und einen asynchronen Modus unter Praxisbedingungen miteinander vergleicht. Die Ergebnisse zeigen nicht eindeutig, ob Bewegungssynchronität erzielt wurde. Dennoch hatte das Experiment positive Effekte, wie ein erhöhtes Gefühl der Verbundenheit und ein insgesamt positives Erlebnis, auf die Teilnehmer. Technische Probleme, wie die Performanz des Spiels, hatten einen großen Einfluss auf den Verlauf der Studie. Es wurde gezeigt, das das Konzept von Multiplayer-AR-Spielen vielversprechend ist und die Idee in Zukunft weiterverfolgt werden sollte, dann jedoch unter besser kontrollierten Bedingungen.

Contents

1. Introduction

Imagine you are about to meet a new colleague or project partner but not in-person, but rather via a video conferencing tool like *Zoom*[1] or *Discord*[2]. This is a scenario which is happening increasingly often, especially since the beginning of the COVID-19 pandemic. The problem is that video conferences can be awkward, especially when meeting for the first time. Part of the reason why, is that communication is not just words, and also a video call cannot replicate all facets of a real interaction. One phenomenon in interactions, that is often unnoticed, is interpersonal synchrony. Interpersonal synchrony happens naturally when humans interact with each other and means that some form of behavior synchronizes between them [1]. It has been studied in group settings [2], between individuals [3–5] and also in VR [6, 7]. It has also been shown that it can affect how likable other persons are perceived or how likely it is for us to cooperate with them [2, 3, 8].

The goal of this thesis is to get more insight into synchrony in video conferencing settings, especially between new acquaintances. The use of video conferencing tools has increased in the last years even before the COVID-19 pandemic began [9]. During the pandemic its use was and still is ubiquitous as a lot of work and school has to be done remotely. That means that many interactions that normally happen face-to-face are now happening virtually. Therefore, it is important to understand what implications this has. To investigate this, this thesis focuses on interpersonal synchrony. Studying it in a virtual conference setting is a novel matter and what is not clearly known is how to reliably induce it in a controlled way.

To study this, we wanted to design and create a type of game that can function as an icebreaker between new colleagues or project partners and that can be used to induce synchrony. Our game should be developed in WebAR, because it is important that the players can see each other in a conference-like setting and we just augment their video feeds with game components. This game could be used as a calibration tool when determining the amount of synchrony between two people. So, for example, two people would play the game and would then have a high amount of synchrony. Then they could be observed talking to each other or doing some work together. It could then be determined if the amount of synchrony dropped while they are performing

[1] `https://explore.zoom.us/en/products/meetings/` (accessed on: 09.09.2021)
[2] `https://discord.com/` (accessed on: 09.09.2021)

this task. If this calibration step is omitted, it is hard to know if the synchronization is an effect of the task or if they are just naturally synchronized.

More generally, the game would try to improve the experience of a video conference. Especially during the COVID-19 pandemic, the use of video conferencing tools has been associated with negative experiences [10]. The term "zoom fatigue" has been created to describe this phenomenon, which might actually be directly related to synchrony [11]. This game could be used as a way of making the experience more fun and improving the mood, similar to a group game, which is played before a group exercise.

What this thesis wants to investigate, is how synchrony affects video conferencing settings and more importantly what aspects are needed to induce it. In order to do this, first literature has been reviewed to get an overview as to what other work has been done on this topic. The focus here was synchrony in games and also generally multiplayer AR applications. Next a game was designed which employs game mechanics with the aim to induce synchrony in the players. Then the implementation of the game is described and different considerations that had to be made are explained. In order to evaluate the game, a user study has been carried out, the results of which are presented and discussed.

2. Related Work

Because interpersonal synchrony is a very important part of this thesis, a short introduction is necessary. Afterwards synchrony in virtual contexts is examined more closely. Finally, AR applications that involve multiple people are discussed.

2.1. Interpersonal Synchrony

Interpersonal synchrony describes behavior between people that is matched in time [1]. Behavior is very broadly defined here as it can vary from heartbeats to walking [1]. A distinction has to be made between behavior matching (also called behavioral mimicry) and interpersonal (also called interactional) synchrony, both of which are a type of interpersonal coordination [1]. The main difference is the timing. Synchrony only occurs if the behavior happens simultaneously, while behavior matching or mimicry means copying of behavior, so there is a time difference [1]. This means, for synchrony to work, people must be able to predict the motion of others [12].

2.1.1. Examples of Synchronization

It has been observed that people synchronize with each other while interacting. Examples of this include synchronization of eye blinks and head nods during conversations even if the partners are not looking at each other [5]. It was also observed that people synchronize phase and speed when sitting in a rocking chair and staring at each other, even when they are told to rock at their own pace [13]. Similarly, a study in [14] was performed where dyads walked side-by-side in various conditions. It was found that half of the dyads who had tactile feedback (i.e. holding hands) walked in synchrony. The above-mentioned examples are types of movement synchrony, but there are also types that are not directly related to movement. A study performed in [15] found that the heart rate of participants synchronized while playing a public goods game, which is a game where participants can either contribute tokes to a public pot or keep them for themselves. Furthermore, in the same study, there was a correlation between heart rate synchrony and return in the game. These observations raise the question why this behavior is observed.

2.1.2. Positive Effects

Making people move in synchrony has been observed to lead to positive effects. In [2] participants, who walked in-step with each other, felt more connected than those who did not. In the same study, participants also took part in a task where they had to move cups and sing along to a song. Those who sang and moved the cups together felt higher team association and also were more cooperative in a following weak link coordination exercise. Another study involving tapping a finger on a drum in and out of synchrony with an experimenter, showed that the participants liked the experimenter more if they were tapping in sync with them [3]. In [4] a study was conducted that showed that rocking a chair in sync with a partner increased their performance in a joint-action task that was performed afterwards. Dyads that rocked in sync needed less time to maneuver a steel ball through a labyrinth than dyads who rocked their chairs out of sync. These positive interpersonal results can already be seen in infants. A study with 14 month old infants showed that they were more likely to help the experimenter when they had bounced in synchrony with them before as opposed to out of synchrony [8]. Lastly, [16] linked synchronization of arm curls with another participant via a video call (which was actually a pre-recorded video of a person doing arm curls at a fixed pace) to higher self-esteem.

Even though many positive effects have been found, negative effects of synchrony have also been observed. In [17] participants took part in a complex joint task (i.e. building a LEGO car). It was observed that synchrony was negatively associated with task outcome (i.e. less aesthetic appeal, fewer pieces used). Also some negative associations with subjective perceptions (e.g. fun, cooperation) have been found, but only in some conditions while in others these were positively associated.

These results show that synchrony can have positive effects on social interactions, but not always on task performance (especially for complex tasks). But the social effects are the main interest of this thesis as they are often lacking in virtual contexts [18]. Therefore, increasing synchrony in such situations seems like a promising way to improve them overall.

2.2. Interpersonal Synchrony in Virtual Contexts

As we have already discussed, synchrony, although not fully understood, is well researched in real-world interactions. This is a bit different for virtual contexts, where only limited research has been done so far. This section outlines some virtual experiences and describes them and their findings.

In [19, 20] a virtual environment was created where the breathing of both participants was visualized using the position, scale and biolumniescence of a jellyfish. Additionally, a glass sponge could be grown if the participants breathed in synchrony. No formal study on the effects has been done, but an initial survey at an art exhibition where the experience could be tested had some positive results. For example, most participants reported co-presence and engagement with their partner.

Another study [21] was conducted where a person was put in a virtual environment with two virtual characters. The movements of the person were replicated by two virtual humans in the VR environment, but with varying amounts of delay. The study found that participants where the virtual characters matched their movements more closely reported greater social closeness.

In [6] a study was done where participants needed to perform a finger pointing task in a virtual as well as a real environment. In both cases synchrony could be measured using hyperscanning, a technique to measure neural activity. This suggests that synchrony has similar effects in virtual and real environments.

[22] describes SynchroMouse, a game that has been created in which players must synchronize their mouse movements. The goal of the game is to move the mouse over the screen as much as possible while staying in sync (same position and movement) with the other players. But no research has been done as to whether this game has any effect on interpersonal synchrony.

In [23] another example of interpersonal synchrony combined with gaming is given. The paper describes an exergame (i.e. exercise game), called ExerSync, in which two people can exercise in synchrony, even when they are using different exercise devices (e.g. stationary bicycle, treadmill) or have different preferred exercise speeds. To do that, the rhythm of a player is extracted, translated into another frequency range (selected by the player) and then visualized to the other player. It was found that using audio-visual rhythmic cues lead to lower perceived workload than just using a speed gauge. But the actual synchronization was not different between the two conditions.

The mentioned research shows that real-world presence is not necessarily required to achieve synchrony or to experience positive effects from it. Up until now there has been no research done about synchrony in video conferencing environments and that is what this thesis wants to investigate.

2.3. Multiplayer AR

Augmented reality (AR) is a technology whose use has steadily increased over the past years. It is used in games, the most prominent example being *Pokemon Go*[1], but also in serious applications like room planning [24, 25]. What is yet not common are

[1] `https://pokemongolive.com/en/` (accessed on: 09.09.2021)

AR multiplayer or online games. This section gives an overview of the research and examples that are available.

In [26] a mixed reality trampoline game, called Super Stomp, is described. The game tracks two players who are jumping on separate trampolines and puts a video feed of their bodies into a virtual environment. The movement of the players is exaggerated in horizontal as well as vertical direction and the goal is to stomp the other player. In a small trial, participants stated high levels of closeness and overall the game was perceived positively.

But there are also some AR projects which work online/remotely (like this game will). An example of this would be Co-sound [27]. It consists of a shared space to which multiple people have access. Everyone can manipulate objects in the space and also trigger audio events which are heard by all users. The project had an overall high acceptance and demonstrates that *WebRTC*[2], and peer-to-peer technology that is discussed later, is suitable for this kind of application.

Another example is described in [28]. In this paper the authors expand a classical video conferencing environment with a second camera feed for each user. In this camera feed, users can put their hand or other objects, which are then automatically cropped and placed over the image of the other person. This leads to an experience which offers more interactability than traditional video conferencing setups. It was also shown that it leads to increased social presence compared to the traditional setup.

2.4. Summary

As can be seen, interpersonal synchrony can have some positive effects, which we want our game to benefit from. Additionally, there have been other games that tried to induce synchrony, some of which succeeded. Most of them used movement synchrony, which will also be used in this game. The reasons for this is that it can be easily detected in a video image and the movement makes for a good game mechanic. Next, the game itself will be described in more detail.

[2] `https://webrtc.org/` (accessed on: 09.09.2021)

3. Prototype

This chapter describes what we designed and built to "calibrate" synchrony between people.

3.1. Game Design

The goal was to design and build a WebAR game to achieve the purpose of movement synchrony. This section describes how the process of designing this game proceeded.

3.1.1. Idea

The idea was to develop a game that has a similar setting to a two person video conference. So there would be two camera feeds, side-by-side, where both players can see each other. Additionally, we wanted to include some AR augmentations to the images to make the experience more immersive. Furthermore, the game had to involve some kind of movement that would be suitable to invoke synchrony between the players.

The focus of this project was accessibility. So as many people as possible should be able to use it. This was done because video conferencing is used by everyone. Therefore, it seems reasonable that such a game, which aims to improve the experience, should also be available to as many people as possible. Subsequently, it was decided only to use a camera and to make the game web-based. This way the game is accessible to the largest number of people as the required technology is present in almost all households and no installation is required. Although this also means that WebAR needs to be used, which is generally less capable than native optimized AR implementations.

Crazy 8

To get some ideas my advisor and I used a brainstorming technique called Crazy 8. Both of us had a paper with 8 segments and had six minutes to fill these segments with sketches of ideas. During this we should not regard any technical limitations whatsoever. We then ended up with 14 different ideas as not all segments could be filled within the given time. To evaluate which ideas to peruse, the ideas were rated in

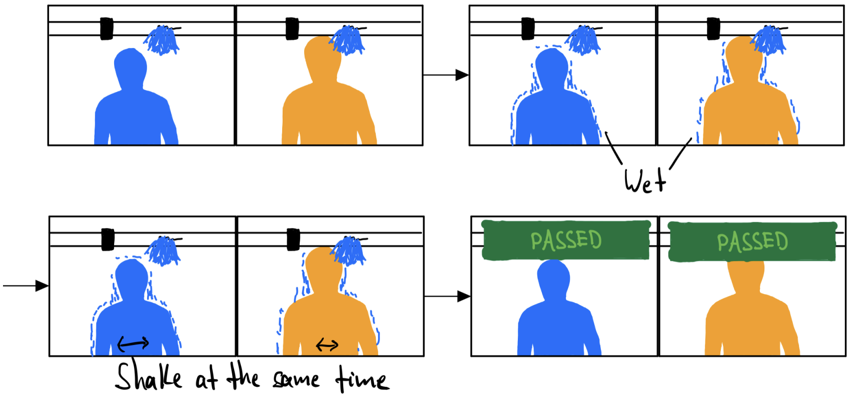


Figure 3.1.: Example of the storyboards that have been created for *Water Spill*

seven mostly subjective categories. These included, for example, technical feasibility, amount of invoked synchrony, fun and which technologies would be needed. All of the parameters were estimated based on prior research and experience. The whole table of the evaluation can be found in the appendix (Table A.1). Of the 14 ideas the three best performing were selected and these are described in more detail below.

Asteroids Both video feeds of the players are overlapping, so the player can see themselves and their partner in the "same" video feed. Then an asteroid falls from the sky and they must get out of its way, although they are also not allowed to overlap with each other.

Water Spill Water spills in a zero gravity environment. The water sticks to the players and they must shake at the same time to get rid of the water.

Rhythmic Task The players must move in a rhythm and match each others movements, but they do not always see each other. In the periods where they do not see each other they must still try to move accordingly. The rhythm should help them with that.

Storyboards

To further develop the ideas, storyboards were made for the ideas *Asteroids* and *Water Spill*. These include the basic game idea as well as ideas for expanding them. An example can be seen in Fig. 3.1. All storyboards can be found in the appendix (Fig. A.1). No storyboards for *Rhythmic task* have been done because the *Water Spill* idea was already liked a lot and *Rhythmic task* was still very vague and had no real gameplay yet.

Final Idea

The *Water Spill* idea was then chosen over *Asteroids* because it incorporated more movement and thus synchrony. It was decided that there should be three parts of the body that can get wet and that they need to be shaken in a specific sequence. The selected parts are: head, left hand, right hand. These parts were chosen because they are visible in the camera and should be reasonably easy to detect (see section 3.3).

3.1.2. Mechanics

In this section the specific mechanics of the game are described. The game is round based. In every round all three of the body parts of both players are wet. They need to shake them to get the water off. But this shaking must happen in a specific pattern given by the game. There are two different game modes which differ in the way the players are instructed to shake. These are explained after we discuss the basic common mechanics below. If they shake successfully, the water droplets will be suspended in the air and float around (because the players are in a spaceship). Then a vent will activate and suck them out of the ship.

For every body part that needs to be shaken, there is a time limit in which both players need to do so. If they manage to do it before the time runs out, the humidity in the spaceship decreases. If they cannot make it in time, the it increases. The available time is inversely logarithmic with regards to the rounds. Hence, the further the players progress the less time they have. If the humidity goes over 90% they lose the game, because their equipment breaks. The number of rounds they successfully complete will then be their score.

3.1.3. UI

As mentioned the game consists of two camera feeds. The own player 's camera feed is located on the left of the screen, while the one on the right belongs to the other player. An example of the website with the game can be seen in Fig. 3.2.

Most of the UI is centered around the player's own camera feed. In the top left there is a timer in red (see Fig. 3.2), which indicates how much time there is left to shake the current body part. At the bottom there is a bar which indicates how high the humidity is and at which point the game will be lost. At the top there is a vent where the suspended droplets will move to. Its activation is indicated by a turning propeller.

Game events, like new rounds or detected shakes, are indicated by a text in the middle of the screen (see Fig. 3.3). Above both camera feeds are the instructions, about the order in which the parts need to be shaken. It is also indicated what the current part is and what parts already have been completed.

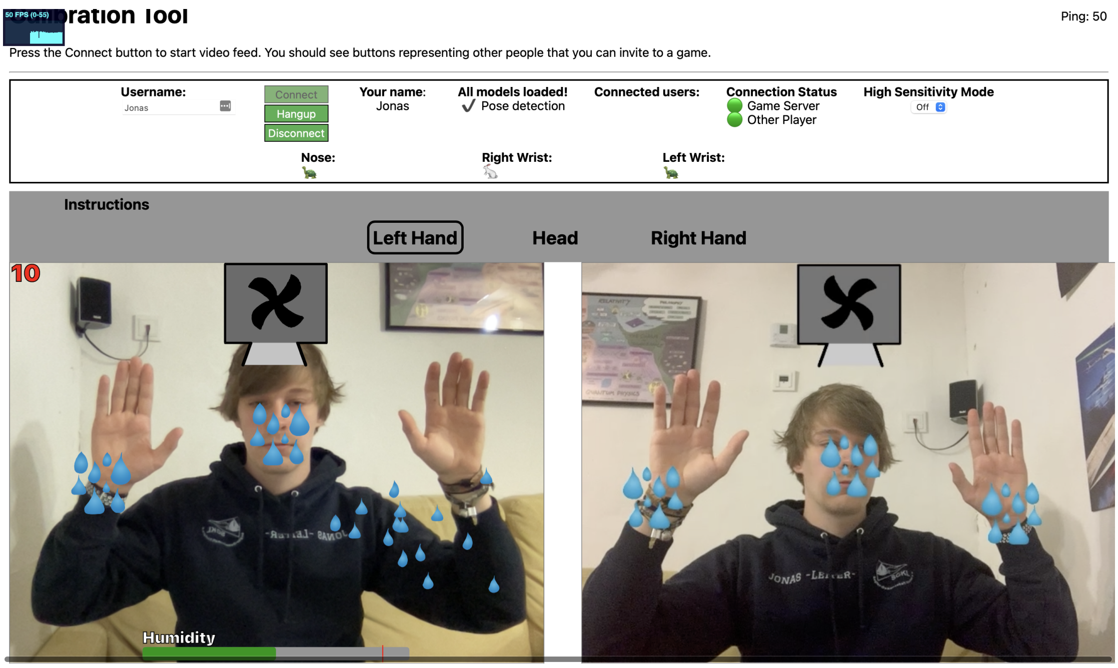


Figure 3.2.: Screenshot of running game
All three body parts of the right player are wet, while for the left player only head and left hand are wet. The right hand has already been shaken and the droplets are suspended in the air. The humidity bar can be seen at the bottom and the vent and timer at the top. Above the two video feeds instructions can be seen. First the left hand must be shaken, then the head and last the right hand. The current objective is to shake the left hand.

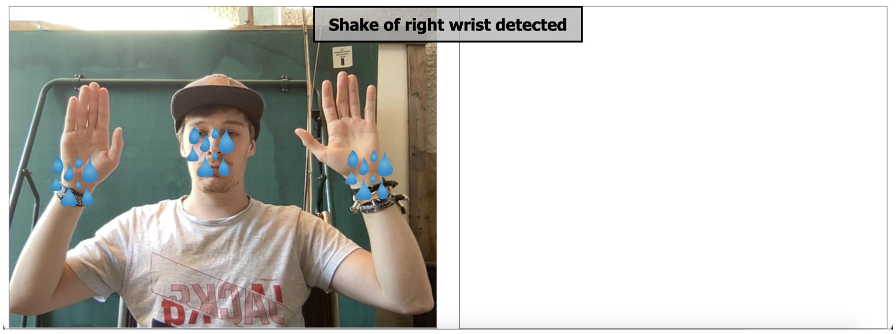


Figure 3.3.: Screenshot of a “shake detected” message

The camera feed of the player itself is mirrored while the camera feed of the other player is not. This is the same behavior as in most video conferencing platforms. Mirroring the player's own video makes seeing yourself more natural, because it is the same as if you were to look in a mirror. By not mirroring the other player, it is as if you were sitting face-to-face with them. By sticking to the most common configuration, it is hoped that mix-ups between left and right will be limited.

3.1.4. Augmentations

To indicate to the players which parts are wet and which are not, augmented water droplets are used. When a part is wet, a bunch of droplets are displayed over it. As soon as a legal shake is detected, the water droplets leave the body part and are now suspended in the air and move randomly. An example of this can be seen on the left of Fig. 3.2. When the vent activates all currently suspended droplets are drawn towards it and are then removed. The augmented water droplets only follow the 2D movement of the body parts (i.e. they do not get smaller when the hands move further away; see section 3.5 for more details).

3.1.5. Modes

As this thesis wants to investigate the effects of synchronized movement, the game needs a way of forcing the players to do that. Specifically, the shaking should happen at the same time. Additionally, a second mode was introduced where the shaking did not need to be at the same time. Both modes are described in the following section.

Synchronous Mode

In the synchronous mode (SM) both players have to shake the same body parts at the same time. Then the vent activates and the part is completed. However, the sequence in which they need to shake the three parts is only shown to one player, who needs to tell the other one what to do. This was done so the players had to engage with each other more and also in the hope that it would lead to the players looking at the other person instead of only at themselves.

For a shake of both players to count as synchronous, both shakes need to be within one second of each other. To determine the time between the shakes, the start of them is taken as reference. It was decided that this was the best option because it is rather easy for the players to time it that way. The end of shakes can vary because it takes different amounts of time on different setups until shakes are detected. In Fig. 3.4 this is illustrated again.

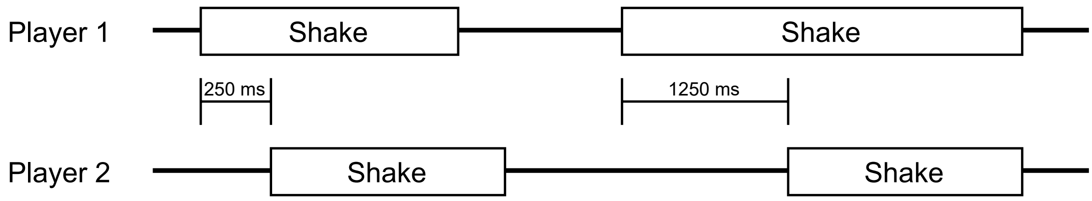


Figure 3.4.: Illustration of synchronous shake detection
The first shake will be counted as synchronous while the second one will not. Even though players stop shaking at the same time, the starts are too far apart.

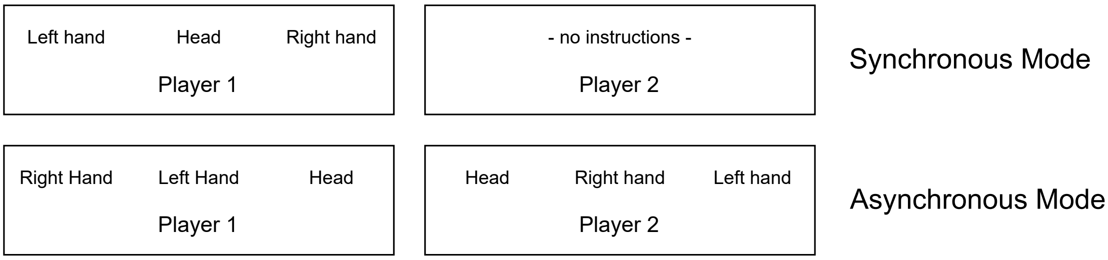


Figure 3.5.: Example illustration of instructions for both game modes

Asynchronous Mode

In the asynchronous mode (AM) the shakes do not need to be at the same time and also the body parts that have to be shaken are always different from each other. So the players never need to shake the same part. When both players have completed their shake, the vent activates and the part is complete.

This mode was chosen as such, because it has the maximum amount of asynchrony compared to the SM. There is no synchrony in the timing of the shakes as well as what the players are shaking.

As both players now have different sequences both sequences need to be displayed. It was decided that the sequence should be displayed to the other player respectively. Consequently, each player must tell the other one what to do. This was done to keep it similar to the other condition where the players need to engage with each other. Fig. 3.5 shows an example of how the instructions differ between the modes.

3.1.6. Accessing the Game

As with every online game, the players need a way of connecting to the game and each other. The process of how this works in this game is described in this section.

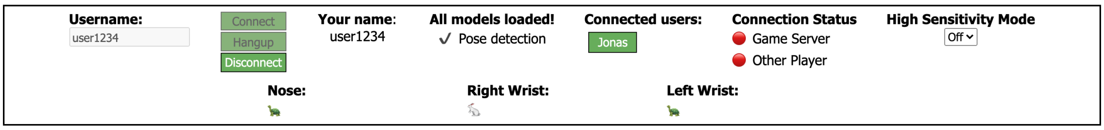


Figure 3.6.: Screenshot of the control bar

When a user opens the website, the first thing they need to do is to enter a username and press connect in the control bar (see Fig. 3.6). This means they will register themselves on the signaling server and can then be found by other people. Additionally, all models that are needed are then loaded to avoid delays later in the game. Once the user is connected, they need to give access to their camera and are then able to see themselves with water droplet augmentations on the left of the screen. They also see a list of other users which they can invite to a game by clicking their names. If they do so, the other user will get an invitation which they can either accept or decline. The control bar also displays the connections status to the game server of both players (see section 3.2) as well as information about the movement speed of the three body parts. This information is mostly displayed for debug purposes, so the experimenter can identify problems better and advice the players on what they need to do differently. The last item in the control bar is a toggle to make the shake detection more sensitive (see section 3.3).

3.2. Networking

As the game is a multiplayer game that should be playable in the browser, there is some networking that needs to be taken care of. It was decided that the game should be playable over the internet and not just in a local network. This was done to make it comparable to traditional video conferencing as well as to simplify the user study. This section explains how the communication between the players and the servers works during the game.

3.2.1. Overview

First of all, there are two parts of the system to consider here. On the one hand, there is the connection between the two players. It is responsible for the video chat between them. On the other hand, there is the connection between the players and the game server. It is responsible for communicating the game logic. Fig. 3.7 shows an overview of the connections during a game.

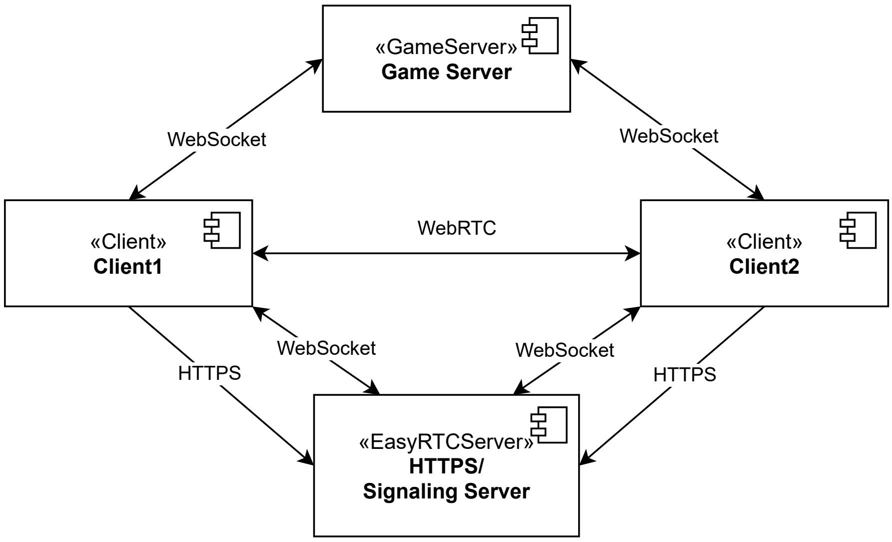


Figure 3.7.: Deployment diagram of all involved server and clients during a game

3.2.2. Video Chat

First, we want to look at the video chat between the two players. It is handled by a WebRTC connection. WebRTC was chosen because it is fast [27], secure [29] and supported by almost all browsers [29]. WebRTC establishes a peer-to-peer (P2P) connection over which the video stream is sent. To simplify the usage of WebRTC the library *Open-EasyRTC*[1] is used. It handles the support of multiple browsers as well as signaling.

Signaling means establishing the connection between the two peers. It looks like the following. Both players connect to the signaling server. The signaling server then notifies all connected clients about all other connected clients. The user initiates a call by clicking the name of another user. This invitation is forwarded to the correct user via the signaling server. If the other user accepts the call, a direct connection between the two clients is attempted. For that the address details of both players are needed, which are determined by Open-EasyRTC using public STUN servers. These address details are exchanged via the signaling server. If a direct connection is not possible, a relay TURN server is used. Once the connection is established, the signaling server is no longer used and all the traffic between the two players is sent via the WebRTC P2P connection.

[1]`https://github.com/open-easyrtc/open-easyrtc` (accessed on: 09.09.2021)

3.2.3. Gameplay

Now we will take a look at how the players get updates about the game state. As seen in Fig. 3.7, both players are also connected to a separate game server using a WebSocket. This connection allows for full-duplex communication, which allows both the game server and the players to push updates. These updates include actions from the player (i.e. starting and finishing shaking) and changes in the game state (i.e. advancing, change in humidity), and are sent via JSON messages. The WebSocket connection is established once both players are connected to each other via WebRTC.

As the framework for the game server, *Play Framework*[2] in combination with *Java*[3] is used. Akka streams and actors, which are a functionality of Play to manage information flow, are used to handle the WebSocket connection.

3.3. Shake Detection

As the main mechanic of the game is shaking water off your body, a reliable method to detect when a player shakes is needed. This section describes how this problem was solved.

3.3.1. Library Selection

First of all, a method is needed to extract the position of the body parts from the camera image. Then, in later steps, the movement of the body parts can be analyzed and a decision can be made. For the first part, different libraries and technologies were considered.

At first *Handtrack.js*[4] was considered as it can detect hands and faces, and moreover it can also differentiate between hand poses (e.g. open, closed, pinch) which could possibly be used for an extension of the game. However, it cannot distinguish between left and right hands and was therefore not suitable.

Next the library *BodyBix*[5] was considered. It can give a pose estimation (PE) of the person in the video. That means it provides a list of 17 keypoints (e.g. nose, left wrist, right elbow) each with a position in the camera image as well as a score on how good the estimate is. In addition, it can do body part segmentation (BPS), which means that for every pixel in the image it can tell which of 24 body parts (e.g. left face, front of left upper arm) it belongs to. Using PE is suitable for the problem as it

[2] `https://www.playframework.com/` (accessed on: 09.09.2021)
[3] `https://www.java.com/en/` (accessed on: 09.09.2021)
[4] `https://github.com/victordibia/handtrack.js/` (accessed on: 09.09.2021)
[5] `https://github.com/tensorflow/tfjs-models/tree/master/body-pix` (accessed on: 09.09.2021)

means differentiation between left and right is possible. As keypoints, the nose has been selected for the head as it moves the most when shaking the head from side to side. For the arms the wrists have been chosen as keypoints for the same reason.

In the end *PoseNet*[6] was used. It is a pre-trained *TensorFlow*[7] machine learning model for PE. It works the same as the PE described above, but to a higher degree of accuracy than BodyPix. See section 3.5 for more details.

3.3.2. Approach

The next step is to take the positions of the three keypoints, analyze them and decide if a shake is currently taking place. The basic approach of how this is done is described in this section.

First of all, the accuracy score of the keypoint is considered. If it is too low, this measurement will be discarded. If it is high enough, the distance the keypoint has moved since the last frame is calculated. This distance is then normalized with respect to the width of the canvas, such that the distance is consistent over all camera resolutions. From this distance a speed is calculated by dividing it by the time that has passed since the last time the detection ran. This speed is then added to an array containing the speeds of the last 20 detections. Afterwards, the average of this array is calculated, which gives a smoothed speed of the given keypoint.

This smoothed speed is the main factor that is used for the shake detection. If it gets above a certain threshold, a shake is started which keeps going until the velocity drops below the threshold. If the velocity is high for long enough (500 ms), a shake is detected and the user is notified. Different thresholds for the head and the arms have been chosen as head movements are slower and shorter than arm movements.

The approach described above works fine, but can lead to many false positive especially on faster systems. Therefore, some refinements had to be made. The main problem was that many regular arm movements were detected as shakes (e.g. moving hand from one side of the screen to the other). This has two reasons: First, the keypoint only has to move fast for 500 ms to be detected, but increasing this time would lead to actual shakes needing to be too long and uncomfortable. Second, it was not considered if the keypoint was actually changing directions, which is the essence of shaking. Therefore, direction changes have been added to decrease the number of false positives.

To do that, first the velocity as a vector is calculated from the current and last positions as well as the time difference between them. Then the velocity is checked to see if it is faster than a given threshold. If it is not, it is discarded. This is done because noise in the measurements can lead to small movements and therefore velocities in

[6] `https://github.com/tensorflow/tfjs-models/tree/master/posenet` (accessed on: 09.09.2021)
[7] `https://www.tensorflow.org/` (accessed on: 09.09.2021)

arbitrary directions. After that, the dot product between the current and the last velocity vectors is taken to calculate the angle between them. If this angle is larger than 90 degrees, a change in direction is assumed.

This algorithm takes into account both properties of shaking, that is fast movements and changes of direction. Because of that, it leads to a reliable detection. However, because hand movements are quite fast, it does not work well on slower systems. This is because the sampling frequency gets too low and some direction changes will be missed. This leads either to shakes needing to be very long to be detected or to shakes not being detected at all. Intensive testing was done to adjust the parameters to make the detection algorithm work on different system setups. As a backup, an option was implemented to ignore the direction changes and only use speed for the detection. This option could be used if the detection was not sensitive enough otherwise.

3.4. Other Libraries

In order to complete the game experience some other libraries have been used.

3.4.1. Merging Video and Augmentations

Because not only the video of the player, but also the augmentations (e.g. water droplets) need to be sent to the other player, these must be combined into one video. The initial approach was to draw the video image on a canvas and then draw the augmentations on top of it on the same canvas. But this leads to sluggish video when the frame rate of the augmentations drops, which impacts the experience. To get around this, the library *video-stream-merger*[8] was used. It can combine multiple media streams into one. Using this feature, the media stream of the video element is combined with the media stream of the canvas, which is acquired using `HTMLCanvasElement.captureStream()`. This combined stream is then sent to the other player. Using this method, a possible slowdown of the PE and augmentations does not influence the frame rate of the video, because the draw loop of the canvas is completely separate from the video.

3.4.2. Stats

To assess the performance of the game on different browsers and devices, some means of measuring it was needed. For this the library *stats.js*[9] was used. Using its FPS monitor, the performance of the game loop (i.e. PE, draw call, shake detection) can be

[8] `https://github.com/t-mullen/video-stream-merger` (accessed on: 09.09.2021)
[9] `https://github.com/mrdoob/stats.js` (accessed on: 09.09.2021)

(a) Water augmentations using BPS[10]

(b) Example background using BPS

(c) Example usage of jeelizFaceFilter[11]

Figure 3.8.: Previous versions of augmentations that have been tried out

evaluated. This was very important to decide which browser should be used for the user study as the results varied a lot.

3.5. Previous Versions and Future Ideas

The project went through many iterations before arriving at the state described above. There were several ideas, which have not made it into the final version, for different reasons. These will be described in this section.

[10]Water texture taken from: `https://www.4dm-works.com/content/porsche_cayman_s_gt4.html` (accessed on: 10.09.2021)

[11]Battle Damaged Sci-fi Helmet - PBR by theblueturtle_, published under a Creative Commons Attribution-NonCommercial license: `https://sketchfab.com/models/b81008d513954189a063ff901f7abfe4` (accessed on: 10.09.2021)

3.5.1. Augmentations

The augmentations in the current version of the game are not considered augmented reality by the definition of Azuma [30] because the droplets are not registered in 3D and are only a video-overlay. However, it is definitely considered mixed reality by the definition of Milgram and Kishino [31]. The original plan was to include real augmented reality, but it was then later decided only to only this trimmed-down version. The reasons for this are explained below.

The first idea was to use BPS from BodyPix. The advantage of this is that not only the position of points are known, but the region that the specified part occupies, too. That means that water augmentations could be drawn over the whole arm and not just around a dot at the wrist. An example of this can be seen in Fig. 3.8a. Another feature that can be done with BPS is drawing a background image by coloring all pixels that do not belong to any body part. This can be seen in Fig. 3.8b.

These ideas worked quite well and could also lead to a more immersive experience because the augmentations match the pose of the player more closely. But this approach was ultimately discarded as it is quite computationally intensive and leads to very low FPS on slower systems. This would still not mean that registration happens in 3D, but at least depth would be considered.

Using *jeelizFaceFilter*[12], which is a JavaScript library for 3D face tracking, was also considered. It allows positioning of 3D objects relative to the head of the user as seen in Fig. 3.8c. It could have been used for more advanced augmentations in the face or other additional augmentations (e.g. game info). This was not done because there were already performance problems, which would have been made even worse.

3.5.2. Shaking

Implementing a more complex version of the shake detection, so that the participants need to put in more effort to get synchronized shakes, was also considered. The first idea was that the shakes needed to be in-phase with each other. But a concern was that the delay between the players was too large for it to work reliably. However, this would definitely be something that should be explored in a future version of the game, as it would probably benefit the synchrony between the players.

Another idea for an extension would be to provide different patterns that the players must follow while shaking (e.g. up/down, side-to-side), this would make the game more like a rhythm or dance game.

[12] `https://github.com/jeeliz/jeelizFaceFilter` (accessed on: 09.09.2021)

4. Method

In order to evaluate if the game actually leads to increased interpersonal synchrony between the players, a user study was set up. This chapter describes how it was performed.

4.1. Participants

Participants were sourced from the author's extended social network as well as from Technische Universität München. They were paired up with someone they did not know beforehand (some dyads had seen each other before, but no one considered the other as an acquaintance) and assigned to one of two conditions. These were synchronous mode first (3 dyads) and asynchronous mode first (4 dyads).

4.2. Procedure

The experiment took place remotely on the personal devices of the participants. Therefore, it was important that the participants had clear instructions on what their setup should look like, in order to get consistent results. Before the experiment began, participants got a document explaining them how their setup should look. This included their spatial setup, as in how much space they needed, how far away from the camera they should sit and how their lighting should be. Further, it also stated which browser they should use and it contained a URL where they could test the website including the augmentations to make sure their computer was powerful enough and to check if there were any other problems. The setup was then checked again by the experimenter, who told the participants if there were any problems (e.g. a light directly behind the head which impacts the shake detection).

The actual experiment took place using Discord. The participants were told that the reason for the study was to evaluate the game which had been developed. They were not told specifically that the synchrony between them was being investigated. An overview of the procedure can be seen in Fig. 4.1. First of all, the participants were introduced to each other and the experimenter via a video call. Then the participants filled out a consent form as well as demographic survey. After that, the main part began.

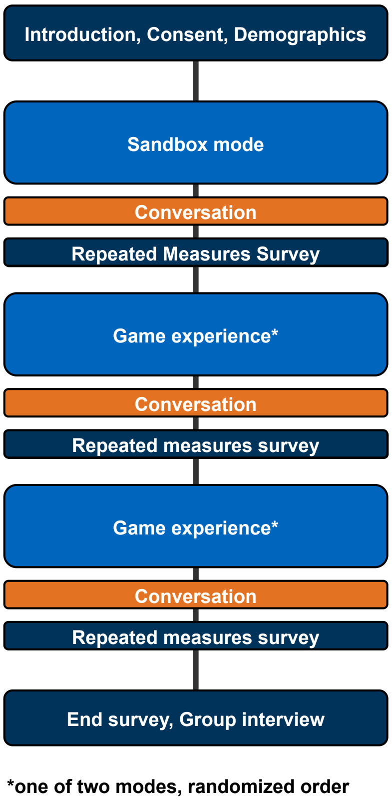


Figure 4.1.: Procedure of user study

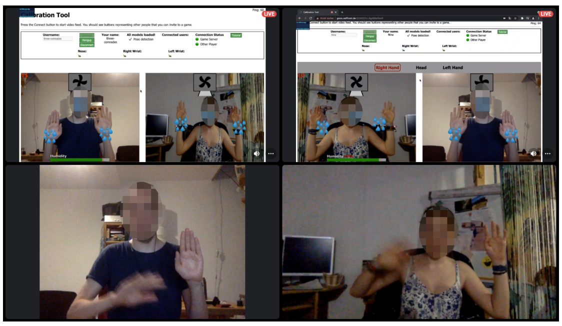


Figure 4.2.: Recording setup

For that the participants were instructed to open the website of the game using *Google Chrome*[1] (one participant used *Safari*[2] on macOS). Chrome was chosen here, because it yielded the best performance of all browsers that were tested. The main part consisted of a sandbox (SB) mode and then two game experiences. The participants shared their screens, so that the experimenter could see the gameplay from both perspectives as well as their cameras (see Fig. 4.2). However, for many participants using the camera in Discord and the game simultaneously did not work. Therefore, they only shared their screen. This was not a big problem as the camera was redundant anyway and all participants were clearly visible in the screen they shared. After each of the three parts, the participants were asked to talk about a neutral topic (e.g. job, vacation, hobbies) for three minutes. Then they needed to complete a repeated measures survey. At the very end, there was a closing survey as well as a semi-structured group interview.

In the SB the participants were introduced to the shake detection and were then given time to try it out and get used to it. The two game experiences consisted of the two participants playing the game in one mode each (see section 3.1.5). The order of the modes was counter-balanced to avoid learning of the game mechanics affecting the results. Every dyad played each condition for at least six minutes. If they finished a game and their playtime was below that, they were asked to play again.

4.3. Questionnaire

The questionnaire consisted of three parts. The first part contained questions about demographics as well as the experience that participants had had with AR/VR before. One question was also about how well the participants knew each other.

The second part was the repeated measures part. It contained questions from three categories which were asked in random order. The questions were answered on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The first category was a modified version of the NASA-TLX (NASA Task Load Index). The second was regarding the system. It contained questions like: “I felt that the computer or internet was too slow,” “This part of the experience was fun.” The last category was about the social interaction between participants. It included questions like: “I feel connected to my teammate,” “I would like to play with the person more.” The whole questionnaire can be found in the appendix (section A.3).

The last part was the closing survey. It was about the study as a whole and its purpose was to evaluate the general concept of the game as well as the technological implementation. It also contained a 7-point Likert scale (1 = strongly disagree, 7 =

[1] `https://www.google.com/intl/en_us/chrome/` (accessed on: 09.09.2021)
[2] `https://www.apple.com/safari/` (accessed on: 09.09.2021)

strongly agree) part. It asked the participants about the following aspects of the game: fun, fairness, game performance, augmentations and shake detection. There were also free-text questions where participants could leave feedback and suggestions and where they could also elaborate on problems they might have had.

5. Results

5.1. Demographics

A total of 14 participants ($m = 11$, $f = 3$) took part in the study. They were grouped in pairs, three pairs were mixed while 4 pairs consisted of two males each. Their average age was 24.07 years ($\sigma = 7.2$, $min = 14$, $max = 48$). Every participant stated that they currently lived in Germany. All of them had heard of AR/VR before but 5 had never used it. The rest had used it at least once and 2 participants stated they used it every week.

5.2. Repeated Measures

In this section, the results from the repeated measures survey of the questionnaire are presented. This is done in three parts. First the system section, next the social questions and last the NASA-TLX part.

In Table 5.2a an overview of the results from the system section can be seen. Notable results are highlighted and are now presented here. The users perceived internet or computer problems more strongly in the SM compared to the other modes. Moreover, the deviation for this question in the SM is the highest of all in the system section, which means these (perceived) problems varied a lot between participants. At the same time, the answers for lag or delay between the movements show very little difference between the modes.

The results from the social section, which are considered next, are presented in Table. 5.2b. First of all, the statement “I would like to continue talking with this person” got very high agreement with low standard deviation. The answers also differ very little between all three modes. So many participants would like to continue talking to each other and that did not change during the study. Similar results can be seen for the statement “I would like to play with the person more.” Another notable result is that there is no difference in connectedness between the SM and AM, but it is higher for both of these compared to the SB. Participants also stated that they were paying less attention to what the other player was doing in the AM compared to the other modes, although this statement also has a high deviation in all three modes. Lastly, the

Table 5.1.: Results of the repeated measures survey between the sandbox (SB), synchronous (SM) and asynchronous mode (AM).
Answers are on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree)

(a) Repeated measures system section

	I felt that the computer or internet was too slow		This part of the experience was fun		I felt that my partner and I had no lag or delay between our movements		I felt that my partner and I were mirrored	
	AVG	σ	AVG	σ	AVG	σ	AVG	σ
SB	1.57	0.73	5.86	0.83	3.64	1.95	4.07	1.71
SM	**4.14**	**2.10**	5.50	0.63	3.93	2.02	5.14	1.36
AM	2.36	1.44	5.93	1.03	4.21	1.61	4.64	1.63
Δ SM-AM	**1.79**	**0.66**	0.43	0.41	**0.29**	**0.41**	0.50	0.28
Δ SM-SB	**2.57**	**1.37**	0.36	0.21	**0.29**	**0.07**	1.07	0.35
Δ AM-SB	0.79	0.72	0.07	0.20	**0.57**	**0.34**	0.57	0.08

(b) Repeated measures social section

	I would like to continue talking with this person		I felt like I was working together with my teammate		I feel connected to my teammate		I was paying attention to what my teammate was doing		My teammate was efficient at their task		I would like to play with the person more	
	AVG	σ	AVG	σ	AVG	σ	AVG	σ	AVG	σ	AVG	σ
SB	**5.93**	**0.70**	5.14	1.51	4.79	0.86	5.57	1.35	5.57	1.12	5.57	0.90
SM	**6.07**	**0.80**	5.93	0.96	5.64	1.11	5.57	1.64	5.64	1.11	5.43	1.12
AM	**5.86**	**0.74**	6.29	0.80	5.64	0.89	4.93	1.49	6.00	0.65	5.71	0.70
Δ SM-AM	**0.21**	**0.06**	0.36	0.17	**0.00**	**0.21**	**0.64**	**0.15**	0.36	0.45	0.29	0.42
Δ SM-SB	**0.14**	**0.10**	0.79	0.54	**0.86**	**0.25**	**0.00**	**0.29**	**0.07**	**0.01**	0.14	0.21
Δ AM-SB	**0.07**	**0.04**	1.14	0.71	**0.86**	**0.03**	**0.64**	**0.14**	0.43	0.46	0.14	0.20

(c) Repeated measures NASA-TLX section

	The task was mentally demanding or complex		The task was physically demanding		I was hurried or rushed while completing my task		I was successful in accomplishing my overall task		I had to work hard to achieve my level of performance		I was irritated, stressed, and annoyed during the task	
	AVG	σ	AVG	σ	AVG	σ	AVG	σ	AVG	σ	AVG	σ
SB	2.07	1.10	1.57	1.05	2.64	1.59	5.79	1.01	2.64	1.63	1.50	0.63
SM	2.93	1.39	3.64	2.22	4.21	1.74	4.36	1.72	4.14	1.85	2.79	1.42
AM	3.21	1.66	3.29	1.79	3.93	1.75	5.50	1.05	4.29	1.75	2.43	1.29
Δ SM-AM	0.29	0.27	0.36	0.43	0.29	0.01	**1.14**	**0.66**	**0.14**	**0.10**	0.36	0.13
Δ SM-SB	0.86	0.29	2.07	1.17	1.57	0.15	**1.43**	**0.70**	**1.50**	**0.22**	1.29	0.80
Δ AM-SB	1.14	0.56	1.71	0.74	1.29	0.17	**0.29**	**0.04**	**1.64**	**0.12**	0.93	0.67

Table 5.3.: Total game rounds played and playtime for the synchronous (SM) and the asynchronous mode (AM)

SM		AM	
Rounds played	Total playtime [s]	Rounds played	Total playtime [s]
62	3406	95	3453

efficiency rating they gave each other are very similar between the SM and SB, but they are slightly higher in the AM compared with the other two modes.

Finally we take a look at the results from the NASA-TLX section which can bee seen in Table 5.2c. It is notable that the participants feel that they were overall less successful in accomplishing their task in the SM compared to the other modes, while the success in the other two modes is rated very similarly. However, very little difference in effort can be seen between the SM and AM, while the SB was perceived to be the mode which needed the least amount of effort.

5.3. Game and Player Performance

In addition to the players' subjective evaluations, objective information about the game has also been collected. This data can be used to evaluate the performance of the dyads in the SM and AM. Logs that have been created on the game server were analyzed to extract this information. The results of these are presented here.

Table 5.3 shows how many rounds were played during all studies in both conditions, and the total playtime for each condition is given as well. It can be seen that the playtime is roughly the same in both modes, but a third more rounds were played in the AM. All but one dyad completed more rounds in the AM compared to the SM, but the one dyad which completed more rounds in the SM had a bug in the AM, which prevented them from scoring higher. Thus, it is not possible to know how many rounds they could have completed in the AM.

In addition to the metrics from the game server, also the ping of the WebRTC connection between the participants was measured. This ping includes the time it takes for a message to be sent to the other player and back, as well as the time it takes for the other player's computer to process the message. This processing time was presumably the most significant contributor, especially on slower systems. Therefore, the ping is different for both players. This ping is a good estimate for the delay between the two

Table 5.4.: Rounds per minute, average ping and percentage of legal shakes (i.e. started less then one second apart) for each trial of the study

Trial	1	2	3	4	5	6	7
Average ping [ms]	65.8	91.9	91	297.6	39.2	115.2	27.5
Rounds / minute	1.11	1.64	1.44	0.85	0.93	0.44	2.80
Percentage of legal shakes [%]	50.00	53.85	79.31	34.78	36.36	7.14	87.25

video feeds and is therefore an important measure to consider. The values presented here are the averages from a few samples that were taken from both players over the whole duration of the respective studies. In Table 5.4 the average ping from each study is compared with the rounds the dyads managed to complete per minute. It can be seen that dyads with a lower ping generally scored more rounds per minute. For instance, study seven had the lowest ping and also scored the most rounds per minute, while study four with the highest average ping scored relatively few rounds per minute. The fact that the ping varied a lot between the participants should also be considered. For all but two dyads, one player had a ping of more than double that of the other player. Moreover, half of the participants had a ping below 50 ms and over 70% had a ping below 100 ms. However, due to the paring of the participants, only less than half the trials consisted of both players having pings less than 100 ms.

5.4. Synchronous Shaking

Another measure of player performance is how successful they were at shaking synchronously. As a reminder, a shake is legal if the starts of both shakes are within one second of each other. From the game log of the SM games, all shakes where both players started a shake within three seconds of each other were extracted. It was assumed that these were the shakes that were intentional while shakes with a larger difference were accidental. This is, of course, an arbitrary cut-off point, but it should still be possible to compare the results.

The average difference between synchronous shakes was 1114 ms ($\sigma = 1171$ ms), which is slightly higher than the one second that is needed to be accepted by the game. In Table 5.4 the percentage of shakes that were withing one second of each other can be seen for each study. Similar to the results for the rounds per minute, it can be seen

Table 5.5.: Results of the closing survey
Answers are on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree)

The experience as a whole was fun		The augmentations (water droplets) improved the experience		The game recognized my hand movements well		The game is fair		The game ran smoothly on my computer	
AVG	σ	AVG	σ	AVG	σ	AVG	σ	AVG	σ
6.29	**0.59**	6.07	0.70	5.43	**1.24**	6.07	**1.33**	6.43	0.90

Table 5.6.: Results of the observations of the conversations from the best and worst performing trials in the synchronous mode
SB: sandbox mode, SM: synchronous mode, AM: asynchronous mode

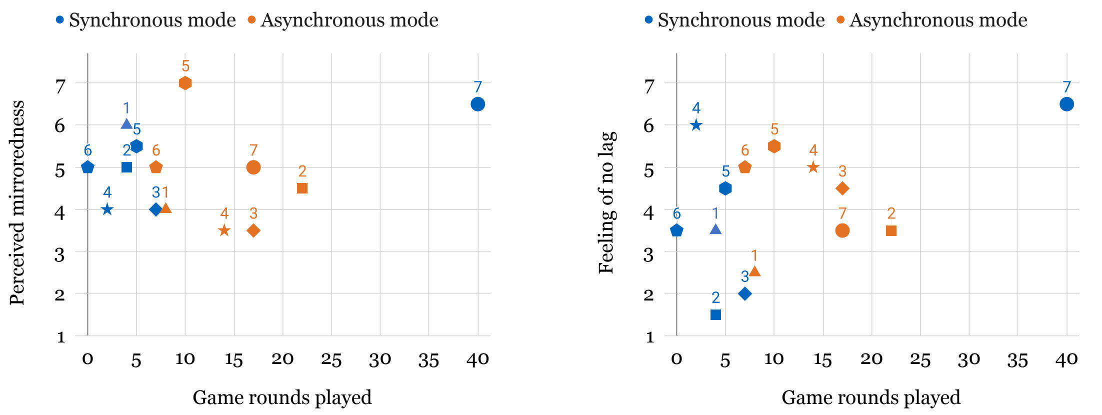
Worst trial		Best trial	
Order	# of sync movements	Order	# of sync movements
SB	1	SB	0
AM	0	AM	3
SM	1	SM	5

that a lower ping is associated with a higher percentage of legal shakes and therefore a better performance.

5.5. End Survey and Conversation Observations

Next we take a look at the results from the end survey as well as from observations that were made while the participants talked to each other after the game modes.

First, we want to take a look at the Likert section of the end survey of which the results can be seen in Table 5.5. The first thing to note is that all statements have a relatively high score. The two statements with the highest deviation are "The game is fair" and "The game recognized my hand movements well" which may be directly linked. Another thing to note is that the experience was rated with a high fun factor with low deviation overall.



(a) Perceived mirroredness

(b) Reported feeling of lack of lag between the movements of the participants

Figure 5.1.: Perceived mirroredness and feeling of lack of lag of the participants in relation to game rounds played in the synchronous and asynchronous modes

Second, the participants were observed while they were conversing after the modes. In particular the best and worst performing trials in the SM have been examined. They managed to get 40 and zero rounds respectively. Some kind of movements that were simultaneous were checked for, so not only interactional synchrony but interpersonal coordination in general was looked for. The results of this can be seen in Table 5.6. Examples of observations that have been made are participants touching their face or nodding their heads. These happened either simultaneously or shortly after one another. What can be seen in the results is that the trial that performed best had more synchronous movements than the worst trial. Further, it can be seen that there were more synchronous movements after the SM than the AM in both of them. Although the SM was also the last mode in both cases. Another observation that has been made is that the participants of the best trial generally moved around a lot more during the conversation, while the ones from the other trial were much more passive.

5.6. Perceived Lag and Mirroredness Between Participants

Another factor to consider while taking a closer look at these two trials is the perceived mirroredness ("I felt that my partner and I were mirrored") they felt as well as the amount of lag or delay they noticed between their movements ("I felt that my partner and I had no lag or delay between our movements"). As can been seen in Fig. 5.1a and

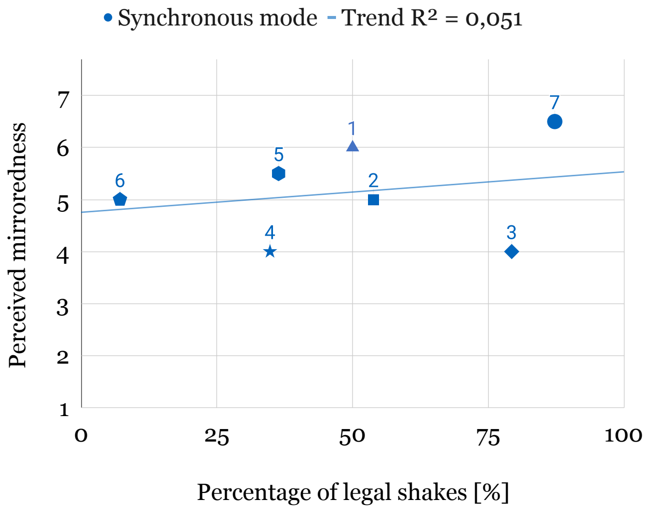


Figure 5.2.: Perceived mirroredness of the participants in relation to the percentage of legal shakes in the synchronous mode

Fig. 5.1b, trial seven, which was the best performing one, had a significantly higher perceived mirroredness and less lag in the SM compared to the AM. On the other hand, trial six, the worst performing one, felt no difference in mirroredness and more lag in the AM.

Fig. 5.1a shows that for most trials the perceived mirroredness is higher in the SM compared to the AM, although apart from trial one, that difference is less than one. In trial five the perceived mirroredness was significantly higher in the AM compared to the SM even though they were the third best performing trial in the SM.

When looking more closely at Fig. 5.1b, no clear pattern emerges. Some trials felt more lag in the SM while others felt more lag in the AM. However, it is to note that trial one felt less lag in the SM similarly to their increased feeling of mirroredness in this mode.

The last thing we want to do is compare the perceived mirroredness to the performance of the trials in the SM. To quantify the performance, the percentage of legal shakes was chosen as before. The results of this can be seen in Fig. 5.2. It shows that there is a small upward trend where better performance in the SM is associated with higher perceived mirroredness between the participants.

6. Discussion

In this chapter the results and their implications are discussed. First, we take a look at the focus of this thesis, which is interpersonal synchrony. Second, we consider comments from the participants about the game and how these could be incorporated. Third, we focus on multiplayer WebAR games in general and last, we discuss future work that needs to be done.

6.1. Success in Calibrating Synchrony

As the main objective of this thesis was to induce interpersonal synchrony between the participants, this section discusses what impact the game had on the participants in this regard.

6.1.1. Achieved Synchrony

To begin with, we want to explore whether the results indicate that synchrony was actually achieved.

Perceived Mirroredness Perceived mirroredness shows a difference between the modes, as for most trials it was higher in the SM compared to the AM. Additionally, a higher percentage of legal shakes in the SM, which indicates a larger amount of movement synchrony, was associated with higher perceived mirroredness. This could indicate that the way of assessing movement synchrony, that is the success of shaking synchronous in the game, is reliable to an extent as it matches up with the perception of the participants. This would mean that performing better in game leads to increased movement synchrony, which was part of our goal. However, it is difficult to deduce social implications from perceived mirroredness, as it is just the observation of movement synchrony from the participants and has no social aspect.

Connectedness There was little difference in the social section of the repeated measures study. For instance, feeling of connectedness, which could be a social effect of interpersonal synchrony, showed no significant difference between the SM and the AM.

However, comparing both of these modes to the SB shows that participants felt more connected to each other after playing them. This could indicate that playing the game, regardless of the mode, did actually increase their connectedness. Yet it also needs to be considered that by that point players have also talked to each other at least twice which, of course, can also impact their feeling of connectedness.

Generally, because the participants had to talk to each other before they filled out the survey, the conversation also affected the survey results. The experiment was designed this way, to maximize the impact of the game experiences on the conversations. However, it is now possible that the effect of the conversations on connectedness overshadow the difference between the two modes.

Performance and Effort Another aspect that was regarded is task performance. It is generally observed that people who experience movement synchrony also perform better, at least in simple tasks [4]. However, this was not the case in this experiment. In general, the participants scored less rounds in the SM than in the AM, while both modes were played for roughly the same amount of time. The same is confirmed by the perception of the participants. This could have multiple reasons, though first of all the two modes are probably not equally difficult. A point can be made that shaking at the same time is harder than shaking independently of each other, as more coordination is needed. It could also have technological reasons, which are discussed in the next section.

Nevertheless, there are some interesting results regarding perception of the participants. As stated above, they felt like they did not accomplish their task as well in the SM, but there was no difference in reported effort between the two modes. This is even though, some dyads really struggled with the synchronous shaking and needed to change their strategy a few times. Moreover, there was little difference in the efficiency ratings the players gave each other between the two modes. It is possible that because of the movement synchrony they experienced, the players perceived it to be less effort than it actually was. Yet it could also just be that because they had the same problems they had to solve together, it reduced their perceived workload.

Awareness of the Other Player Something else that was different between the modes was that participants payed less attention to their partners during the AM compared to the other modes. It is not clear why this was because it would seem that paying attention if the other person is executing your orders correctly, is important. It could be that because the other person is doing the same movements, it makes it more likely to look at them. Another reason could be that the participants did pay more attention to each other in order to check if they really started shaking at the same time. Nevertheless, it made the participants more aware of each other, which is a positive effect.

Observations Because many trials did not perform well in the SM, a closer look has been taken at the best performing trial in this mode. First of all, some synchronized movements could be observed during the conversations and even more could be observed after the SM in this trial. The same was also true for the worst performing trial, although the difference was smaller and there were also less synchronized movements in general in this trial. It could be that in the best trial, the participants were more naturally synchronized than in the worst trial and therefore exhibited more synced movements. This could also be a factor in their good performance. However, no synced movement could be observed in their first conversation after the SB, which would indicate that the game had an effect. But what is also important to remember here is that the SM came after the AM in this trial and therefore they had already been talking to each other for longer when the conversation after the SM happened. Moreover, they spent 53% more time playing the SM compared to the AM.

This trial also showed large differences in perceived mirroredness and feeling of lag between the SM and AM, which could be an indication that there was some synchrony between them. However, it could be that one of the players in this trial misread the statement "I felt that my partner and I had no lag or delay between our movements" in the AM and therefore thought the scale was the other way a round. This is thought because the answers from this participant vary a lot between the modes and are also not consistent with the ones from their partner. Lastly, it is worth considering that this is of course only one example that has been looked at and therefore can only provide indications and more rigorous testing is definitely needed.

Differing Amounts of Movement Synchrony The observations from the best trial lead to another question and that would be how the relationship between movement synchrony and real interpersonal synchrony looks like exactly. Most studies to date only compare conditions where either movement synchrony is present or not, but the question is how the behavior in between would look like. It might be the case that there is a linear relationship and slowly increasing movement synchrony also increases interpersonal synchrony. But on the other hand, something which also seems more likely, would be that it acts more like a switch. In this case a certain amount of movement synchrony would be required and only then interpersonal synchrony would be achieved. It is possible that most participants did not get over this threshold in our experiment. Maybe the best trial exceeded this amount which is the reason why some effects could be observed in it.

Overall Social Effects All that being said, the experience still had positive social effects on the participants overall. This is indicated because participants stated that they wanted to continue talking and playing with their partners. Participants also stated that they felt they were working together. But because the answers did not vary between the modes, it is not clear what the reason for this is and if synchrony plays a role in it. It could be just because they played the game together or because the participants talked to each other. Nevertheless, these are desirable social effects and they will be more thoroughly discussed in section 6.3.

Synchrony Unrelated to Movements As discussed, interpersonal synchrony can also be induced between people when they are just interacting naturally [5]. Therefore, it could be the case that playing the game increases the synchrony between the players regardless of the movement synchrony they exercise. It is speculated that this synchrony would be stronger if the players performed better in the game. If this was the case, it would explain why the best performing trial exhibited more synchronized movements during the conversation than the worst performing one. Additionally, this synchrony could also be part of the reason why the above-mentioned positive social effects were observed.

In conclusion, only limited positive effects could be found that could be directly linked to interpersonal gestural synchrony. Therefore, not much insight into whether a game like this could be used as a calibration tool for synchrony could be gathered. However, it is also not clear if the experiment successfully induced movement synchrony between the participants. This is also discussed in the next two sections.

6.1.2. Technological Challenges

As stated above, the overall performance in the SM was worse then in the AM. One big suspected reason for this are technological challenges. This is supported by the fact, that the performance of the participants was directly impacted by the ping. This is especially true for the SM (percentage of legal shakes). This section discusses in what way this impacted the experiment and how this can be improved.

Impact on Experiment A large ping can make it hard for the participants to shake simultaneously. This is because the game server just takes the absolute time difference between the two shakes starts into account. If there is a delay between the video feeds, a shake might look simultaneous for a participant while it actually is not and therefore it is not counted by the game. Furthermore, the ping is not the same for both players and thus it can happen that it looks simultaneous for one player but not for the other.

Some dyads really struggled to perform in the SM and the most probable reason for that is that they experienced large delay. This can also be seen in the results where participants with larger ping had lower percentage of legal shakes then those with smaller ping. One dyad did not manage to finish a single round even though it looked like they were shaking at the same time (at least on one of the participants' screen). This can really impact the experience of the players as they do not know why it does not work. This is also indicated by the results, because internet or computer problems were stated much more for the SM compared to the other modes. The fact that the fairness of the game had high deviation could also be an indicator for this. Players who did badly because of these problems probably considered the game as less fair. This was also brought up in the interview, for instance, one participant said: "I don't know if the video was really starting at the same time for both of us. Maybe that was a reason for not having success as well."

Therefore, it is evident that the experience in the SM was far from optimal for most dyads, but there still were no significant differences in perceived fun, effort or frustration. It could be that, had it not been for these technological problems, the SM would have done even better in these categories and that any possible positive social impact was diminished by these problems.

Measures to Reduce Problems There are a few things that can be done to reduce these problems. It would be possible to add a margin for the delay in the shake detection, but then also more not synchronous shakes would be detected. At the core it is most important that the shaking looks simultaneous for both players. This is only possible if both players have a low ping. Therefore, this must be assured, either by enforcing system requirements or by optimizing the game.

Another aspect to consider is how the simultaneous shake detection works. The fact that it only looks at the start of the shakes was a problem in some cases. Sometimes the participants would both continue shaking their body parts for a long time when they started too far apart, as they thought the shake was just not counted yet. Therefore, it might be a good idea to also consider different methods of detecting a simultaneous shake, such as looking at the phase or a defining a minimal amount of time for which both players must shake together. This could lead to a more consistent and intuitive game experience.

Research has been done to approximate the amount of delay from which on the experience of games is impacted. In [32] a study was done with a VR FPS in which negative effects (i.e. subjective and objective) appeared above 100 ms of delay. This can give an indication of how much delay is still acceptable, although it might be the case that movement synchrony requires less delay. An indication that this might be the case is that during music performances some negative effects already appear at 30 ms of delay [33].

In general, it can be said that the performance impact of slower systems on the game was underestimated, although it was considered during the development phase and was also the reason for the rather basic augmentations. However, our results in combination with previous research indicate that these problems can be reduced to an unnoticeable point. But to achieve that, both participants cannot exceed the delay limits stated above. This means that there needs to be a higher requirements for the performance of the computers of the participants, which, of course, limits the accessibility of the game.

6.1.3. Limitations

In addition to the technological challenges, there are also other methodological limitations to this experiment. First of all, the sample size of the experiment was too small to come to generalizing conclusion. However, the results can still point to interesting effects that can be studied in the future.

As stated earlier, the placement of the conversations before the questionnaire was filled out meant that no differentiation between the effects of the game and the effects of the conversation can be done, although the conversation was present in all modes and therefore, should impact all modes the same way. Nevertheless, it may be the case that the impact of the conversation was bigger than the game and therefore it is possible that differences in the modes are not measurable.

For interpersonal synchrony to occur, there are a few requirements to be met. First, the people must be able to anticipate the movement of the other person [12]. In order to do that, often visual contact is required [12]. Second, for an interaction to count as interactional synchrony it is often required for it to happen at the same frequency [12]. Furthermore, often only interactions are looked at which happen either in- or anti-phase, because these are the only phase shift conditions that can be maintained easily [34]. It has also been shown that positive effects do not appear if the movement does not happen at the same frequency, even if it is done at the same time [3, 4].

The SM of the game allowed people to anticipate the movement of the other player and also allowed for visual contact, but as discussed earlier there were no limitations on frequency or phase. Therefore, the type of movement induced by the game might not be considered interactional synchrony, although it is still interpersonal coordination in the broader sense [1], which is also associated with positive social effects.

The AM did not require players to shake simultaneously, although this does not mean players did not do that. Many dyads adapted a strategy where they would first both tell the other one what part they needed to shake and then start shaking. Additionally, the limited amount of time forced the players to started shaking as fast as possible. This could lead to both players shaking at the same, but of course different parts of the body. However, synchrony does not strictly require the movements to be the same [12, 23]. Nevertheless, the synchrony in movements in the AM was generally less than in the SM, at least from observations. Often players would already start moving while instructing the other player, or they would only tell the other player what they need to do after they finished shaking.

6.2. Comments from Participants

Many participants left comments about the game, either as they played or they wrote them in the survey. These comments are discussed in this section.

6.2.1. Shake Detection

Some participants stated that the shake detection sometimes did not work or that is was inconsistent. One participant wrote: "Detection varied a lot, so I wasn't sometimes sure if a bad detection was intended."

The reasons for shakes not being detected varied. Some participants had rather slow systems, which made the shake detection worse. But also sometimes participants did not shake the way they were instructed to. One thing that often happened was that participants shook their hand by turning it, instead of moving it from side to side. This cannot be detected by the current implementation as it only tracks the position of the wrist, which does not move if you shake this way. Often participants would shake the "correct" way at first, but then, as the game got hectic, changed to the "wrong" way. Another thing that sometimes happened was that the hands of the participants left the camera image while they were shaking, which also led to some shakes not being detected.

Hence, to make the game more fun and consistent for the players, the shake detection needs to be more robust and be able to detect different variations of shaking (e.g. turning hand, moving hand from side to side). Alternatively, it would also be possible to encourage the "correct" way of shaking through game design (e.g. moving boxes on the screen that need to be followed).

6.2.2. Head Shaking

The most common comment was that the head shaking was uncomfortable and made them dizzy. One participant said: "One thing was really annoying, the head movement [...] it makes you dizzy." This was also something participants said to each other during the game. Another participant stated that while moving the head you cannot read the what is on the screen. This makes it hard to know when the shake is detected and also to follow the game in general.

That the head shaking can be uncomfortable was already noticed during the development. Therefore, the speed at which the head has to move was lowered to make it more pleasant. But as the participants did not know how fast they had to move, most of them moved their head quite fast to be sure it was detected. Nevertheless, even when moving slower it still becomes uncomfortable after playing 40 rounds (which was the high score during the user study) and not being able to see what is on screen is not optimal either.

This means that either another body part should be chosen or the required movement needs to be made even smaller (e.g. just looking left and then right, small nod).

6.2.3. Game Design

Lastly, there were some comments about various aspects of the game design that are discussed shortly.

Game Events One common comment was that the players wanted more feedback from the game. For instance, it was suggested that there should be sound cues when a shake is detected or when the vent activates. This would makes it easier to recognize when the player can stop shaking. One player also suggested that the timer should become larger when it is close to expiring to make sure the player notices it.

In general, it can be said, that game events were communicated on a very basic level via text messages on the screen. Because of that, they were often ignored by the players. Hence, the game experience could be improved by developing more natural ways of communicating these events. These could include sound cues for completed shakes and finished rounds. But also augmented animations when a shake is detected (e.g. a ring that grows around the detected part).

Tasks Players also commented about the tasks they had to do. It was stated that the task became boring after a time and that more additional complex tasks would be nice. One participant suggested that there could be tools (e.g. clock, metronome) which can be used by the players if they liked. But in order to use these tools, they would first need to be calibrated. This would leave the players the choice if it is worth taking the time to synchronize these tools.

The tasks in the game are very basic, which is for two reasons. First, it makes it easier to study the effects of synchrony while not including other effects like shared spaces.

Second, is the limited development time of the game, though it is definitely possible to make more complex tasks for the game, like the ones discussed in section 3.5. Also elements alongside the tasks, like the tools suggested above, could be a good addition.

Besides shaking there are also a lot of other movement tasks that could be implemented while still requiring the players to be synchronized. For instance, there could be circles that appear on screen and the players need to move their hands on top of them. Only if players have their hands placed correctly, they would advance. Different colored circles could be used for different body parts.

Setting The last notable comment that was made was to add a background to the video feed: "Maybe a 'Spaceshuttle' background for the video." This idea could make the game more immersive and also remind the players, that they are in a zero gravity environment.

This was actually considered during the design phase and an example of this can be seen in Fig. 3.8b. The body part segmentation from BodyPix would allow for this to be done. As it was then decided against using BodyPix, for performance reasons, the idea was discarded. Although a similar effect could be achieved simply by adding a spaceship themed frame around the image at considerable lower computational cost.

In conclusion, it can be said that there are some things in the game's design that should be improved in a future version of the game that can help to improve immersion and the general gameplay experience.

6.3. Viability of Multiplayer WebAR Game in General

This section discusses if multiplayer WebAR games are viable in general, unrelated to the topic of synchrony. As these kinds of games have been rare until now, this game can give a first insight into this topic.

WebAR is not yet used widely, which has multiple reasons. First, AR in general is not yet adopted in everyday use, although, as discussed earlier, there are some applications for it (e.g. room planning). As often, there are also games which have been one of the first ones to adopt the technology. But secondly, most of these applications, game or not, use a native AR library like *ARKit*[1], rather than WebAR. Because WebAR is confined to the web browser, not all of the performance or sensors (e.g. LiDAR) of a device can be used. This is probably the main reason why native implementations are used more often.

[1] `https://developer.apple.com/documentation/arkit` (accessed on: 09.09.2021)

Nevertheless, we are now at a point where this restricted power is still powerful enough to use the technology effectively. Despite this, the shortcomings WebAR can still be seen in our experiments, as discussed above. This means, to really be able to enjoy the game, a fairly powerful setup is needed, even though the augmentations in this game are rather basic.

Therefore, to achieve a good experience for as many people as possible, different performance modes could be introduced. In the highest mode, all augmentations would be active. And in the lower modes some of these augmentations would be exchanged for more basic UI elements. Additionally, for each mode, a system specification should be provided, so that everyone can choose a mode appropriate to their setup.

From a social perspective, this experiment shows that such games can work. First, this can be seen because most participants agreed that the experience was fun overall. Second, the participants felt more connected to each other after playing the game compared to after they only did the SB. But as discussed, the conversation could have also had an effect on this. Moreover, the results also suggest that the augmentations of the water droplets improved the gameplay. These findings suggest that such a multiplayer WebAR game does work in general and also that it can have positive effects.

Therefore, it is possible that such a game could be used in order to improve the perception of video conferences. So, for example, such a game could be played before a video conference starts in order to lift the mood similarly to a real world group game as theorized earlier. Such a game could also be imagined for groups. An example of this would be a game where all cameras are arranged in a grid. Then a virtual ball floats across these camera feeds and the people can interact with it. For such a setting, the above-mentioned performance modes would be very important, as everybody should be able to participate.

In conclusion, this experiment shows that such multiplayer WebAR games are viable, but care must be taken as to how computational demanding the application is and how powerful the devices of the users will be. It also seems like such games can have positive social effects, but more research is needed to confirm this and also to investigate the difference between these and classic computer games.

6.4. Future Work

As discussed above, many questions remain unanswered and this section discusses what future work could help answer them.

As it is not clear if interactional synchrony was actually achieved in this experiment, further experiments should be done where this aspect is controlled more closely. This could, for example, include requiring in-phase shaking. A small test with an algorithm which checks for the concurrency of direction changes as an indication of in-phase shaking was done. The test indicates that such detections can work, but it was also done in optimal conditions with powerful computers. What was also noticed during this test was that trying to keep in-phase requires looking at the other player at all times.

In any case, a future experiment would need to happen in an environment where the performance and internet infrastructure is more closely controlled to ensure that there are only small delays between the participants. What is also still unclear is what influence the game had in comparison to the conversation, so this is something that needs to looked into individually.

In lab environments, more complex methods to measure interpersonal synchrony in this setting could also be used. Examples for this would be hyperscanning similar to the study in [6] or measuring head movements and eye blinks like in [5]. Such experiments could also give more insights into where a game like this could be used as a calibration tool for synchrony as theorized in the beginning.

Another question that was brought up by our research was the impact which differing amounts of movement synchrony has. This is something which can be investigated by future work to figure out if the relationship between movement synchrony and interpersonal synchrony behaves linearly or more like a switch.

As the game had positive effects as an icebreaker on the participants, this could be an interesting topic of further research. Future work could consider whether playing such a game with someone before getting to know them increases the social relationship. In addition, the concept of video conference games could also be further investigated and also expanded to group settings as discussed above.

7. Conclusion

In summary, it can be said that this game shows that multiplayer WebAR games are viable in general. Computational performance is still an important aspect to consider, but that will become less important in the future. The game had positive effects on the participants generally, which leaves the topic of video conferencing games as an interesting research topic. This could facilitate video conferences being seen as a more enjoyable experience overall.

Even though no direct effects of interpersonal synchrony could be observed in this experiment, some interesting results were still present. Future work could continue to investigate this in a more controlled setting with fast computers and an established internet infrastructure to ensure only small delays and also extend it to a group setting. This thesis can provide helpful insights to future experiments on the topic of synchrony in video conferencing environments and whether it can be used as a form of calibrating synchrony between new acquaintances who are meeting online.

A. Appendix

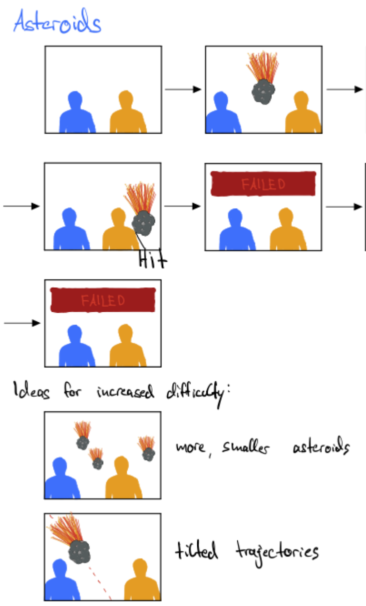
A.1. Evaluation of Ideas

Table A.1.: Evaluation of initial ideas

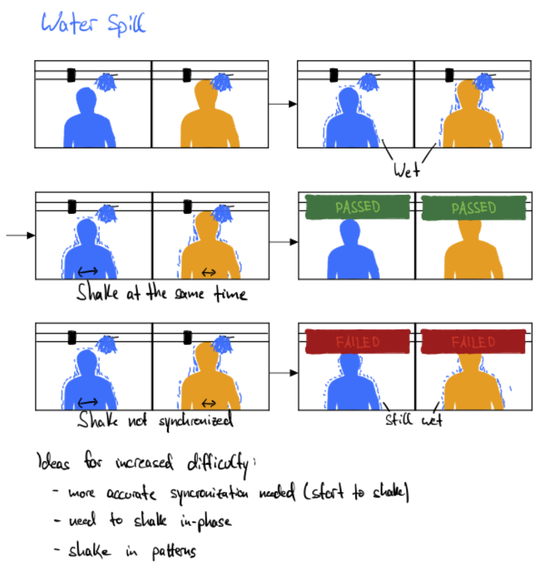
Idea	Technic. feasible	Synchrony invoking	Fun	Duration	Easy to visualize	Good Story / Motivation	Technologies needed	Type of motion
Asteroids	5	3	4	Continuous	4	4	BPS	S
Wave	3	4	4	Continuous	3	3	BPS/PE/FT	T
Punch Aliens	3	1	5	Continuous	2	5	BPS, PE	T
Lab Facial Gestures	2	3	3	Continuous	5	4	FE	S
Throw & Catch	4	4	4	One-off	4	3	HT	T
Difference	2	5	5	Continuous	2	3	PE, FE	T
Water Spill	4	4	4	One-off	2	5	BPS/PE, (HT, FT)	T
Hand Clapping	3	5	5	Continuous	2	?	HT	T
Pose Mimic	4	3	3	Continuous	5	?	PE	S
Falling Words	3	1	1	Continuous	2	3	PE, HT	T
Rhythmic Task	3	5	4	Continuous	4	?	PE, (HT)	?
Face Swap	4	2	2	Continuous	4	?	FT	?
Protect the Baby	3	1	2	Continuous	3	4	FT, (PE, HT)	?
Dance Commu-nication	?	4	3	Continuous	?	?	BPS, PE	?

'/': both technologies would work
',': both technologies are needed
'?': not filled in because it was not known yet
BPS: Body part segmentation
PE: Pose estimation
FT: Face tracking
HT: Hand tracking
S: Static motion (i.e. they need to do something static like a pose)
T: Temporal motion (i.e. they need to do some movement)

A.2. Storyboards



(a) *Asteroids*



(b) *Water Spill*

Figure A.1.: Storyboards that have been created to develop the ideas

A.3. Questionnaire

L = 7-point Likert scale (1 = strongly disagree, 7 strongly agree)
F = free-text answer

Demographics

1. How old are you?
2. What is your gender identity?
3. Which country you are currently living in?
4. Where did you grow up?

5. What is your occupation?
6. What is the highest level of education you have completed?
7. Do you have experience with Augmented or Virtual Reality?
8. Please tell us about your experience with Augmented and Virtual Reality (F)
9. How familiar is the other player to you?

Repeated Measures

System

10. I felt that the computer or internet was too slow (L)
11. This part of the experience was fun (L)
12. I felt that my partner and I had no lag or delay between our movements (L)
13. I felt that my partner and I were mirrored (L)

Social

14. I would like to continue talking with this person (L)
15. I felt like I was working together with my teammate (L)
16. I feel connected to my teammate (L)
17. I was paying attention to what my teammate was doing (L)
18. My teammate was efficient at their task (L)
19. I would like to play with the person more (L)

NASA-TLX

20. The task was mentally demanding or complex (L)
21. The task was physically demanding (L)
22. I was hurried or rushed while completing my task (L)
23. I was successful in accomplishing my overall task (L)
24. I had to work hard to achieve my level of performance (L)
25. I was irritated, stressed, and annoyed during the task (L)

Closing

26. The experience as a whole was fun (L)
27. The augmentations (water droplets) improved the experience (L)
28. The game recognized my hand movements well (L)
29. The game is fair (L)
30. The game ran smoothly on my computer (L)
31. If the game didn't run smooth please describe what the problem was (F)
32. Do you have suggestions that would improve the game? (F)
33. Do you have anything else you want to tell us? (F)

List of Figures

List of Tables

Bibliography

[1] F. J. Bernieri and R. Rosenthal, "Interpersonal coordination: Behavior matching and interactional synchrony," 1991.

[2] S. S. Wiltermuth and C. Heath, "Synchrony and cooperation," *Psychological science*, vol. 20, no. 1, pp. 1–5, 2009. DOI: 10.1111/j.1467-9280.2008.02253.x.

[3] M. J. Hove and J. L. Risen, "It's all in the timing: Interpersonal synchrony increases affiliation," *Social Cognition*, vol. 27, no. 6, pp. 949–960, 2009. DOI: 10.1521/soco.2009.27.6.949.

[4] P. Valdesolo, J. Ouyang, and D. DeSteno, "The rhythm of joint action: Synchrony promotes cooperative ability," *Journal of Experimental Social Psychology*, vol. 46, no. 4, pp. 693–695, 2010. DOI: 10.1016/j.jesp.2010.03.004.

[5] A. Gupta, F. L. Strivens, B. Tag, K. Kunze, and J. A. Ward, "Blink as you sync," in *Proceedings of the 23rd International Symposium on Wearable Computers*, K. Farrahi, R. Harle, and N. Lane, Eds., New York, NY, USA: ACM, 2019, pp. 66–71. DOI: 10.1145/3341163.3347736.

[6] I. Gumilar, E. Sareen, R. Bell, A. Stone, A. Hayati, J. Mao, A. Barde, A. Gupta, A. Dey, G. Lee, and M. Billinghurst, "A comparative study on inter-brain synchrony in real and virtual environments using hyperscanning," *Computers & Graphics*, vol. 94, pp. 62–75, 2021. DOI: 10.1016/j.cag.2020.10.003.

[7] Y. Sun, O. Shaikh, and A. S. Won, "Nonverbal synchrony in virtual reality," *PLOS ONE*, vol. 14, no. 9, e0221803, 2019. DOI: 10.1371/journal.pone.0221803.

[8] L. K. Cirelli, S. J. Wan, C. Spinelli, and L. J. Trainor, "Effects of interpersonal movement synchrony on infant helping behaviors: Is music necessary?" *Music Perception*, vol. 34, no. 3, pp. 319–326, 2017. DOI: 10.1525/mp.2017.34.3.319.

[9] Forrester Research, Inc., *Streamlined videoconferencing strategy drives business results*, 2018.

[10] D. Serhan, "Transitioning from face-to-face to remote learning: Students' attitudes and perceptions of using zoom during COVID-19 pandemic," *International Journal of Technology in Education and Science*, vol. 4, no. 4, pp. 335–342, 2020. DOI: 10.46328/ijtes.v4i4.148.

[11] B. K. Wiederhold, "Connecting through technology during the coronavirus disease 2019 pandemic: Avoiding 'zoom fatigue'," *Cyberpsychology, behavior and social networking*, vol. 23, no. 7, pp. 437–438, 2020. DOI: `10.1089/cyber.2020.29188.bkw`.

[12] J. L. Lakin, "18 behavioral mimicry and interpersonal synchrony," in *Nonverbal Communication*, J. A. Hall and M. L. Knapp, Eds. De Gruyter Mouton, 2013, pp. 539–576. DOI: `doi:10.1515/9783110238150.539`.

[13] M. J. Richardson, K. L. Marsh, R. W. Isenhower, J. R. L. Goodman, and R. C. Schmidt, "Rocking together: Dynamics of intentional and unintentional interpersonal coordination," *Human Movement Science*, vol. 26, no. 6, pp. 867–891, 2007. DOI: `10.1016/j.humov.2007.07.002`.

[14] A. Z. Zivotofsky and J. M. Hausdorff, "The sensory feedback mechanisms enabling couples to walk synchronously: An initial investigation," *Journal of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, p. 28, 2007. DOI: `10.1186/1743-0003-4-28`.

[15] P. Mitkidis, J. J. McGraw, A. Roepstorff, and S. Wallot, "Building trust: Heart rate synchrony and arousal during joint action increased by public goods game," *Physiology & Behavior*, vol. 149, pp. 101–106, 2015. DOI: `10.1016/j.physbeh.2015.05.033`.

[16] J. Lumsden, L. K. Miles, and C. N. Macrae, "Sync or sink? Interpersonal synchrony impacts self-esteem," *Frontiers in Psychology*, vol. 5, p. 1064, 2014. DOI: `10.3389/fpsyg.2014.01064`.

[17] S. Wallot, P. Mitkidis, J. J. McGraw, and A. Roepstorff, "Beyond synchrony: Joint action in a complex production task reveals beneficial effects of decreased interpersonal synchrony," *PLOS ONE*, vol. 11, no. 12, e0168306, 2016. DOI: `10.1371/journal.pone.0168306`.

[18] C. Fullwood, "The effect of mediation on impression formation: A comparison of face-to-face and video-mediated conditions," *Applied Ergonomics*, vol. 38, no. 3, pp. 267–273, 2007. DOI: `10.1016/j.apergo.2006.06.002`.

[19] J. Desnoyers-Stewart, E. R. Stepanova, P. Pasquier, and B. E. Riecke, "JeL: Connecting through breath in virtual reality," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, S. Brewster, G. Fitzpatrick, A. Cox, and V. Kostakos, Eds., New York, NY, USA: ACM, 2019, pp. 1–6. DOI: `10.1145/3290607.3312845`.

[20] E. R. Stepanova, J. Desnoyers-Stewart, P. Pasquier, and B. E. Riecke, "JeL: Breathing together to connect with others and nature," in *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 641–654. DOI: 10.1145/3357236.3395532.

[21] B. Tarr, M. Slater, and E. Cohen, "Synchrony and social connection in immersive virtual reality," *Scientific Reports*, vol. 8, no. 1, p. 3693, 2018. DOI: 10.1038/s41598-018-21765-4.

[22] V. Wikström, M. Falcon, S. Martikainen, and K. Saarikivi, "SynchroMouse: A game of improvised joint action," in *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, E. Gilbert and K. Karahalios, Eds., New York, NY, USA: ACM, 2019, pp. 418–422. DOI: 10.1145/3311957.3359482.

[23] T. Park, U. Lee, B. Lee, H. Lee, S. Son, S. Song, and J. Song, "ExerSync: Facilitating interpersonal synchrony in social exergames," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW '13, New York, NY, USA: Association for Computing Machinery, 2013, pp. 409–422. DOI: 10.1145/2441776.2441823.

[24] Inter IKEA Systems B.V., *Ikea App Page - IKEA*, Accessed on: 10.09.2021. [Online]. Available: https://www.ikea.com/au/en/customer-service/mobile-apps/say-hej-to-ikea-place-pub1f8af050.

[25] BeLight Software Ltd., *Live Home 3D — Home Design App for Windows, iOS, iPadOS and macOS*, Accessed on: 10.09.2021. [Online]. Available: https://www.livehome3d.com/.

[26] L. Lehtonen, M. D. Kaos, R. Kajastila, L. Holsti, J. Karsisto, S. Pekkola, J. Vähämäki, L. Vapaakallio, and P. Hämäläinen, "Movement empowerment in a multiplayer mixed-reality trampoline game," in *CHI PLAY '19*, J. Arnedo and L. E. Nacke, Eds., ser. ACM Digital Library, New York, NY, USA: Association for Computing Machinery, 2019, pp. 19–29. DOI: 10.1145/3311350.3347181.

[27] K. Inokuchi, M. Tsukada, and H. Esaki, "Co-sound: An interactive medium with WebAR and spatial synchronization," in *Entertainment Computing - ICEC 2020*, N. Jardim Nunes, L. Ma, M. Wang, N. Correia, and Z. Pan, Eds., ser. Information Systems and Applications, incl. Internet/Web, and HCI, Springer International Publishing, 2020, pp. 255–263. DOI: 10.1007/978-3-030-65736-9_22.

[28] I. d. S. Almeida, M. A. Oikawa, J. P. Carres, J. Miyazaki, H. Kato, and M. Billinghurst, "AR-based video-mediated communication: A social presence enhancing experience," in *2012 14th Symposium on Virtual and Augmented Reality (SVR 2012)*, IEEE, 2012, pp. 125–130. DOI: 10.1109/SVR.2012.4.

[29] B. Feher, L. Sidi, A. Shabtai, R. Puzis, and L. Marozas, "WebRTC security measures and weaknesses," *International Journal of Internet Technology and Secured Transactions*, vol. 8, no. 1, pp. 78–102, 2018. DOI: 10.1504/IJITST.2018.092138.

[30] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997. DOI: 10.1162/pres.1997.6.4.355.

[31] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. E77-D, no. 12, pp. 1321–1329, 1994.

[32] Vlahovic Sara, Suznjevic Mirko, and Skorin-Kapov Lea, "Challenges in assessing network latency impact on QoE and in-game performance in VR first person shooter games," in *2019 15th International Conference on Telecommunications (ConTEL)*, 2019, pp. 1–8. DOI: 10.1109/ConTEL.2019.8848531.

[33] P. F. Driessen, T. E. Darcie, and B. Pillay, "The effects of network delay on tempo in musical performance," *Computer Music Journal*, vol. 35, no. 1, pp. 76–89, 2011. DOI: 10.1162/comj_a_00041.

[34] R. C. Schmidt, C. Carello, and M. T. Turvey, "Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 16, no. 2, pp. 227–247, 1990. DOI: 10.1037/0096-1523.16.2.227.