University of Stuttgart
Germany

Institute for Parallel and Distributed Systems
- Analytic Computing Department -

Universitätsstraße 32
70569 Stuttgart

Bachelor Thesis

Gaze and voice driven hands free gaming

Mike Lauer

Study program:	B.Sc. Informatik
Examiner:	Prof. Dr. Steffen Staab
Advisor:	Ramin Hedeshy, M.Sc.
Start date:	15.03.2021
End date:	15.09.2021

Acknowledgment

I would like to thank Prof. Dr. Steffen Staab for his feedback during the initial phase of the thesis. I am very grateful to Ramin Hedeshy and Dr. Chandan Kumar for their continuous advice and guidance throughout this project. I would like to thank my family for their support and for testing the game. I must also thank all the study participants for their time and participation, as well as the institutions that were very interested and shared the information with their members. These include in particular the "Körperbehinderten-Verein Stuttgart e.V.", the "Margarete-Steiff-Schule Stuttgart" and the "Kinder- und Jugendheim St. Kilian Walldürn".

Zusammenfassung

Das Spielen von Computerspielen ist eine der häufigsten und unterhaltsamsten Aktivitäten im modernen Lebensstil. Menschen mit motorischen Beeinträchtigungen, die Maus und Tastatur nicht benutzen können, sind jedoch nicht in der Lage, mit Computergeräten so zu interagieren, wie es das Spieldesign erfordert. In dieser Hinsicht können neuartige Interaktionstechniken eine freihändige Steuerung mit Blick und Stimme als Eingabemodalitäten ermöglichen, um Menschen mit motorischen und sprachlichen Beeinträchtigungen zu unterstützen. Dies ist die erste Evaluierung einer Videospiel-Steuerungsmethode, die aus einer Kombination von Eye-Tracking und nonverbalen Sprachbefehlen (z.B. Summen) besteht und in einer 2D-Jump-and-Run-Spielumgebung angewendet wird, die wesentliche räumlich-zeitliche Interaktionen beinhaltet. Um die Machbarkeit der Interaktion zu beurteilen, bestand die Evaluationsstudie sowohl aus qualitativen als auch aus quantitativen Maßnahmen. Zusätzlich wurde die Machbarkeit mit einer Nutzerzielgruppe von Menschen mit motorischen Beeinträchtigungen validiert. Insgesamt deuten die Ergebnisse auf eine geringere, aber konkurrenzfähige Leistung bei gleichzeitiger Erhöhung des Spaßfaktors hin.

Abstract

Playing computer games is one of the most common and enjoyable activities in the modern lifestyle. However, people with motor impairments who cannot use mouse and keyboard are not able to interact with computing devices as required by the game design. In this regard, novel interaction techniques can enable hands-free control using gaze and voice as input modalities to assist people with motor and speech impairments. This is the first evaluation of a video game control method consisting of a combination of eye tracking and non-verbal voice commands (e.g. humming) applied in a 2D jump-and-run game environment involving essential spatio-temporal interactions. To assess interaction feasibility, the evaluation study consisted of both qualitative and quantitative measures. In addition, the feasibility was validated with a target user group of people with motor impairments. Overall, the results indicate a lower but competitive performance while increasing the fun factor.

Contents

1. Introduction

Restrictions in motor movement have a severe impact on everyday life. Depending on the effect, simple household activities cannot be carried out at all or only with the help of others. This does not only apply to those who have such a disability from birth; it can affect anyone. For example, strokes or accidents with spinal cord injury can quickly lead to paralysis. While physical activities such as cycling or swimming are often not readily available, digital media at first glance offer great potential for entertainment. The computer, in any form, has become an indispensable part of modern life. It is not for nothing that video games are an established means of entertainment for many people [6]. In the pandemic year 2020, there were about 2.7 billion gamers worldwide, representing about a third of the world's population [31]. Video games can be found on every platform: The classic desktop PC, the game console, or simply on the smartphone.

However, if we look at how we interact with these devices, we quickly realise that this promising entertainment value is not accessible to people with motor impairments. The mouse for guiding the pointer requires precise movements of the hand and arm. The keyboard seems more straightforward at first, but it also requires certain dexterity to hit the right keys. The controllers of game consoles require a lot of sensitivity and flexibility of the fingers as well. In summary, computer interaction in general, but especially regarding video games, is not an easy obstacle to overcome if no precise motor movement is possible.

Novel computer interaction technologies can solve this problem. Eye tracking allows the user's point of view to be captured, thus replacing the physical mouse. It requires no motor movement of the extremities but offers comparable accuracy. Voice control is often used as a secondary modality. With modern algorithms, words or sentences can precisely execute all kinds of commands. Phrases like "Turn on the coffee machine" or "Play my music" are nothing new these days. For smart assistants, this offers a good range of functions, but new problems arise. Users who are not able to speak clearly will be excluded again. Furthermore, verbal commands need a relatively long time to be spoken aloud and can only be processed afterwards. But there are many more ways to process voice as input. Instead of pronouncing complex words, non-verbal noises are much more accessible. For example, humming can be detected and processed by the computer in real-time, while most people with disabilities can make sounds.

1.1. Aims and Objectives

This thesis investigates the feasibility of hands-free multi-modal control in computer games. A game specially developed for this purpose is the core of a user study to record playability and persuasiveness during play. The aim is to determine whether this can be a possible future of entertainment for people with motor impairments or whether the gaming experience is still limited due to the input modalities and the current state of technology.

Objectives

1. Develop a game in a suitable and representative format that makes use of spatio-temporal qualities.
2. Design the game so that hands-free modalities are applicable and efficient.
3. Evaluate the reliability and performance of novel interaction techniques in video games, both individually, i.e. gaze for spatial orientation and humming for temporal bracketing, and in combination as a complete interaction system.
4. Evaluate the different perceptions of the game through the interaction techniques.
5. Evaluate the feasibility of novel interaction modalities for people with motor impairments.

1.2. Hypotheses

The research study focused on the reliability, engagement, and feasibility of gaze and non-verbal input as a spatio-temporal channel in human–avatar interaction. Specific hypotheses tested are:

1. Game Control
    a) Gaze is a reliable input (as accurate as mouse movement) for **spatial orientation** of an avatar in gameplay.
    b) Humming is a reliable input (as accurate as key press-release) for **temporal bracketing** of the actions of an avatar in game play.
    c) Gaze and Hum is a reliable multi-modal input for **spatio-temporal control** in human-avatar interaction.

To test the reliability, the focus was mainly on quantitative measures like time, distance travelled, etc. to quantify the control for gaze, hum, mouse and keyboard. The perceived control from subjective assessment strengthened the hypothesis test.

2. Game Engagement
    a) Gaze and Hum as natural input provides a more **immersive** (also referred to as perceived realism, sensory and imaginative immersion) experience compared to a physical device control intervening between human and avatar.
    b) Gaze and Hum as natural input provides a better **flow and persuasiveness** (e.g., time went by quickly while playing this game) compared to physical device control.
    c) Gaze and Hum as a novel input provides a more **challenging** experience (contributing to joy) compared to physical device control.

To test the engagement, the focus was mainly on qualitative measurements. The correlation from quantitative measurement strengthened the hypothesis test.

3. Feasibility
   a) Gaze and Hum as a hands-free interaction medium provides a **reliable** input for human-avatar interaction for people with motor impairment.
   b) Gaze and Hum as a hands-free interaction medium provides an **engaging** game play experience for people with motor impairment.
   c) Gaze and Hum as a hands-free interaction medium **enables** people with motor impairment to perform as good as other individuals using gaze and hum for game play.

To test the feasibility, the focus was mainly on the subjective feedback of perceived control and engagement from the participants, who could represent the target group where hands-free control for gaming could be useful in future.

1.3. Approach

For a systematic analysis of the characteristics of the new modalities, two user studies were conducted. The purpose of the first study was to find out how the new hands-free control performs compared to the conventional combination of mouse and keyboard. In the pilot study, which served to test the structure and procedure of the study, but also to analyse the difficulty of the game, 13 young people took part. In the main study there were a total of 15 participants. The second study explicitly focused on usability for people with motor impairments. They represent the target group of this research and are thus an essential part of the feasibility analysis. Ten people with motor and partly also speech impairments participated. Although they all enjoyed the game and received a good impression, due to various circumstances, which will be explained in more detail later, only the quantitative data of five participants could ultimately be used for a more detailed data analysis.

In order to investigate the subjective perception of playing as well as the performance in detail, both studies relied on qualitative and quantitative measurements. Previous work has looked chiefly at either user experience or performance compared to standard interaction methods. In this work, both were investigated to get a concrete impression of the developed control system.

The quantitative data was automatically collected by the game, which ran permanent measurements in the background. A wide variety of parameters were measured, which provide conclusions about each movement and the general movement behaviour. A detailed list follows in the evaluation chapter. Two examples are each movement's start and stop coordinates and the duration for which a button was pressed. Measurements were taken in milliseconds, as temporal actions can be very fast in video games.

The qualitative evaluation consisted of a questionnaire that was specially compiled based on recommended questionnaires and previous work. Besides the participant information, it consisted of twelve questions about experience and control in form of a 5-point Likert scale.

The two-tailed t-test was used to make statements about the significance of differences in the data. Detailed values can be found in the appendix, whereas simplified representations are used in the following.

1.4. Contribution

This paper shows how computer games with hands-free interaction can be made accessible to people with motor disabilities using affordable means and without complicated customisation. Based on previous research, eye tracking is now combined with continuous non-verbal voice commands and presented in a specially developed game. Two studies, which examine both the comparison to conventional modalities such as mouse and keyboard, but also the feasibility for people with disabilities, show great potential for the topic of accessible gaming.

The next chapter gives an overview of computer accessibility and explains previous work in the field of hands-free gaming to illustrate the current state of affairs. Chapter three then presents the implementation of the control system and describes the game design. Chapters four and five are devoted to the studies, with the former first explaining the structure in more detail, while the latter presents the results. In the last parts, the results are discussed and traced back to the hypotheses.

2. Background and Related Work

The topic of accessible computer interaction is by no means new. There are many products that attempt to simplify human-computer interaction, often through simplified physical devices that require less motor ability than standard peripherals. Equally new is the attempt to transfer accessibility to video games. Various approaches and possibilities in this field are discussed in this chapter.

2.1. Auxiliary Peripherals for Computer Interaction

The constancy of the input methodology makes it difficult to introduce new modalities and thus build compatible peripherals that people with motor limitations can operate. Individual organisations such as *Able Gamers* [4] are campaigning for accessibility in video games. Some work goes into simplified mechanical devices for computer interaction, but as individualised as the effects of disabilities are, so too must the devices be. The resulting costs are probably the biggest reason this area is not explored more deeply by the industry.

According to the World Health Organisation (WHO), about one billion people worldwide (15% of the population) live with some kind of disability [27], and about 65 million people (1% of the population) need a wheelchair [26]. The most common causes are stroke, multiple sclerosis, spinal cord injury and cerebral palsy [5]. For some, the effects are more pronounced than for others and not everyone requires hands-free computer interaction. For those who still have some mobility in their hands, a trackball mouse (figure 1a) is a good choice. Alternatively, joysticks (figure 1b) offer similar control and require even less accuracy.

(a) Trackball mouse (b) Joystick mouse

Figure 1: Assistive products for computer interaction [21]

However, none of them can be operated hands-free. One of the best-known products in the field of hands-free yet physically controlled peripherals is the *Quadstick* [22], shown in figure 2 on the following page, which is controlled with the mouth. This project raised over $27.000 on Kickstarter and is available for $449. This device allows paralysed people to play on nearly game controller performance. Nevertheless, not everyone can cope with it or even operate it in the first place.

Figure 2: Quadstick for playing video games operated with the mouth [22]

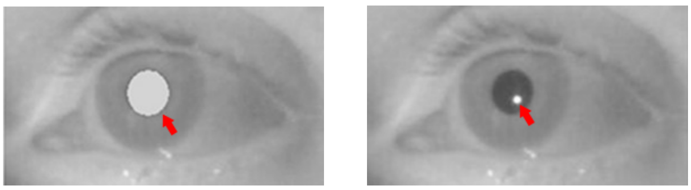
The *Footmouse* [29] in figure 3 is designed for a different part of the body. As the name suggests, this device can be controlled with the feet, which is ideal for those who cannot move their arms but can move their feet. With the help of this device, a computer can be fully controlled as it provides a slipper for mouse input and buttons for various actions. In addition to the model shown, there are many other foot pedals, which are also known from dance games. They can be relatively easily converted for computer interaction if no specially designed device is available.

Figure 3: Footmouse for operation with the feet [29]

2.2. Applications of Gaze Analysis

In 1879, the eye's movement was studied for the first time through direct observation. Louis Émile Javal found that our eyes do not move evenly when we read, but in small steps.

Nowadays, there are highly developed eye tracking devices for this purpose, allowing the user's point of view to be captured precisely. The most common are optical trackers, which use cameras and infrared light to determine the gaze point. They achieve this by analysing the pupil and the corneal reflection in the eyes, both shown in figure 4.



(a) Bright pupil (b) Corneal reflection

Figure 4: Principles of optical eye trackers [8]

Eye trackers are most commonly used to obtain information about the focus behaviour. For example, it is easy to create heat maps of where a user looks in the results of a search engine or which points in a CV stand out first. Since the devices have now reached a high level of accuracy, they can also be used for entirely different purposes. For example, the mouse pointer can easily be controlled with the help of eye tracking, which opens up completely new possibilities for hands-free computer interaction. Another example is the gaze control based keyboard called "EyeSwipe" [17], which uses gaze paths to determine words, analogous to swipe typing on smartphones. Especially people with paralysis get the possibility to operate a computer more efficiently. However, it should be mentioned that visual disorders such as squinting or eye twitching considerably reduce the accuracy of eye tracking. Glasses, on the other hand, are usually not a big problem.

Some modern games use eye tracking as an additional input modality to enhance the gaming experience by making the environment more interactive and responsive to the player's point of view. As of June 2021, 162 games support products from one of the largest manufacturers of eye trackers for games. While "enhanced vision" is progressing, the enormous potential of accessible computer interaction remains largely untapped.

2.3. Games with Eye Tracking

The study of Isokoski and Martin [13] shows how gaze control can be used for aiming in shooter games. They compared the three input methods: *mouse-keyboard*, *mouse-keyboard-eyetracking* and a *controller* (Xbox360). Eye tracking was used for aiming and the mouse and keyboard for movement. The setup with eye tracking performed worse than the setup without, but better than the controller. They described rotations with the avatar of up to 180 or 360 degrees, with the help of which the environment is observed, as particularly difficult to implement. Although gaze control did not perform better, eye trackers are becoming increasingly popular in modern video games as a supporting input modality. For example, in the genres and games [24]:

- Simulator: Euro Truck Simulator 2
- Racing: Formula 1 2020
- Other categories: Assasin's Creed Valhalla, Shadow of the Tomb Raider

While the support role of eye tracking increases, games that only make use of eye tracking are lacking — for a reason. Gaze alone is the same as a mouse without buttons. Only minimalist games can make use of it, such as the 3D flying game in figure 5 on the next page that has been used to compare the performance between mouse and eye tracking [19]. Mouse interaction was considered easier to control and less demanding. However, gaze control was more entertaining.

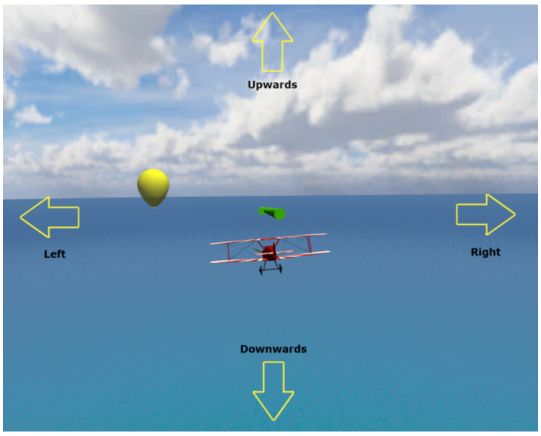


Figure 5: 3D flying game [19]

In an evaluation with a special test environment (figure 6) instead of a game, the performance between mouse and eye tracking was measured for the tasks target detection and target tracking [1]. The gaze input almost reached the performance of a mouse. Two different sizes of targets were used, and the larger one "had a significant effect on the completion time" [1]. The conclusion is that the user interface plays an essential role in the design of a game with gaze control: "aesthetics become ergonomics" [28]. Therefore, elements need to be scaled for user interaction to compensate for the inaccuracy of the eye tracker.

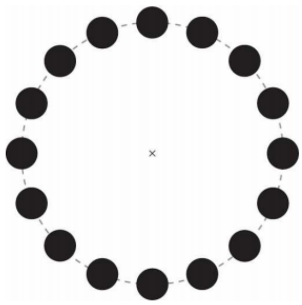


Figure 6: Test environment: circular layout [1]

2.4. Applications of Voice Analysis

Controlling something by voice commands is not a new idea of hands-free interaction. As early as the 18th century, in the play "Les Mille et une nuits" by Antoine Galland, the phrase "Sesame, open yourself" appeared. With the development of technology, the first attempts to make voice control a reality also began. The first research on this topic appeared in the 1950s, but it probably became popular mainly through the science fiction films that were only too happy to simulate this gadget.

Today, there are sophisticated algorithms that can interpret exact commands in many languages. Verbal voice commands offer a vast range of functions, but they are based on the principles of conversation. First, one speaks and then waits for a response or answer.

In 2001, Igarashi and Hughes [11] showed a more direct approach of voice control, which can be compared to a keystroke: Non-verbal voice commands. This may at first seem very limited compared to verbal commands, but it also offers some advantages. Sounds such as humming can be detected and processed by the computer in real time, which means that a reaction can also be displayed in real time. In their work, they mentioned the control of the TV volume as a simple application scenario with the command "Volume up, ahhhhhh" from a combination of verbal and non-verbal commands. During the "ahhh" the volume is increased and stops when it falls silent — just like holding a button. Besides the technical advantages, non-verbal control also allows people with speech disorders to operate computers by voice. This can go much further than just increasing the TV volume, as will be shown below.

2.5. Games with Voice Control

Voice control can be divided into verbal commands (pronounce words) and non-verbal commands (make noises or hum).

Verbal Commands For speech commands, whole words are spoken aloud as if one was talking to the computer. This method has been used to create a game helping people with speech sound disorders by including speech therapy exercises [18].

Speech recognition is, just like eye tracking, popular as a supporting input modality. Some virtual reality games use speech commands for certain tasks as no keyboard is available. There are only a few games that are controlled by speech alone. The modified Tetris version of XiaoJie Yuan and Jing Fan [30] is an example. While speech commands are powerful, they also come with drawbacks. People who have difficulty speaking cannot use them, and in fast-paced games, the delay between issuing a voice command and the reaction of the game is a noticeable disadvantage [20] [23].

Non-verbal Commands Humming can be much better in terms of time and accuracy. In a study comparing both types of voice commands using the game Tetris, non-verbal voice commands were much more efficient and accurate than traditional speech recognition, especially on advanced difficulties [23]. The voice pitch distinguished the commands, as outlined in figure 7 on the next page. Humming a sound from high to low moves the Tetris block to the left. A short buzz turns the block.

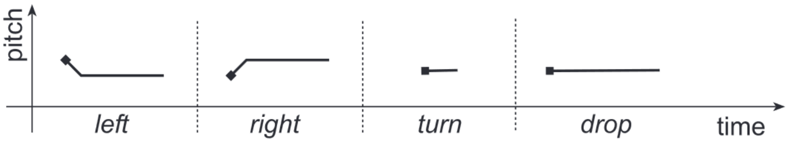


Figure 7: Non-verbal Tetris [23]

Harada et al. [9] developed a tool called *Vocal Joystick Engine*, shown in figure 8, allowing to control the mouse cursor with voice only, but in a special way. Certain sounds have been assigned certain functions, such as the sound "ck" from "click" for a left click. Continuous voice input opens up a way to execute much more complex commands and be more efficient at the same time [9]. One conclusion was that non-verbal commands are up to 50% faster than speech commands [9].

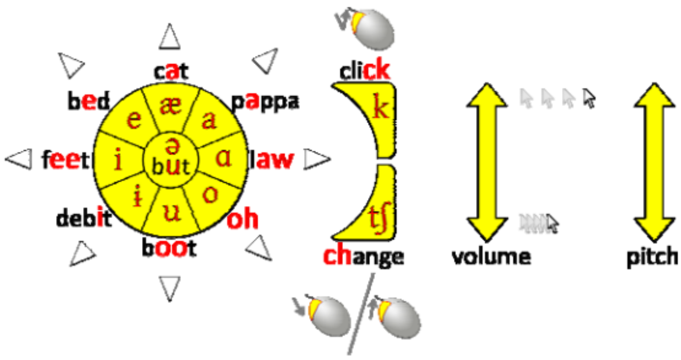


Figure 8: Vocal Joystick Engine [9]

The fact that making sounds can also be more fun is confirmed by the mobile app "Scream Go Hero" (figure 9 on the next page) from *Ketchapp* [15], which uses quiet and loud noises as controls in a jump-and-run style game. If there is a soft noise, the avatar runs, and if there is a loud noise, it jumps. The app reached over ten million downloads in the *Google Play* store. However, the use of only one modality limits the complexity of the game. The avatar can only move in one direction and cannot perform any complex movements. While this makes the game entertaining as a mini-game, it is not a perspective for larger games.

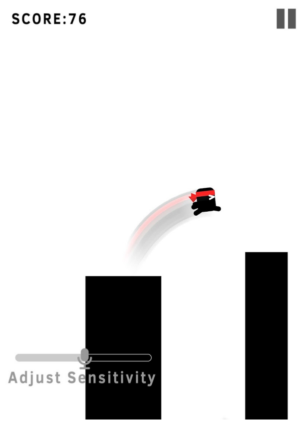


Figure 9: Mobile game: Scream Go Hero [15]

2.6. Combinations of Gaze and Voice

One input method is not enough, which is why we typically use a mouse **and** a keyboard. Following on from this point, O'Donovan et al. [20] developed a game that is controlled by gaze and voice together. Their game *Rabbit Run*, screenshot in figure 10, is a 3D first-person maze that uses gaze for spatial orientation and voice commands for movement and interaction.

Video games in the first-person perspective are games whose displayed image corresponds to the view of the avatar. Thus, the player sees the game through the eyes of the avatar.

To change the orientation in the game of O'Donovan et al., semi-transparent buttons [3] are placed at the edge of the screen. Voice commands such as "walk" and "stop" are for movement. Players with mouse and keyboard achieved better results and navigated the maze faster. However, 75% of the participants said that the eye tracking variant was more enjoyable. They noticed a delay in voice commands, which they solved by adding a manual delay between the user input and executing a command to compensate for the timing.

Figure 10: Rabbit Run [20]

Another study, which focused more on different game mechanics, shows how eye tracking and other hands-free input methods could be integrated into video games [28]. As a result, four mini-games have been developed, each with its unique mechanic. For example, the selection of objects using gaze and voice, or smart text that disappears when the programme recognises that it has been read. This work illustrates that there are many different game types besides first-person that can integrate eye tracking and voice commands as long as the user interface is designed accordingly [28].

2.7. Other Variants of Hands-free Interaction

An excellent example of an alternative game control is again a version of the popular mini-game Tetris. This time it is controlled by gaze and brain activity using a brain-computer interface (BCI) that allows the game to be played entirely hands-free [16]. There was no comparison with other input methods, but the feedback gathered suggested that it was easy to play, and the focus was on strategy rather than control. An important factor, however, is that excitement affects the BCI and thus also the control. So the more exciting a game is, the more difficult it becomes to play. It's a challenge and a fun factor at the same time, much like voice control.

3. Implementation

The primary input method on a computer is the physical mouse because it is suitable for both **indicating locations** (example: desktop environment with mouse pointer; the game moorhuhn) and **indicating directions** (example: any first-person game). A joystick or game-pad is excellent for directional orientation only and is the main input control for consoles (Xbox, Playstation and similar). In this work, we use eye tracking, which is why we focus on indicating locations, as this corresponds to the point of view and thus enables intuitive use.

The most common interaction in gaming environments is spatio-temporal, which combines both spatial and temporal qualities. One style of game that exploits spatio-temporal interaction in a simple but popular design is *Jump and Run*. There are many different games based on this principle, one of which is the famous *SuperMario*. The advantage of such 2D environments is that they can be controlled by locations, which correlate with gaze control. Hence, we investigated the feasibility of hands-free control in a two-dimensional jump and run scenario.

3.1. Movement and Control

The difficulty of hands-free interaction lies in the design of the controls, as computers are not standardised for this type of input. Therefore, a precise adaptation of controls to movements is essential for natural interaction.

3.1.1. Required Types of Movement

A jump-and-run game, as the name suggests, has a manageable variety of movements: Jumping and running:

Run The simplest form of movement is running. The avatar can run to the left and the right. Different speeds are an option but may require more variety in controls. A separate button for sprinting is a standard solution. However, for the sake of clarity, we consider walking at a constant speed.

Jump A jump is a simple upward movement to gain height that allows the player to climb onto objects or take different paths. Jumping to the left or right is usually performed with a run-jump combination, which is the most natural form of jumping, as a long jump in the real world works the same way.

A jump can vary in height or distance. It is usually implemented by variable acceleration time represented by a short or long key press.

Fall When the avatar goes over a cliff, he starts to fall. There are two ways to implement falling: Either control is allowed in the air, or the avatar falls straight and without any control. The lack of the possibility to move in the air is more in line with reality. However, in video games, it can be very advantageous to move in the air, especially in a jump-and-run game where the challenge is dexterous movement.

3.1.2. Available Controls

The standard input devices for PCs are a mouse and a keyboard, whose functions are straightforward and familiar to most people. Spatial movement of the mouse pointer by physically moving the mouse and temporal actions by pressing keys on both the mouse and the keyboard is the standard way of interaction. With alternative input methods, the control implementation is usually not immediately apparent. They often have to be mapped to the principles of a mouse or a keyboard.

Eye tracking With eye tracking, the user's point of view on the monitor is determined with the help of a corresponding device. It is very similar to the characteristics of a mouse but is less precise, which is due to the way our eyes and our vision work [25]. Thus, eye tracking instead reflects an area at which the user is looking. With good calibration, however, eye tracking can undoubtedly be used as a mouse. It just needs to be taken into account in, for example, the icon size on the desktop or the scaling in video games. Gaze control thus offers good qualities for spatial control in the sense of indicating locations.

Since eye tracking has neither a left nor a right-click, one quickly encounters the Midas Touch Problem. It describes the difficulty of determining an interaction since our eyes move or focus both consciously and unconsciously [14]. Often auxiliary means are used such as [14]:

- Dwell time: Looking at a spot for a certain time to interact
- Wink: One wink corresponds to a click
- Extra input modalities: For example, a button or voice control

Non-verbal voice Non-verbal voice commands are sounds such as humming or buzzing. Compared to verbal voice commands, they can be processed immediately, and the programme does not have to wait for the end before interpreting the input [23]. This fact shortens the programme's reaction time considerably and makes it much more responsive — a tremendous advantage in games.

This type of control can be implemented as binary input (sound or no sound), but also in more complex forms such as:

- Pitch and its progression
- Tone duration
- Tone volume and its progression

In this thesis, the focus is on a binary implementation of short and continuous voice commands. They offer enough control options and require hardly any training, which makes them very intuitive to use.

3.1.3. Mapping Controls to Movements

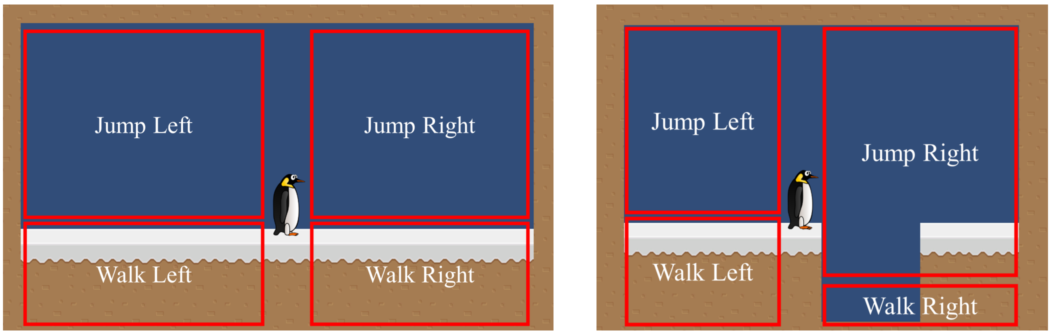
Typically, the arrow keys and the space bar are used to control jump-and-run games. However, these would not make use of eye tracking and could not be compared with it. Therefore, we are now introducing a new concept based on indicating locations. This concept allows control with both conventional and hands-free interaction techniques.

Indicating locations The *control point* determines the direction and type of the movement. This point can be indicated either with the mouse or gaze. The control system is shown in Figure 11 and means the following: For the movement direction applies:

- Point to the left of the avatar ⇒ Left
- Point to the right of the avatar ⇒ Right

For the movement type applies:

- Point is on the ground or below ⇒ Walk
- Point is above the ground or over a cliff ⇒ Jump



(a) On flat ground (b) At a cliff

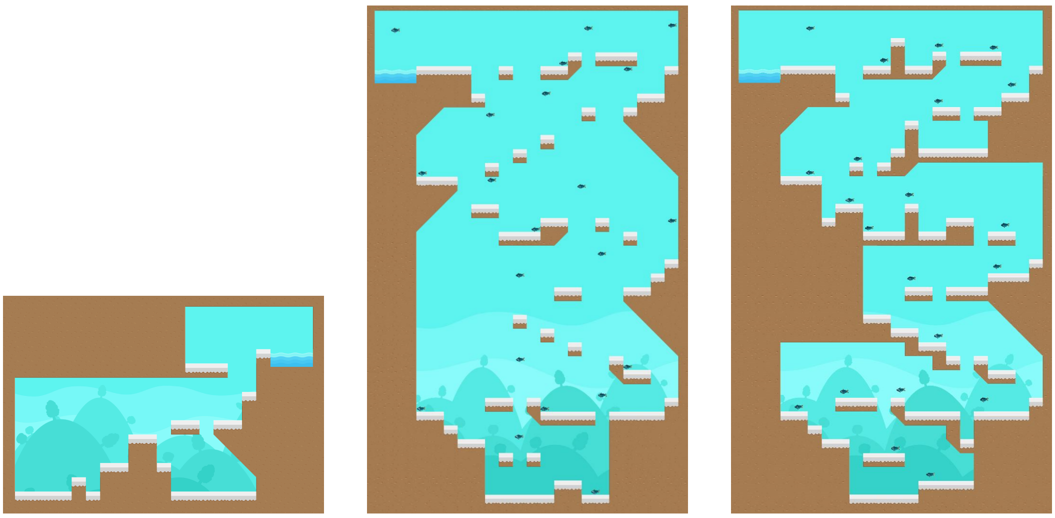
Figure 11: Spatial control in the game

Temporal actions The direction and type of movement can be determined by indicating locations, but no movement is executed. For that, the non-verbal voice commands become essential. If an action is triggered, the movement specified by the control point is carried out. The beginning of a sound or hum corresponds to the pressing of a key, the silencing corresponds to the release of a key. This enables a temporal-stepless triggering of actions and makes small and long steps or jumps possible.

The advantage of this type of control is not only that hands-free implementation is possible, but it can still be operated with the mouse and keyboard, allowing the different modalities to be compared with each other.

3.2. Game Design

The basic structure of the game is different from conventional jump-and-run games. Instead of reaching the end on the far right, the player has to make it to the top, making eye control more crucial as the player has to keep moving left and right. On the way there, hills and cliffs have to be overcome without falling. The finish line is represented by water into which the penguin has to jump. Before the actual game begins, there is a small training world. In this world, the players can familiarise themselves with the controls. This makes the game easier to play and reduces the training effect when playing several times, as experience can be gained at the beginning. Figure 12 shows a comparison of the main and the training world. The actual game world is a lot bigger, but the training world is sufficient for its purpose. The world was created with a free package from Bayat Games [7]. The simple structure with blocks is ideal for eye tracking, as no super precise control is required, it offers little distraction, and it can be scaled as desired.

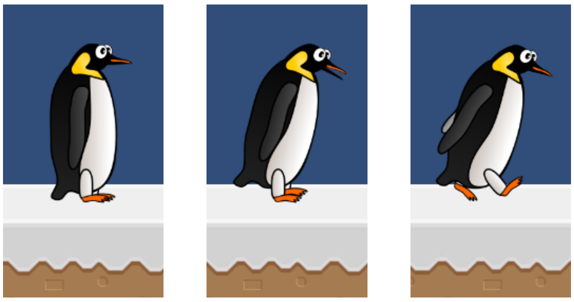


(a) Training world (b) Original level (c) New level

Figure 12: Comparison of the original and adapted level design

The idea behind the game is that it is easy to understand, requires little training, and still offers some challenge. This challenge is difficult to assess, as a too easy game produces blurred data, while one that is too difficult reduces the fun and leads to frustration. The pilot study, which will be explained in more detail in the next chapter, showed that the game world was too difficult, so it was adapted. The differences can be seen in figure 12 on the previous page. More blocks were added to reduce falling in general. Furthermore, the level has been adjusted so that when the avatar falls, it is noticeable, but it does not fall too far, which allows quick recovery. The figure also shows that the positioning of the fish has changed. Initially, they were intended as an additional challenge and were therefore positioned in difficult spots. In order to simplify playing, they have been placed in the new design in such a way that they support the player in some places by serving as a focus point. The motto, "Look at the fish, and you make the jump", makes some jumps more manageable.

A penguin was chosen as the main avatar to make the game attractive and amusing. A waddling flightless bird increases the fun of the game and thus the bond between player and avatar. The penguin and the fish were designed and animated by ourselves. In order to make the movement behaviour as transparent as possible to the player, the penguin uses its body posture to indicate whether it will walk or jump. If it leans forward slightly (figure 13b), it will walk when an action is triggered.



(a) Penguin idle

(b) Penguin indicating walking

(c) Penguin walking

Figure 13: The penguin

The movement of the camera, which means the field of view, is also different from conventional. Usually, the avatar is permanently in the centre of the image, which was our first approach. However, this leads to the problem that when jumping, the gaze point is no longer in the original spot, and the player thus tends always to jump further than intended. It also leads to more distractions because the world is constantly in motion, which can be fatal when using eye tracking. Therefore, the camera movement was set so that it always follows with a delay, i.e. during a movement, the image is almost still, and as soon as the avatar has landed, the camera moves forward.

3.3. Technical Details

The game was developed with the Unity game engine (2019.4.16f1). The sound detection was integrated directly into the game, whereas the gaze control was reduced to the functionality of the mouse pointer. This means that the game can be controlled solely with the mouse as the primary input and with the keyboard or voice as the secondary input. In order to be able to control the game with the gaze, more software is required. First of all, the core software (version 2.16.6) from Tobii was used to be able to address the eye tracker. Next, the programme *Gaze Point*, also from Tobii, was set up to control the mouse pointer with gaze. This allows the physical mouse to be replaced by the eye tracker. With the help of this software, games can be implemented particularly efficiently because the eye tracker is not directly addressed and therefore does not have to be taken into account in the development of the control system. Many participants in the studies found it easier when they could see the gaze point. This is a feature of the core software that displays a floating circle (figure 14) representing the approximate viewpoint.

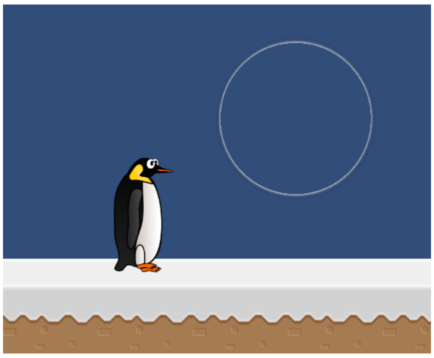


Figure 14: Gaze trace provided by the Tobii software

4. Evaluation

Two studies were conducted to gain an accurate insight into the conscious and unconscious user experience. The former is used to compare the hands-free variant to conventional methods of computer interaction. The second study focuses on the feasibility for people with motor and partly also speech impairments. It determines whether the game control presented can be used effectively by the target group. In this chapter, the structure and procedure of the studies are explained in more detail, and information about the participants is presented.

4.1. Apparatus

We used the Eye Tracker 4C from Tobii and attached it to a laptop with a screen diagonal of 13.5 inches which is in range of the supported screen size. The laptop's built-in microphone array was used for audio capture, so no extra headset was necessary, making our control system more accessible to use. Especially for some of the participants with disabilities, as it did not require any modification of the headrest. In addition, the microphone arrays sensitivity could be fine-tuned to filter as much background noise as possible while maintaining accurate voice detection. A mouse pad was used to avoid interference from different table materials when playing with mouse and keyboard. We increased the size of the mouse pointer by setting its size in *Windows* to four to improve visibility in the game. The chosen key for performing actions was the space bar, as this is the easiest to use and is more differentiated than if a mouse button had been used. It also allows the analogy of using eye tracking with a button instead of voice.

4.2. Procedure

The study took place both at a fixed location and at the participants' homes, as it was more convenient for some participants due to the pandemic, especially for those with disabilities. Hygiene measures such as disinfecting the equipment and spacing rules were observed throughout the study. In some cases, masks had an effect on the eye tracker for participants wearing glasses, as they fogged up more easily. Therefore, the participants were free to take off their mask while playing.

In the beginning, the participants were shown videos about the game in general and the controls. This way, each participant received the same explanation and guidance. A full game run consists of completing the small training and the main world, both by the penguin jumping into the water. The first time the participants took their time in the training world to familiarise themselves with the controls. After that, the training world only served to get back into the game in all further runs and was quickly mastered. Since a lot of practice could be done initially, no significant training effect on the game was expected. Any difference in performance between the two control methods should therefore be due to the type of control.

In the first study and after the first two runs, participants were shown a third video explaining the other control method, with which they also played two runs. For counter-balancing, 8 of the 15 participants played first with **Mouse+Keyboard** (*MK*) and then with **Gaze+Voice** (*GV*). Finally, the questionnaire comparing the two control types was completed. In the study with people with disabilities, the participants only answered the questions about the hands-free variant.

4.3. Measurements

The questionnaire consists of three parts: the participant information, the experience and the control. The experience section is about enjoyment, challenge and focus. In contrast, the control part focuses on the individual modalities and their rating. The questions in both of these sections are answered for both *MK* and *GV*. The complete questionnaire with scales is listed in table 3 on page 42. There are many well-structured game experience questionnaires [10]. However, they focus on the perceived experience from the game only. The control and the experience it creates are usually not taken into account. However, it is precisely the perception by the control that is the focus of this work, which is why we had to compile a new questionnaire based on recommended questionnaires [12][2] and previous work [20].

The game took quantitative measurements while the participants played. Only the main world was considered, while movements in the training world were not recorded. The following data was collected for each run:

- Time to reach the finish line
- Time the avatar was in motion
- Number of collected fish
- For every movement has been collected:
  - Time the avatar was in motion
  - Time how long the button was pressed and held down
  - Time to the last movement
  - Start coordinates (take-off point)
  - Stop coordinates (landing point)
  - Movement was a walk or jump

This data could then be processed further so that, for example, the average distance of a movement or the fatal movements that led to a fall could be determined. With the help of the coordinates, each course of the game could be reconstructed for analysis.

4.4. Pilot Study

First, a pilot study was carried out with 13 young people aged between 11 and 16. It was noticed that the game was a lot too difficult because hardly anyone had made it to the finish line, which had an apparent effect on the fun while playing. Therefore, we adapted the level design (figure 12 on page 16) to make the game easier. In addition to the challenging game design, many external influences, such as high ambient noise, made playing hands-free significantly more difficult. Based on this experience, we have further optimised the microphone settings and ensured good environmental conditions. The task was to assess which microphone sensitivity was sufficient to reliably pick up humming while at the same time ignoring background noise.

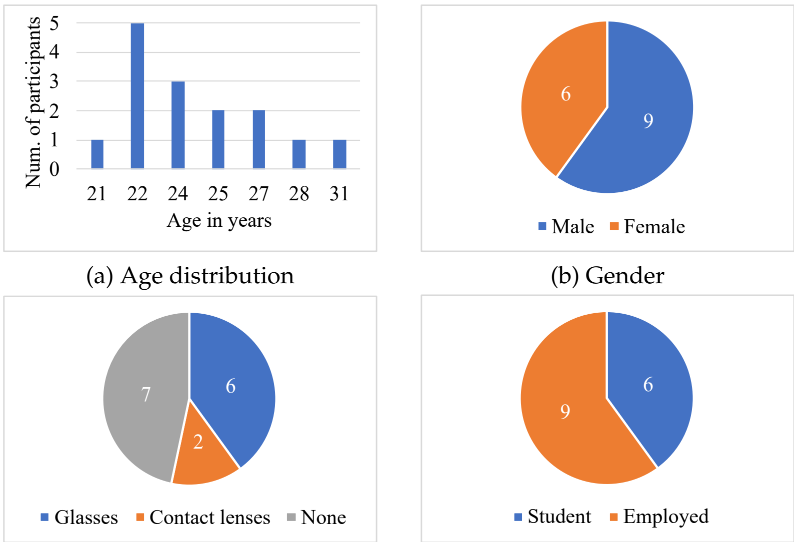
4.5. Main Study

The first study focused on comparing the modalities *MK* versus *GV*. The control with *MK*, which is familiar to the participants, serves as a reference line with which the difference in performance can be determined and thus the practicality confirmed. *GV* represents the maximum of hands-free interaction, which is why a combination of eye tracking and keyboard, although practically relevant, is not investigated. The effects of disabilities are very different, and substitutions to physical devices are always possible. For example, a physical button can easily replace voice control for players who can move their fingers but have difficulty speaking. Further reasons for an investigation of only two modalities are the lower effort for the participants, as two control systems are easier to compare than three, and the total time spent by a participant averaging 30 instead of 45 minutes also reduces the degradation of motivation.

4.5.1. Participants

The first study involved 15 people averaging the age of 24. The participants consisted of friends and acquaintances who are either studying or working and are active in various fields other than computer science. Five of them said they are well experienced with video games, while the other ten have low to medium experience. None of the participants used eye tracking or voice-activated programmes before. More demographic details are shown in figure 15 on the following page.

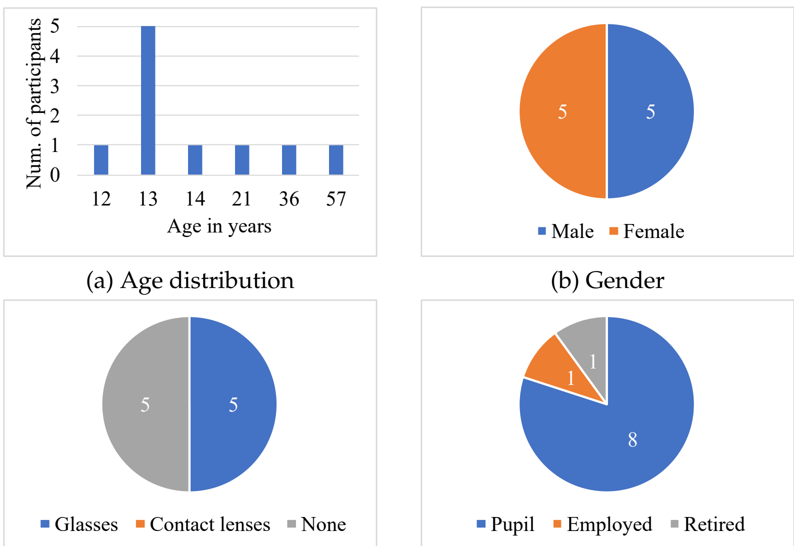
We reached a comparable number of participants for the second study through various institutions such as special schools and associations for people with disabilities. Ten people with motor and partly speech impairments took part and said they have about medium experience with video games. However, this "medium" cannot be compared with the one from the first study, as the situation is entirely different. Through discussions with the participants, it was determined that these are elementary games that cannot be compared to typical popular video games. However, it also shows that there is interest in games and that the participants have knowledge by which they can judge our developed game better. The most common diseases among the participants were duchenne muscular dystrophy, cerebral palsy and spastic tetraparesis. None of the participants used eye tracking or voice-activated programmes before. Detailed participant information is shown in figure 16.



(a) Age distribution (b) Gender

(c) Visual aid (d) Occupation

Figure 15: Study 1: Demographic information



(a) Age distribution (b) Gender

(c) Visual aid (d) Occupation

Figure 16: Study 2: Demographic information

5. Results

This chapter presents the data collected for both studies. In the end, the two studies are also compared with each other, using only the data from the first study in which *GV* was played first to keep the comparison fair.

5.1. Study 1: Comparison of Input Modalities

Results of the first study comparing the two modalities *MK* and *GV* each in combination and individually. First, the collected data is analysed and then the questionnaire is examined.

5.1.1. Quantitative Measures

All values and the significance of the differences are visualised in table 1. A colour difference, i.e. green to red, between *MK* and *GV* means that there is a significant difference in the comparison of the two control modes. If the colours differ within a modality, i.e. different shades of green or red, it means that there is a significant difference between the two runs within that modality. Detailed information about the two-tailed t-tests can be found in table 4 on page 43, 5 on page 43 and 6 on page 44.

Measure	MK 1	MK 2	GV 1	GV 2
Time to reach finish line (min:sec)	1:41	1:22	3:06	2:18
Proportion of motion to total time	63.09%	66.95%	47.81%	51.90%
Time between movements (ms)	495	427	992	850
Number of fatal movements	2.2	1.87	4.8	3
Walking percentage	20.88%	20.26%	25.05%	23.48%
Duration of temporal action (ms)	637	661	716	742
Distance of a movement (blocks)	2.19	2.31	1.93	2.05

Table 1: Study 1: Average values of collected data. Colours indicate significant difference for MK-vs-GV, MK run 1-vs-2 and GV run 1-vs-2.

Overall, the players finished 44% faster with *MK* (1:32 min) than with *GV* (2:43 min). While there was no significant improvement within *MK* ($p = .2547$) and *GV* ($p = .0506$), players tended to improve more with the hands-free control, which makes sense as this was new to the participants, resulting in a larger learning curve. The proportion of motion in the total time is nearly identical for both runs with the conventional input control ($M_{MK1}^{\text{motion}} \approx 63\%$, $M_{MK2}^{\text{motion}} \approx 67\%$, $p = .3192$). It is, however, distinctly larger ($p < .001$) compared to the hands-free variant, which means that players were more active when playing with mouse and keyboard. In the second run with *GV*, the ratio increased to about 52% — a clear improvement on the first run ($M_{GV1}^{\text{active}} \approx 48\%$, $p = .0465$). Analogously, the proportion is reflected in the time between the individual movements. With *GV* ($M = 944$ms), participants spent twice as much time compared to *MK* ($M \approx 460$ms). Here, too, a significant ($p = .0427$) improvement of over 140ms can be observed in the second run of the hands-free variant. In case of inaccurate movement, the avatar could fall down. With the help of the collected data, each complete game sequence could be reconstructed and thus the fatal jumps could be counted. With mouse and keyboard, the players fell about two times, while with eye tracking and voice control they fell almost four times on average. Within both control variants, there was no significant improvement in the second run ($p_{MK} = .7020$, $p_{GV} = .0897$). The percentage of *walking* in all movements was approximately the same across all modalities. With 21% for *MK* and 25% for *GV*, it was higher for the hands-free variant, but not significantly ($p = .2531$). Participants often mentioned that walking with eye tracking was easier, but this seems to have had only a small effect. When conducting the study, it was observed that the participants made longer temporal inputs with voice control than with the keyboard. This is also reflected in the data: the average time of action, i.e. the duration for which a key was pressed, lasted 650ms with the keyboard and 724ms with voice control ($p = .0407$). At first glance, one could conclude that the distances of the movements with *GV* are greater, but looking at this parameter shows the opposite. On average the distance of a movement was 2.25 blocks with *MK* and 1.97 blocks with *GV* ($p = .0027$). A block is the fixed unit in the game world. The last two parameters of the table thus again indicate better dexterity with mouse and keyboard.

The heat maps in figure 17 on page 26 represent the movement-stop coordinates of all data of a control type. Overall, the two diagrams look very similar, which in principle points to similar movement sequences and thus to equivalent control. For a closer look, ten areas were marked which will now be explained:

1. The most prominent area is the first one. There are much darker spots in *GV*, which has to do with the lower-left corner. There are many more data points, which means that many players fell when jumping over the gap above. By repeating the route, the number of data points increases.

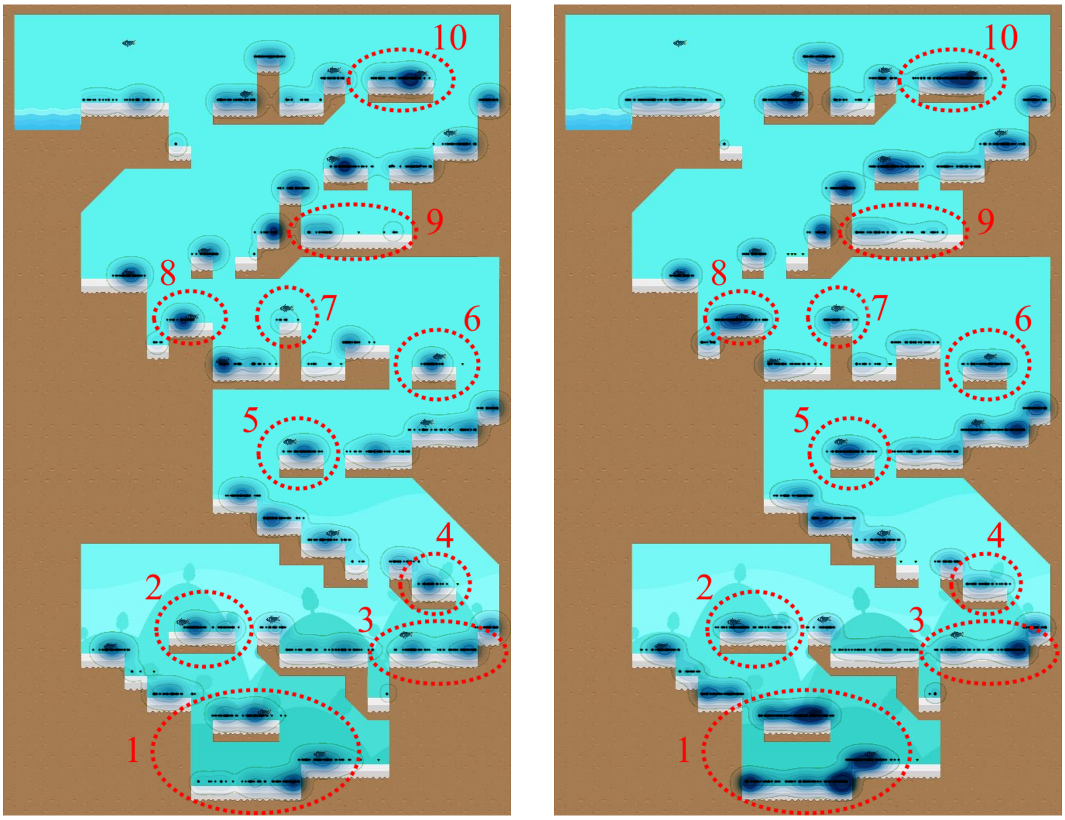
2. This area reflects the whole diagram. In *MK*, the points are much closer together, which means that all the players moved in a similar way. This is unlikely to be a coincident, so we can conclude that *GV* seems to be less accurate.

3. Similar to the last area, the data points are closer together, plus this area is split in two in *MK*. There is a clear gap in the middle, indicating that the players covered this path with two movements. In *GV*, the dots are more spread out

and the darker spot on the right is the result of many players not making the jump from area three to four.

4. The denser movement of the players with *MK* is again confirmed here. This pattern does not apply to every area, but to many, which further confirms the assumption of more precise control with *MK*.
5. This area shows that the two control variants are also very similar. There is again a denser accumulation at *MK*, but compared to other places within the tolerance.
6. The difference is more obvious in this area, as the data points in *GV* are distributed over the entire surface, whereas in *MK* the points gather exclusively in the centre.
7. In some places, the different control also led to quite different behaviour, as this area shows. While with *GV* most jumped on the post, with *MK* most skipped the post completely. Only a few landed on the post and proceeded from there.
8. In this area, the denser movement of *MK* is present again, and we can also observe more fatal movements with *GV* near this area. At the bottom left of the circle marking area eight, we can see many more data points in the pit. Since the player has to jump back to area eight before reaching the next higher point, this area is darker in *GV*.
9. The left side of this area is very similar between the two modalities. However, the right side shows more data points in *GV*, which means that the first gap was a problem for both control types, while the second gap was mainly a problem for players using the hands-free variant.
10. Here, too, the often made statement about similar movement sequences with the classical control is confirmed. Not only are the points more distributed in *GV*, but also the core (dark blue) extends over a larger area.

In summary, the graphs are similar overall, but show differences in detail and indicate more precise control with mouse and keyboard. The movements of all participants with *MK* were very similar, indicated by denser data points. With *GV*, the points are usually more widely distributed and fatal movements occur more often.

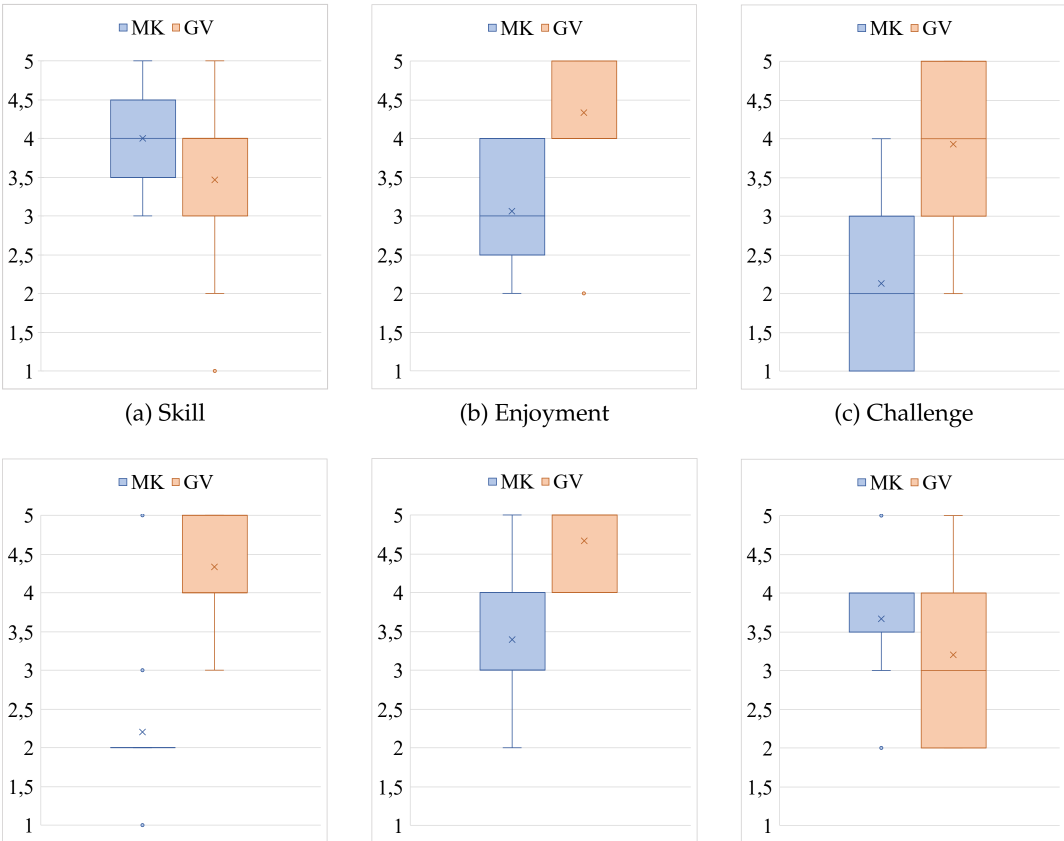


(a) Mouse + Keyboard (b) Gaze + Voice

Figure 17: Heat maps comparing *MK* and *GV*

5.1.2. Qualitative Measures

The core of the questionnaire is divided into *Experience* and *Control*. The answers of the first part are visualised in figure 18 on the next page. When asked how well the participants thought they had played (18a), the conventional *MK* ($M = 4$, $SD = 0.75$) performed only slightly better ($t = 1.47$, $p = .15$) compared to *GV* ($M = 3.47$, $SD = 1.19$). In contrast, the result of the second question on how much fun the players had with the respective control methods (18b) is significant ($t = -4.29$, $p < .001$). Here, hands-free control ($M = 4.33$, $SD = 0.82$) achieved much higher scores than the familiar mouse and keyboard ($M = 3.07$, $SD = 0.8$). The same applies to how challenging (18c) the game was to play ($t = -4.71$, $p < .001$). While the traditional interaction methodology ($M = 2.13$, $SD = 1.06$) is not very difficult, the participants found eye tracking in combination with voice control ($M = 3.93$, $SD = 1.03$) to be a greater challenge. Question number four about the excitement (18d) during the game confirms that the increased challenge can definitely contribute to joy ($t = -6.96$ $p < .001$). While *MK* scored just 2.2 on average ($SD = 0.94$), *GV* achieved a rating of 4.3 ($SD = 0.72$). Only a few found the classic computer interaction similarly entertaining. However, the higher level of fun also has its price: concentration. That is the result of the fifth question on how much the participants focused while playing (18e). The participants had to concentrate only moderately on *MK* ($M = 3.4$, $SD = 0.98$), but very strongly on *GV* ($M = 4.67$, $SD = 0.49$). There was no significant difference ($t = 1.20$, $p = .24$) between the control types in the category of feeling the time (18f). The time spent playing the game passed at a similar average rate for *MK* ($M = 3.67$, $SD = 0.97$) and *GV* ($M = 3.2$, $SD = 1.15$), but responses for the hands-free variant were more widely distributed.

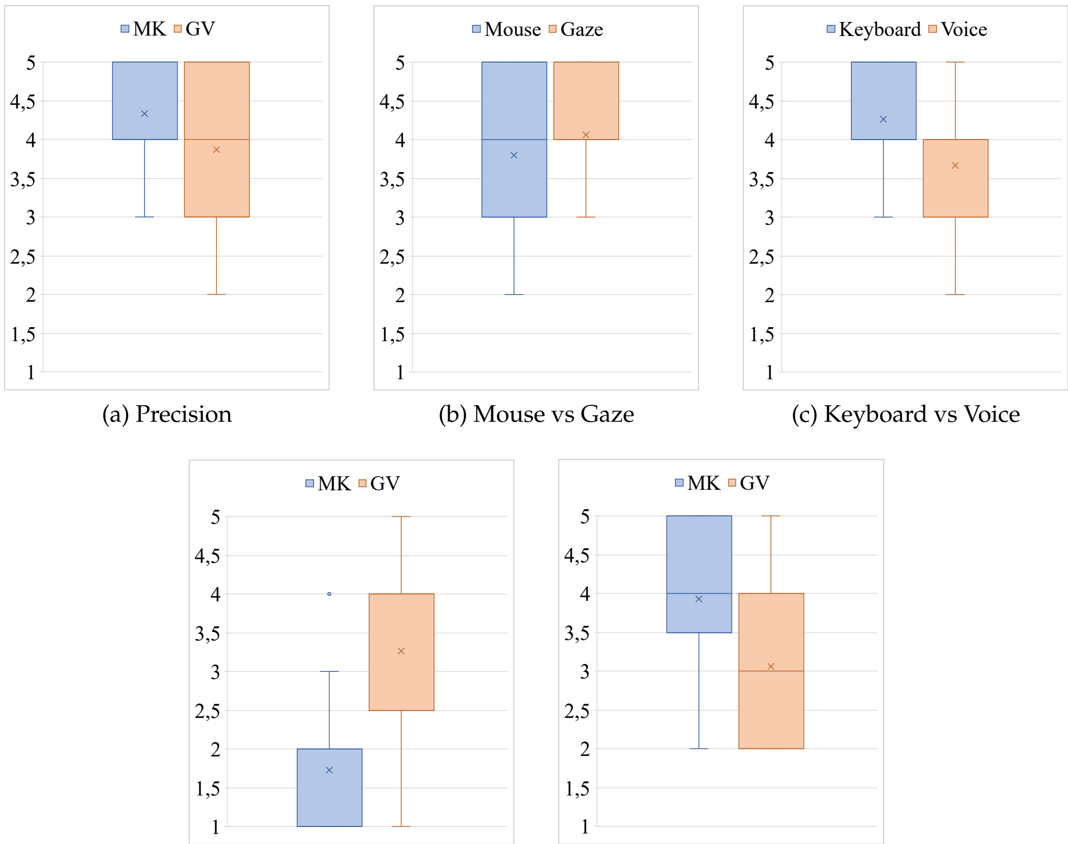


(a) Skill (b) Enjoyment (c) Challenge

(d) Excitement (e) Focus (f) Sense of time

Figure 18: Study 1: Questionnaire result on experience

The results of the second part are shown in figure 19 on the following page. The combination of modalities was examined, as well as each individual modality in direct comparison. When considering the entire control system, i.e. *MK* and *GV*, they performed very similarly ($t = 1.47$, $p = .15$) in terms of the precision of the penguin's control (19a). *GV* ($M = 3.87$, $SD = 0.99$) follows closely behind *MK* ($M = 4.33$, $SD = 0.72$), but is more widely distributed. In the direct comparison of mouse against gaze (19b), there is also no obvious winner ($t = -0.77$, $p = .45$). On average, the mouse scored 3.8 ($SD = 1.15$) and eye tracking only slightly more with 4.07 ($SD = 0.71$). The evaluation of keyboard versus voice is a little bit different. The average values ($M_{MK} = 4.27$, $SD_{MK} = 0.71$, $M_{GV} = 3.67$, $SD_{GV} = 0.9$) are again close to each other ($t = 2.03$, $p = .05$), but in figure 19c there is definitely a tendency towards the keyboard, which is confirmed by the verbal feedback we received. The reason for this result can be found in Figure 19e, which shows the assessment of exhaustion by the control. Eye tracking and voice commands are with 3.27 ($SD = 0.96$) significantly ($t = -4.06$, $p < .001$) more strenuous than mouse and keyboard with 1.73 ($SD = 1.1$). Despite the higher effort, the participants felt that they progressed significantly ($t = 2.23$, $p = .03$) faster (19e) with *MK* ($M = 3.93$, $SD = 1.03$) than with *GV* ($M = 3.07$, $SD = 1.1$). In an overall comparison, two-third of the participants preferred gaze and voice over mouse and keyboard for this game scenario.



(a) Precision (b) Mouse vs Gaze (c) Keyboard vs Voice

(d) Exhaustion (e) Pace

Figure 19: Study 1: Questionnaire result on control

5.2. Study 2: Investigation of Feasibility for People with Disabilities

The results of the second study, which investigates the practical usability for people with disabilities, will now be presented. Again, first the quantitative measurements and then the qualitative feedback will be discussed. Only the data from five of the ten participants could be used for a more in-depth analysis. For the others, there were various difficulties, which is why a complete game run was only possible to a limited extent or not at all. One reason for this was the eye tracker, which could not work correctly depending on the participants' limitations (e.g. squinting, eye tremor). These participants were able to get an impression of the game and the controls, but a more precise data analysis would be prone to errors. That is why the data set for the quantitative study consists of only five participants ($N = 5$), but that of the questionnaire consists of ten participants ($N = 10$).

5.2.1. Quantitative Measures

Following the same scheme as in the previous section, table 2 shows all values and the significance of the differences. To keep the comparison fair, only the data from study 1 was used, in which the participants had first played with *GV*. A colour difference, i.e. green to red, between study 1 (*S1*) and study 2 (*S2*) means that there is a significant difference between the results of the studies. If the colours differ within a study, i.e. different shades of green or red, it means that there is a significant difference between the two runs within that study. Detailed information about the two-tailed t-tests can be found in table 7 on page 44, 8 on page 45 and 9 on page 45.

Measure	S1 (1)	S1 (2)	S2 (1)	S2 (2)
Time to reach finish line (min:sec)	3:14	2:16	7:01	4:47
Proportion of motion in total time	45.63%	52.39%	39.33%	37.52%
Time between movements (ms)	1125	850	1588	1660
Number of fatal movements	4.7	3	7.8	5.6
Walking percentage	18.90%	18.10%	19.12%	16.54%
Duration of temporal action (ms)	749	763	778	732
Distance of a movement (blocks)	1.91	2.05	1.73	1.88

Table 2: Study 2: Average values of collected data. Colours indicate significant difference for S1-vs-S2, S1 run 1-vs-2 and S2 run 1-vs-2.

Overall, it took the participants of the second study (5:54 min) about twice as much time as the participants of the first study (2:55 min). Although the participants improved on average in the second run, this is not significant for either *S1* ($p = .09$)

or *S2* ($p = .2085$). The proportion of motion in the total time is higher in *S1* ($M_{S1} = 0.48$, $M_{S2} = 0.38$), which explains the significant difference in the time between movements ($p = 0.012$). The participants in *S1* rested about 1.0 second, whereas those in *S2* rested about 1.6 seconds before issuing the next command. The players with disabilities thus progressed more slowly overall. However, the increased time to complete results not only from the activity, but also from the increased number of fatal movements ($M_{S1} = 4$, $M_{S2} = 6.7$), which extend the playing time. A look at the last three parameters shows that apart from time, the players in both studies were very similar in the way they moved. Of all movements, the participants in *S1* walked 19% and in *S2* 18%, an insignificant difference ($p = .7791$). The same applies to the duration of temporal actions, which are almost the same ($p = .9047$) at 743ms in *S1* and 755ms in *S2*. The average distances of a movement are equally similar ($p = .2478$). In the first study this was 1.95 blocks and in the second study 1.8 blocks. In both studies, the movement distance increased by about 0.14 blocks in the second run.

For visual comparison of the two studies, heat maps are again shown in figure 20 on the next page. The map for study 1 only contains the data of the participants who first played with GV. It should be noted that due to the different data sets, each heat map has its own colour coding. They are adapted to each other, but in a direct comparison it should be taken into account. Overall, the two graphs are very similar, which confirms the equal movement behaviour. Nevertheless, there are also minor differences in detail which are explained below:

1. Immediately after the first jump, it can be seen that in *S2* the players tended to move a little further to the right. This is not helpful for the next jump and is thus a sign of imprecise control.

2. This area illustrates that players in *S1* liked to skip this step, whereas players in *S2* took every single step. The point here is to balance time gain against safety and illustrates the different approaches of the players.

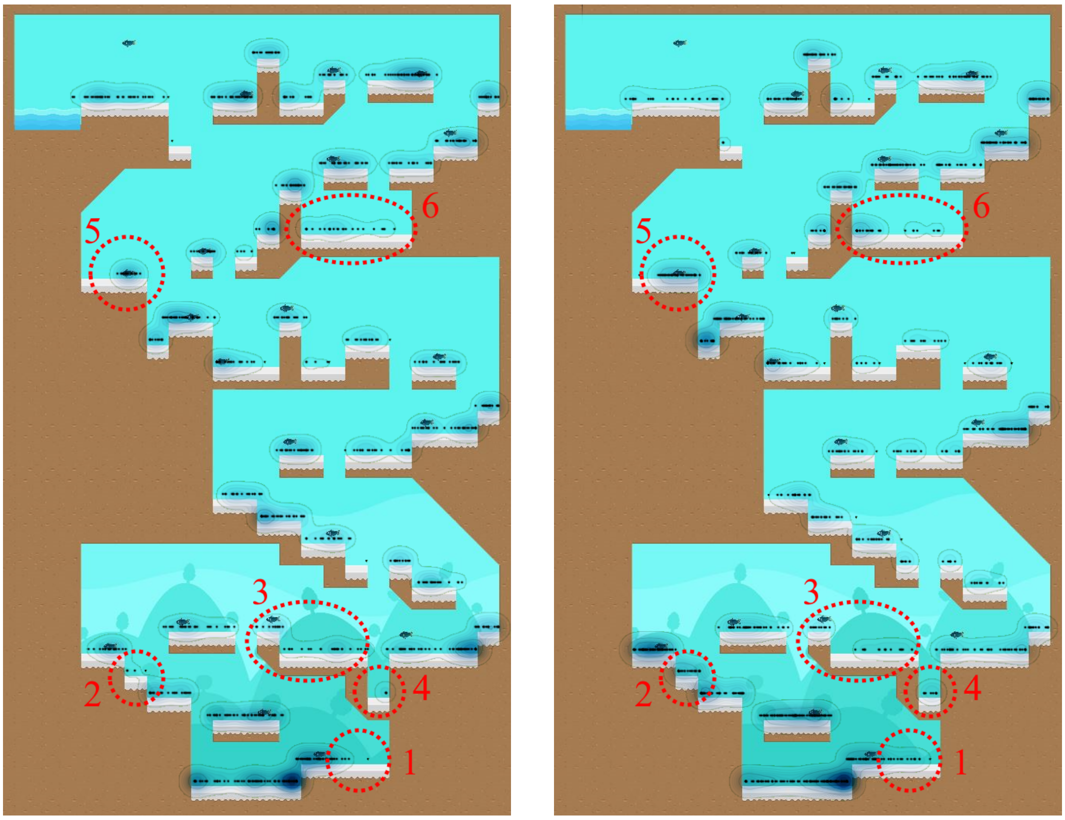
3. In contrast to area 1, it is easy to see that the participants of *S2* had moved more precisely at this location. This can be recognised by the denser distribution of the points.

4. In table 2 we saw that participants in *S2* had more fatal movements. This can be seen very well in this area, as there are more points in *S2* than in *S1*.

5. Once again, a denser cluster of dots is clearly visible in the left image, which means that the players have moved in the same way, implicitly indicating more precise control.

6. In both studies, the participants had similar problems in this area. In *S2*, however, the participants came out of the pit more elegantly.

It can be summarised that the two heat maps are very similar, which indicates a comparable control behaviour. The participants of the second study generally played a little slower, but the movement sequences were the same as in the first study. If we take into account that the participants of *S2* had an increased level of difficulty due to the impact of their disabilities, we can conclude that they perform competitively.



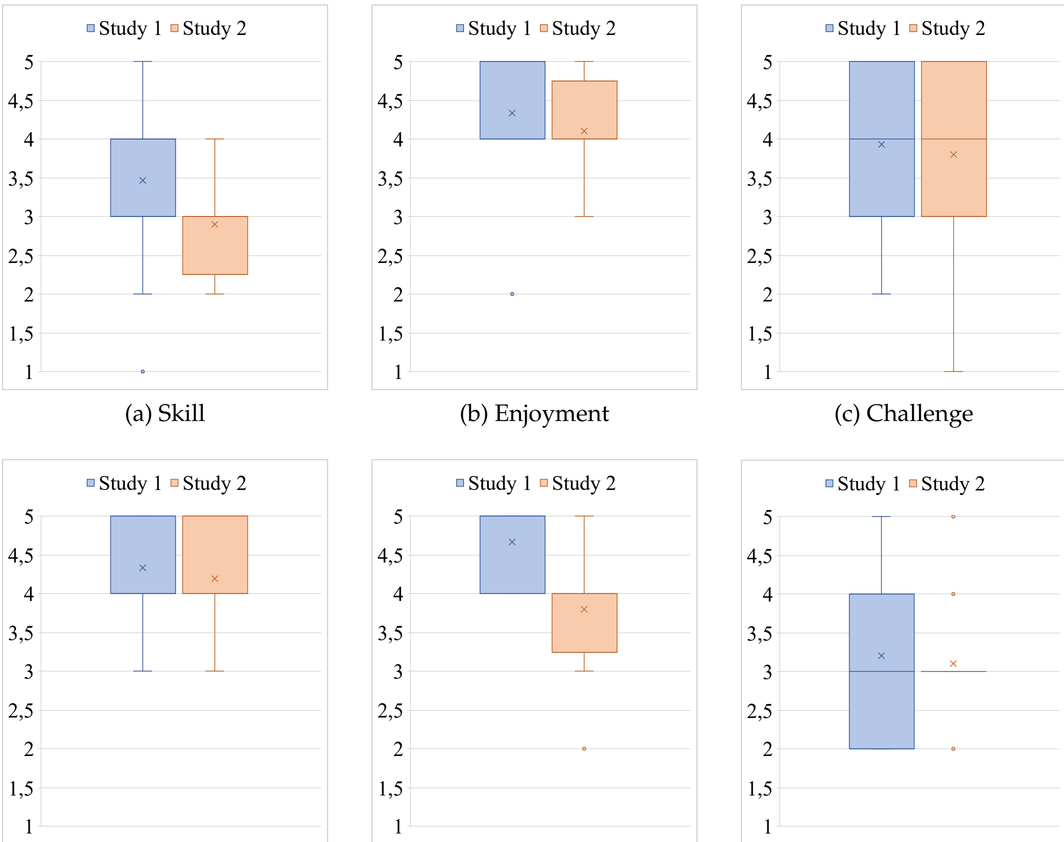
(a) Study 1 (*GV* first) (b) Study 2

Figure 20: Heat maps comparing study 1 (*GV* first) and study 2

5.2.2. Qualitative Measures

The questionnaire was almost the same as that of the first study. The only difference is that the questions about comparison with mouse and keyboard were omitted. In order to be able to classify the answers, the results of the first study on *GV* are listed again. These include all participants of the first study, as it does not matter for the questionnaire which control type was played first.

When asked how well the participants felt they had played (21a), the participants from *S2* ($M = 2.9$, $SD = 0.74$) had rated themselves slightly worse ($t = 1.34$, $p = .1929$) than those from *S1* ($M = 3.47$, $SD = 1.19$). However, everyone was unanimous ($t = 0.73$, $p = .4748$) on the question of whether the game was fun (21b). With a score of 4.33 ($SD = 0.82$) for *S1* and 4.1 ($SD = 0.75$) for *S2*, these ratings are almost identical. This is pleasing, as a game that is not fun would not be of much use. Furthermore, all participants were of the opinion that the game is quite a challenge with this control system ($t = 0.28$, $p = .7794$). The graphs in figure 21c look nearly identical and *S1* ($M = 3.93$, $SD = 1.03$) and *S2* ($M = 3.8$, $SD = 1.32$) only differ in minor details. The same applies to the excitement (21d) while playing, which was very high for all participants ($t = .44$, $p = .6672$). The second study achieved a score of 4.2 ($SD = 0.79$), which is close to the first study's score of 4.33 ($SD = 0.72$). The answers to the question about focus (21e) were different ($t = 3.08$, $p = .0053$), with participants in the first study ($M = 4.67$, $SD = 0.49$) feeling they had to focus more on the game than participants in the second study ($M = 3.8$, $SD = 0.92$). All the more surprising is that the sense of time (21f) was very similar for all participants ($t = 0.23$, $p = .8174$). Both studies had outliers in all directions, but on average, study 1 ($M = 3.2$, $SD = 1.15$) and study 2 ($M = 3.1$, $p = 0.88$) achieved strikingly similar results with approximately normal time perception. In summary, the results of the questionnaire are almost identical, which means in particular that this type of hands-free interaction offers great entertainment value regardless of physical limitations.

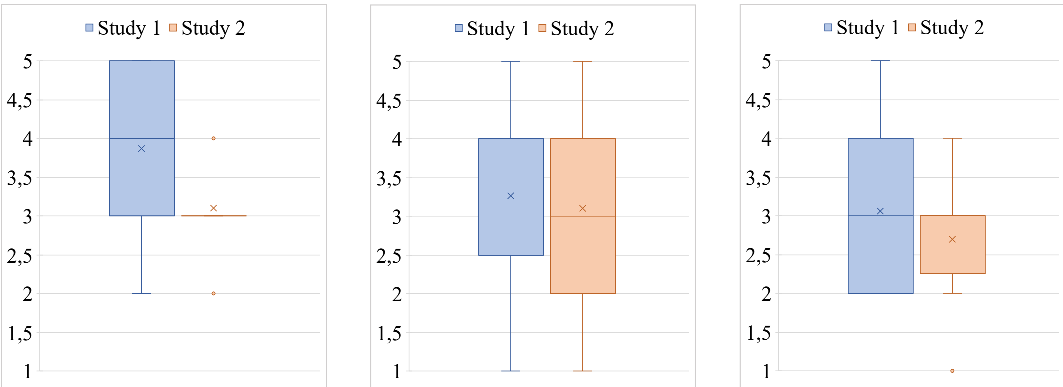


(a) Skill (b) Enjoyment (c) Challenge

(d) Excitement (e) Focus (f) Sense of time

Figure 21: Study 2: Questionnaire result on experience

Figure 22 shows the results of the control section. Again, the answers from the first study with *GV* are given as a reference. The precision of the control (22a) was rated worse ($t = 2.21$, $p = .0375$) by the participants of the second study ($M = 3.1$, $SD = 0.57$), which also corresponds to our observation. The lower score can be attributed to several factors based on individual impairments. Examples are squinting and twitching of the eyes. The results of the first study ($M = 3.87$, $SD = 0.99$) show the potential of hands-free control and thus that eye-tracking and non-verbal voice commands are not suitable as a supportive means of computer interaction for every type of disability. All participants described the effort required by the game (22b) in roughly the same way ($t = 0.32$, $p = .7466$). In figures, the first study scored 3.27 ($SD = 1.1$) and the second slightly less, namely 3.1 ($SD = 1.45$). This shows that despite the higher physical demands, even less trained people do not have major problems with continuous voice control. This becomes particularly important when it is no longer possible to breathe independently and ventilators are used. These can make humming long tones much more difficult. In the last question concerning pace, *S2* ($M = 2.7$, $SD = 0.83$) again scored minimally worse ($t = 0.90$, $p = .3788$) than *S1* ($M = 3.07$, $SD = 1.1$).



(a) Precision (b) Exhaustion (c) Pace

Figure 22: Study 2: Questionnaire result on control

6. Discussion

The results show a better performance and higher accuracy for *MK* compared to *GV*, whereby the latter can be considered quite competitive, considering that the participants are much more familiar with *MK* and never used eye tracking or non-verbal voice commands before. That first-time users of hands-free computer interaction can achieve comparable results shows the potential that can be realised with more practice. It is unlikely, however, that this will result in exactly the same performance as with *MK*, since a keystroke both gives better feedback and can be performed faster and with less effort. A combination of eye tracking and a physical device is both promising and realistic, as many people with motor disabilities still have some mobility in their extremities. An example of such a device would be a simple button for the hand or pedals for the feet.

In the second study with people with disabilities, difficulties repeatedly arose when using the eye tracker. For example, depending on the type of eye twitches, the accuracy was significantly reduced. The effects of the disabilities must therefore be taken into account, and by no means everyone in the world can benefit from this control system. That is why products for people with motor impairments are very individualised, as a generalisation is difficult. However, it was also found that participants with severe paralysis coped very well with gaze control.

Many participants had initially tried, both consciously and unconsciously, to control the penguin with head movements. It took them a while to focus on the gaze point rather than the direction of the head. The eye tracker from Tobii supports the recognition of the head position and orientation, which is why it would be possible to implement this type of control. For players without cramps or muscle twitches, this could be an intuitive alternative to eye tracking. However, it must be taken into account that more complex control is not possible, since only four directions (up, left, down, right) can be indicated.

The participants described the most difficult challenge as not being able to look at the penguin, as otherwise, it would not move any further. The gaze trace function (part of the Tobii software, figure 14 on page 18), which displays a circle representing the rough point of view, made it easier for many to control the eye tracker as it gave them a better feel for it. In addition, we observed that the fish, which were specifically placed as a focus point, made it easier to move around because the player could focus on the fish and not on the need to look at the penguin. For the same reason, the camera (the scene displayed to the player) movement has been adjusted after some testing. In the beginning, the camera moved exactly with the penguin, as it is usual in jump-and-run games. However, this brought the difficulty that if one looked at a certain point and then jumped, the point of view no longer corresponded to the originally aimed point and thus tended to jump too far. This was solved by having the camera stay almost still while the penguin moves and only catch up when the penguin has landed. This makes it possible to focus on a spot, jump and land precisely on the focused spot. Another advantage is that the player is not distracted by the constantly moving game world.

No sound effects or music were used in this game to create the best possible conditions for the microphone. With the help of personalised voice recognition, they could be added to the game. Using headphones is an easier solution, but for some people this is not an option due to special headrests. In the scenario of a multiplayer game over the internet, a special hum recognition would also be advantageous, as players could communicate with each other without unintentionally giving instructions to the game.

Overall, two-third of the participants in the first study preferred the hands-free version and described it as much more fun. In the study with people with motor impairments, on the other hand, all found the game very enjoyable and could well imagine this type of control in the future. It was surprising how little eye tracking technology is known among this target group. It could make simple tasks on the computer much more accessible in their daily life.

7. Conclusion

This work presents a hands-free video game whose control system is based on eye tracking and non-verbal voice commands. It is the first time that these two modalities have been combined into a complete video game control system that has been compared with conventional interaction techniques, but also the feasibility for the target group has been examined more closely. Promising results were noted, including a competitive performance of the hands-free version and a much higher fun factor. However, there were difficulties for some participants with disabilities, and a few could not finish a game run, which was due to the many different effects of the disabilities. The use of eye trackers thus sets the frame for the user group, as full operation is not possible in the case of intensive visual impairment. It should be mentioned that the game shown is straightforward, so one cannot conclude feasibility for all modern video games that are far more complex in control. This is a topic for future work to integrate these controls into normal video games.

After all, we look at the results obtained together with the hypotheses made at the beginning. In the first category about game control, we can conclude that Gaze is a reliable input method and can keep up with the performance of the mouse. Making non-verbal sounds (humming) is a viable method for temporal interaction with the computer. It takes more effort and is not quite as precise as a keystroke, but it enhances the fun of the game. Together, gaze and hum confirm themselves as a well-functioning multi-modal input system for spatio-temporal control. The second category regarding game experience shows that gaze and hum as a natural input offer a more immersive and exciting experience than physical device control. That being said, the flow and pace of the game is slower with hands-free controls, but similar persuasive as with mouse and keyboard. It is also more sophisticated compared to conventional peripherals. The last category answers important questions about the feasibility for people with disabilities. The whole hands-free control system would be of no use if the target group does not find an application for it. But as it turns out, gaze and hum prove to be a reliable input methodology for people with disabilities, as long as the effects of the disabilities allow operation in general. Severe disturbances of the eye and head movement can considerably limit the functionality of the eye tracker. Overall, playing with eye tracking and voice provides engaging and entertaining game experiences for people with motor impairments. The performance turned out to be lower compared to the players without physical limitations but is still in the range of competition.

In conclusion, computer interaction in games based on gaze and hum has potential for further projects and can already offer high value for people with motor impairments. Simple mini-games are easy to develop for the presented control methodology and can thus offer high entertainment value to many people.

References

[1] Javier Agustin, Julio Mateo, John Hansen, and Arantxa Villanueva. Evaluation of the potential of gaze input for game interaction. *PsychNology Journal*, 7:213–236, 01 2009.

[2] Jeanne H Brockmyer, Christine M Fox, Kathleen A Curtiss, Evan McBroom, Kimberly M Burkhart, and Jacquelyn N Pidruzny. The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of experimental social psychology*, 45(4):624–634, 2009.

[3] Emiliano Castellina and Fulvio Corno. Multimodal gaze interaction in 3d virtual environments. *Cogain*, 2008.

[4] The Able Gamers Charity. *Combating Social Isolation Through Play*, 2004. URL `https://ablegamers.org`. Last visited: 22.06.2021.

[5] *Stats about paralysis*. Christopher and Dana Reeve Foundation, 2013. URL `https://www.christopherreeve.org`. Last visited: 23.06.2021.

[6] Muriel Daniel and Crawford Garry. *Video games as culture: considering the role and importance of video games in contemporary society*. Routledge, 2018.

[7] Bayat Games. *Free Platform Game Assets*, 2017. URL `https://assetstore.unity.com`. Last visited: 26.06.2021.

[8] Massimo Gneo, Maurizio Schmid, Silvia Conforto, and Tommaso D'Allessio. A free geometry model-independent neural eye-gaze tracking system. *Journal of neuroengineering and rehabilitation*, 11 2012. doi: 10.1186/1743-0003-9-82.

[9] Susumu Harada, Jacob O. Wobbrock, and James A. Landay. Voice games: Investigation into the use of non-speech voice input for making computer games more accessible. In Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2011*, pages 11–29, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-23774-4.

[10] Johan Högberg, Juho Hamari, and Erik Wästlund. Gameful experience questionnaire: Measuring the gamefulness of service use. In *Frontiers in Service 2019, National University of Singapore, 18-21 Juli 2019.*, 2019.

[11] Takeo Igarashi and John F Hughes. Voice as sound: using non-verbal voice input for interactive control. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 155–156, 2001. doi: 10.1145/502348.502372.

[12] Wijnand A IJsselsteijn, Yvonne AW de Kort, and Karolien Poels. The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven*, pages 3–9, 2013.

[13] Poika Isokoski and Benot Martin. Eye tracker input in first person shooter games. In *Proceedings of the 2nd Conference on Communication by Gaze Interaction: Communication by Gaze Interaction-COGAIN 2006: Gazing into the Future*, 2006.

[14] Erika Jönsson. If looks could kill–an evaluation of eye tracking in computer games. *Unpublished Master's Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden*, 2005.

[15] Kechapp. *Scream Go Hero*, 2020. URL `http://www.ketchappgames.com`. Last visited: 22.06.2021.

[16] Laurens R Krol, Sarah-Christin Freytag, and Thorsten O Zander. Meyendtris: A hands-free, multimodal tetris clone using eye tracking and passive bci for intuitive neuroadaptive gaming. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 433–437, 2017.

[17] Andrew Kurauchi, Wenxin Feng, Ajjen Joshi, Carlos Morimoto, and Margrit Betke. *EyeSwipe: Dwell-Free Text Entry Using Gaze Paths*, page 1952–1956. Association for Computing Machinery, New York, NY, USA, 2016. ISBN 9781450333627. doi: 10.1145/2858036.2858335.

[18] Marta Lopes, João Magalhães, and Sofia Cavaco. A voice-controlled serious game for the sustained vowel exercise. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, ACE '16, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450347730.

[19] Anders Møller Nielsen, Anders Lerchedahl Petersen, and John Paulin Hansen. Gaming with gaze and losing with a smile. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, page 365–368, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450312219.

[20] J O'Donovan, J Ward, S Hodgins, and V Sundstedt. Rabbit run: Gaze and voice based game interaction. In *Eurographics Ireland Workshop, December*, 2009.

[21] *Assistive Technology & Industrial / Commercial Trackballs & Joysticks*. Pretorian Technologies Ltd, 2021. URL `https://www.pretorianuk.com`. Last visited: 23.06.2021.

[22] *A game controller for quadriplegics*. Quadstick, 2014. URL `https://www.quadstick.com`. Last visited: 23.06.2021.

[23] Adam J Sporka, Sri H Kurniawan, Murni Mahmud, and Pavel Slavík. Non-speech input and speech recognition for real-time control of computer games. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, pages 213–220, 2006.

[24] *Enhance PC games with eye tracking*. Tobii Gaming, 2021. URL `https://gaming.tobii.com/games`. Last visited: 23.06.2021.

[25] *Unity SDK*. Tobii Technology, 2021. URL `https://developer.tobii.com/pc-gaming/unity-sdk`. Last visited: 07.04.2021.

[26] *Guidelines on the provision of manual wheelchairs in less resourced settings*. WHO, 2008. URL `https://www.who.int`. Last visited: 23.06.2021.

[27] *Disability and health*. WHO, 2020. URL `https://www.who.int`. Last visited: 23.06.2021.

[28] Tom Wilcox, Mike Evans, Chris Pearce, Nick Pollard, and Veronica Sundstedt. Gaze and voice based game interaction: the revenge of the killer penguins. *SIGGRAPH Posters*, 2008.

[29] Disabled World. *Foot Mouse with Programmable Pedal*, 2017. URL `https://www.disabled-world.com/assistivedevices/computer`. Last visited: 01.09.2021.

[30] XiaoJie Yuan and Jing Fan. Design and implementation of voice controlled tetris game based on microsoft sdk. In *2011 International Conference on Multimedia Technology*, pages 275–278, 2011.

[31] Arthur Zuckerman. *50 Video Game Statistics: 2020/2021 Industry Overview, Demographics & Data Analysis*. CompareCamp, 2020. URL `https://comparecamp.com/video-game-statistics`. Last visited: 23.06.2021.

List of Figures

List of Tables

A. Questionnaire

Question	Scale
How old are you?	Number
Are you wearing glasses or contact lenses right now?	Glasses, contact lenses, None of them
What is your gender?	Female, Male, Diverse
What is your current occupation?	Pupil, Student, Employed, . . .
Do you have experience with video games (mouse and keyboard) ?	1 (Never really played) - 5 (I play very often)
Do you have experience with Eye Tracking?	Never used, used a few times, use it often.
Do you have experience with voice-controlled games?	Never used, used a few times, use it often
How well do you think you played?	1 (Not good at all) - 5 (Very good)
How much did you enjoy playing?	1 (Very boring) - 5 (Very entertaining)
Was it challenging to play?	1 (Not at all) - 5 (Very challenging)
Was it exciting to play?	1 (Not at all) - 5 (Very exciting)
Did you focus deeply on the game?	1 (Not at all) - 5 (Very concentrated)
Have you lost track of time? (Time went by . . .)	1 (Very slowly) - 5 (Very quickly)
Did you get out of breath?	Not at all, sometimes, often
Did the penguin move exactly the way you wanted it to?	1 (Very inaccurate) - 5 (Very accurate)
How would you rate Mouse and Eye Tracking overall?	1 (Very bad) - 5 (Very good)
How would you rate Keyboard and Voice overall?	1 (Very bad) - 5 (Very good)
Was it exhausting to play?	1 (Not at all) - 5 (Very exhausting)
How fast did you play?	1 (Slow) - 5 (Fast)

Table 3: Questionnaire

B. Detailed Data Analysis

Measure	MK avg(1,2)	GV avg(1,2)	MK vs GV
Time to reach finish line (ms)	M=91592.57 SD=40529.55	M=163441.77 SD=56109.68	t=-4.0203 p<.001
Proportion of motion in total time	M=0.65 SD=0.1	M=0.49 SD=0	t=5.2131 p<.001
Time between movements (ms)	M=460.54 SD=207.22	M=944.03 SD=201.31	t=-6.4817 p<.001
Number of fatal movements	M=2.03 SD=1.86	M=3.97 SD=2.52	t=-2.3939 p=.0236
Walking percentage	M=0.21 SD=0.1	M=0.25 SD=0.1	t=-1.1668 p=.2531
Duration of temporal action (ms)	M=649.36 SD=81.70	M=724.41 SD=108.06	t=-2.1455 p=.0407
Distance of a movement (blocks)	M=2.25 SD=0.22	M=1.97 SD=0.24	t=3.2943 p=.0027

Table 4: Study 1: *MK* compared to *GV*. Gray fields indicate significant difference by two-tailed t-test

Measure	MK (1)	MK (2)	MK (1 vs 2)
Time to reach finish line (ms)	M=100980 SD=56682.88	M=82205.13 SD=26389.24	t=1.1630 p=.2547
Proportion of motion in total time	M=0.63 SD=0.1	M=0.67 SD=0.1	t=-1.0141 p=.3192
Time between movements (ms)	M=494.53 SD=219.64	M=426.54 SD=198.04	t=0.8904 p=.3808
Number of fatal movements	M=2.2 SD=2.91	M=1.87 SD=1.64	t=0.3866 p=.7020
Walking percentage	M=0.21 SD=0.1	M=0.2 SD=0.1	t=0.1648 p=.8703
Duration of temporal action (ms)	M=637.45 SD=86.93	M=661.27 SD=93.22	t=-0.7238 p=.4752
Distance of a movement (blocks)	M=2.19 SD=0.25	M=2.31 SD=0.26	t=-1.204 p=.2388

Table 5: Study 1: Both runs of *MK* compared with each other. Gray fields indicate significant difference by two-tailed t-test

Measure	GV (1)	GV (2)	GV (1 vs 2)
Time to reach finish line (ms)	M=180398.53 SD=65378.68	M=138012.5 SD=43048.72	t=2.0458 p=.0506
Proportion of motion in total time	M=0.48 SD=0.2	M=0.52 SD=0.17	t=-2.0865 p=.0465
Time between movements (ms)	M=991.6 SD=205.74	M=850.38 SD=143.87	t=2.1272 p=.0427
Number of fatal movements	M=4.8 SD=2.83	M=3 SD=2.66	t=1.7605 p=.0897
Walking percentage	M=0.25 SD=0.1	M=0.23 SD=0.1	t=0.4283 p=.6718
Duration of temporal action (ms)	M=716.67 SD=111.33	M=742.66 SD=116.63	t=-0.6141 p=.5443
Distance of a movement (blocks)	M=1.93 SD=0.32	M=2.05 SD=0.2	t=-1.23 p=.2284

Table 6: Study 1: Both runs of *GV* compared with each other. Gray fields indicate significant difference by two-tailed t-test

Measure	GV first (1)	GV first (2)	GV first (1 vs 2)
Time to reach finish line (ms)	M=194047.86 SD=67956.98	M=136896 SD=34340.23	t=1.8586 p=.0900
Proportion of motion in total time	M=0.46 SD=0	M=0.52 SD=0	t=-2.1698 p=.0528
Time between movements (ms)	M=1125.47 SD=196.00	M=850.93 SD=161.53	t=2.7241 p=.0198
Number of fatal movements	M=4.71 SD=2.29	M=3 SD=2.45	t=1.3039 p=.2189
Walking percentage	M=0.19 SD=0.1	M=0.18 SD=0	t=1633 p=.8733
Duration of temporal action (ms)	M=749.02 SD=108.13	M=763.12 SD=115.91	t=-0.2268 p=.8248
Distance of a movement (blocks)	M=1.91 SD=0.26	M=2.05 SD=0.17	t=-1.0964 p=.2963

Table 7: Study 1: Both runs of *GV* compared with each other of those who played it before *MK*. Gray fields indicate significant difference by two-tailed t-test

Measure	Study 2 (1)	Study 2 (2)	Study 2 1 vs 2
Time to reach finish line (ms)	M=420848.4 SD=591212.28	M=287428.2 SD=354884.49	t=1.37 p=.2085
Proportion of motion in total time	M=0.39 SD=0.1	M=0.38 SD=0.1	t=0.35 p=.7392
Time between movements (ms)	M=1588 SD=396.78	M=1660 SD=536	t=-0.24 p=.8174
Number of fatal movements	M=7.8 SD=2.49	M=5.6 SD=4.88	t=0.90 p=.3953
Walking percentage	M=0.19 SD=0.14	M=0.17 SD=0.1	t=0.34 p=.7443
Duration of temporal action (ms)	M=778 SD=228.64	M=732 SD=199.46	t=0.34 p=.7448
Distance of a movement (blocks)	M=1.73 SD=0.1	M=1.88 SD=0.17	t=-1.54 p=.1624

Table 8: Study 2: Both runs compared with each other. Gray fields indicate significant difference by two-tailed t-test

Measure	Study 1 *GV* first avg(1,2)	Study 2 avg(1,2)	Study 1 vs Study 2
Time to reach finish line (ms)	M=174629.36 SD=59127.92	M=354138.3 SD=141041.87	t=-3.0573 p=.0121
Proportion of motion in total time	M=0.48 SD=0	M=0.38 SD=0.1	t=2.1779 p=.0544
Time between movements (ms)	M=1037.52 SD=241.70	M=1623.97 SD=421.05	t=-3.0622 p=.012
Number of fatal movements	M=4 SD=2.04	M=6.7 SD=3.62	t=-1.6585 p=.1283
Walking percentage	M=0.19 SD=0.1	M=0.18 SD=0.1	t=0.2882 p=.7791
Duration of temporal action (ms)	M=743.35 SD=110.54	M=754.86 SD=213.61	t=-0.1228 p=.9047
Distance of a movement (blocks)	M=1.95 SD=0.25	M=1.8 SD=0.14	t=1.2274 p=.2478

Table 9: Both runs of study 1 (*GV* first) and 2 compared with each other. Gray fields indicate significant difference two-tailed by two-tailed t-test

Erklärung

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Druck-Exemplaren überein.

Stuttgart, 14.9.2021

(Ort, Datum) (Unterschrift)

Declaration

I hereby declare that the work presented in this thesis is entirely my own. I did not use any other sources and references that the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted hard copies.

Stuttgart, 14.9.2021

(Place, Date) (Signature)