Back to Contents

The Golden Bullet: A Comparative Study for Target Acquisition, Pointing and Shooting

Katharina Anna Maria Heydn,
Marc Philipp Dietrich,
Marcus Barkowsky,
Götz Winterfeldt
Deggendorf Institute of Technology
University of Applied Sciences
Deggendorf, Germany
katharina.heydn@stud.th-deg.de,
firstname.lastname@th-deg.de

Sebastian von Mammen
Games Engineering
Julius-Maximilians University of Würzburg
University of Würzburg
Würzburg, Germany
sebastian.von.mammen@uni-wuerzburg.de

Andreas Nüchter
Robotics and Telematics
University of Würzburg
Julius-Maximilians University of Würzburg
Würzburg, Germany
andreas.nuechter@uni-wuerzburg.de

***Abstract*—In this study, we evaluate an interaction sequence performed by six modalities consisting of desktop-based (DB) and virtual reality (VR) environments using different input devices. For the given study, we implemented a vertical prototype of a first person shooter (FPS) game scenario, focusing on the genre-defining point-and-shoot mechanic. We introduce measures to evaluate the success of the according interaction sequence (times for target acquisition, pointing, shooting, overall net time, and number of shots) and conduct experiments to record and compare the users' performances. We show that interacting using head-tracking for landscape-rotation is performing similarly to the input of a screen-centered mouse and also yielded shortest times in target acquisition and pointing. Although using head-tracking for target acquisition and pointing was most efficient, subjects rated the modality using head-tracking for target acquisition and a 3DOF Controller for pointing best. Eye-tracking (ET) yields promising results, but calibration issues need to be resolved to enhance reliability and overall user experience.**

***Index Terms*—Human-Computer Interaction, Head-Mounted Device, Modality, Ray Casting, Eye-Tracking, Virtual Reality, User-Centered Design**

I. Introduction

Human-computer interaction (HCI) plays a fundamental part in today's high-tech world. Interaction techniques are defined by the use of physical input devices to perform certain tasks in the human-computer dialogue [1]. With the introduction of VR-technologies, HCI has been expanded by a third dimension. A head-mounted Display (HMD) is often used as a visualization device in a VR user interface, it detaches the user from the surroundings by isolating the user's field of vision. However, using HMDs, keyboard and mouse are no intuitively usable input devices for interaction. The observer expects the HMD system to respond in a natural way to head and body movements, i.e. displaying corresponding scenery content with minimum delay. Therefore, there is a need for new input modalities.

In the past, various modalities have already been investigated and compared with each other in terms of precision, accuracy and speed. But these investigations were limited to either the two-dimensional [2], [3] or the three-dimensional space only [4]–[8].

We want to investigate how these VR modalities behave compared to the conventional modalities in performing the same task. But we don't want to look at only one specific interaction like behaviour of modalities for navigation or selection. Rather, we want to observe behavior in a sequence of an interaction tasks.

Therefore we perform the same interaction sequence in a DB as well as in a VR user interface. The task is to find moving objects, to target them and to shoot them. These steps (target acquisition, pointing and shooting) may be considered as main components of a first-person-shooter (FPS) gaming scenario. The consecutive steps are partially relying on different modalities, including 3D controller, gyroscope, mouse and keyboard input as well as monoscopic DB and stereoscopic, tracked VR output.

In the DB user interface the scenario is shown on a standard flat screen display, input is done with keyboard and mouse. The VR user interface uses a HMD, a gyroscope and a handheld 3D controller as input. In parallel to the gyroscope, controller, mouse and keyboard, new concepts such as ET are also being investigated [1], [9]–[15]. Hence we include ET in both user interfaces for the pointing and shooting step.

The effectiveness and efficiency of the modalities is evaluated by measuring the time for target acquisition, pointing and shooting as well as the net gaming time.

The study revealed shorter gaming times for the VR user interface when compairing all their modalities to all modalities in the DB user interface. Also the VR modalities showed more efficiency in target acquisition and pointing, only shooting took longer. The modality using head-tracking for target acquisition, pointing and shooting yield best efficiency because of shortest times in pointing and target acquisition. Modalities using ET showed worst efficiency, most likely due to calibration problems. Their pointing times were the longest and had the highest variances, also using ET influenced the target acquisition in the VR user interface. Using a screen centered mouse for target acquisition, pointing and shooting was the best modality in DB user interface. Target acquisition was as fast as obtained with the head-tracker, but pointing time was slower. Modalities using the keyboard showed slower results for target acquisition.

The paper is organized as follows: First in Sec. II a brief overview is given on research concerning interaction modalities. Then we present our goals and methods in Sec. III and describe which combination of input devices and visualization output is used by each modality. We describe the testing scenario, its task and implementation as well as the participants, testing conditions and data recording in Sec. IV. Afterwards the results of the testing scenario and analysis of the recorded data are shown in Sec. V. Finally we take a closer look on the results in Sec. VI and give a summary of them and an outlook to future work in Sec. VII.

II. State of the Art

A. Interaction Modalities and Techniques

There are many approaches to comparing interaction techniques in a virtual environment. In [8] Bowman et al. compare gaze-directed steering (input with head orientation) and hand-directed steering (input with hand orientation) for first-person motion control in an immersive virtual environment. Their tasks include travelling to an explicit target object and travelling to a point relative to a reference

978-1-7281-4540-2/19/$31.00 ©2019 IEEE

object, measuring the time required for rating the efficiency of each interaction technique.

Travelling techniques belong to the navigation techniques, it also includes the rotational change of the viewport, which is usually achieved by the movement of the head. Considering this Ragan et al. investigate the effects of accelerated head-tracking [16], which corresponds to a semi-natural physical view control. It enables a 360^o viewpoint with even small head movements. Motion sickness, however, occurs more frequently using that interaction technique.

Bowman et al. extend their investigation of travel techniques by techniques for object selection and manipulation in [7]. In two testbeds, one for selection and manipulation, the other for travelling, they compared different VR interaction techniques with each other. For the selection and manipulation tasks, input was done with a three-button joystick and a tracking device. For a fast selection of remote objects, ray casting technique was recommended.

Lee et al. use also ray casting as an input method and examined four different interaction techniques based on this input method [6]: i) the direct image plane selection (2D virtual mouse as input device), ii) head-directed pointer, iii) hand-directed pointer and iv) head-hand-directed pointer. The time required to perform a selection and dragging task was measured. In this study ray-casting is not used in combination with a HMD, but with a projection display. Using the 2D virtual mouse and hand directed selection turned out to be the fastest methods.

Teather and Stuerzlinger also examinate ray-casting in their comparison of interaction techniques [2]. They compare mouse-based and remote-based pointing techniques and incorporate stereo and mono-rendering of the cursor in their considerations.

Kopper et al. noticed issues with hand and tracker jitter while using the ray-casting technique for selection. So for the special case of selecting small, remote objects they propose their progressive refinement method and compared it to ray-casting technique [12]. It works with sphere-casting refined by a QUAD-menu. In contrast to the ray casting method, the effects of hand jitter do not have any negative effects with this method.

B. Eye-Tracking

ET is well suited for selection tasks, as Kaufman et al. have already proven in their early work [9].

Regarding performing a selection process, Sibert et al. used the combination of ET and Dwell Time. The times achieved were much shorter than using a computer mouse in combination with a mouse click [12].

Vertegaal describes similar results [15]. Selecting with ET and a separate click trigger was slower than with ET and Dwell Time. But it provides the best ratio between speed and accuracy compared to a mouse input. This method was also rated the most user-friendly by the participants [13].

In [14] Kasprowski and Niezabitowoski replaced the mouse input with the ET input in a shooting game context. They measured higher values in the aiming time due to the lack of accuracy of the ET method.

Chun et al. [5] also included ET as a method for object manipulation in their comparison besides the classical devices such as mouse and keyboard and extended this additionally with the input via a brain-computer-interface (BCI) [5]. ET yield five-times faster performance time for selection task than using the mouse.

Qian and Teather [17] compared the selection times of head-based selection with those of an eye-based selection in a VR environment. They are using the FOVE, a HMD with integrated ET, for combining VR and ET. It was shown that the Head-Based Selection had more advantages regarding the selection time.

III. Research Question

The purpose of this research is an overall evaluation of an interaction sequence performed with six different modalities. We compare the effectiveness and efficiency of the modalities by measuring the needed times for target acquisition, pointing and shooting at moving objects, which present the steps of the interaction sequence. For the modalities, two visualization devices are used, a conventional display in the DB and a HMD in the VR user interface. In total five different input devices are used: keyboard and mouse in the DB user interface, a handheld-controller and head-position-tracker in the VR user interface, and an eye-tracker in both user interfaces. Different combinations of the input and visualization devices results in total of six modalities (s.f. Sec. IV-B).

Within the scenario the subject stays fixed at one location. He can rotate his viewpoint to the left and right, as well as to the top and bottom, performing the rotational movements yaw and roll. The subject is also able to point a black colored cursor at the actual sphere and fire over a trigger. Depending on the used modality, the subject has one or two devices for target aquisition and pointing at the same time. A trigger button for shooting is always available. The subject has unlimited shots at any time. Further details are given in Sec. IV-A.

The test scenario distinguishes between a navigation process and an interaction process according to Bowman et al. [7]. The navigation process is limited to rotational movement resulting in landscape-rotation and describes the target acquisition, the interaction process includes pointing and shooting.

In total three steps S1 to S3 are distinguished in the task described in Sec. IV-A. We now introduce the different steps, namely target acquisition, pointing, shooting, as follows:

S1: target acquisition (type: navigation) which is the rotational movement of the users viewpoint. This requires an interaction with at least 1 degree of freedom (DOF).

S2: pointing (type: interaction) requires pointing with a cursor at a moving object. This requires an interaction with at least 2DOF.

S3: shooting (type: interaction) consists of triggering shots while remaining targeted at a moving object. At least 2DOF and a trigger interaction is required.

Keyboard and a mouse are used as input devices in the DB user interface. In order to emphasize the perception point of view, it is worth stating that the mouse requires translational movements of the hand while interaction with a keyboard is done by pressing down a key with a finger. Using the mouse is a continuous input technique which allows 2DOF. Using the keyboard is a discrete input technique triggered by pressing the keys W, A, S, D with the finger, two keys are used in conjunction for one coordinate axis (A/D and W/S), thus allowing for a fixed velocity interaction in 2DOF. For the trigger interaction, the mouse button or the space bar on the keyboard is frequently used.

In particular in the VR user interface, a head-tracker is often used for navigation [8]. This is usually implemented using a gyroscope or external tracking sensors. In this study gyroscope-based head tracking instead of outside-in tracking was used, because we valued typical input devices over high precision of head tracking. This allows for continuous 3DOF in rotational coordinates (yaw, pitch, roll) by rotational head movement. For input in the VR user interface, a special handheld controller is often supplied [8]. It enables continuous

Back to Contents

tracking of rotational and translational hand movements resulting in 6DOF. In addition, it features a trigger button.

ET may be used as input device in both, the DB user interface and the VR user interface. The rotational movement of the eyes is usually interpreted as pointing on a 2D plane, i.e. the gaze point, thus resulting in 2DOF. Blinking may be used as trigger but following Jacob´s line of argument [1], it was not used in this study. According to Jacob an input using blinking is rather awkward, because it represents an unnatural input for the user and the attention is focused on an actually subconscious process. He therefore recommends dwell time or pressing a key for confirming interaction. We are using a button trigger, to reach similar trigger times compared to the other modalities. E.g. Meena and Wong-Li also use a soft switch in combination with ET to interact with their multimodal user interface for controlling a wheelchair [11].

The participants performed six trials, one for each of the six modalities. Three modalities use the DB user interface, three use the VR user interface. Tab.I shows the human body part, its type of movement and the number of degrees of freedom that has been used in the three steps S1 to S3 with the six different modalities M1 to M6.

TABLE I
MODALITIES (M) SPECIFYING THE DISPLAY TYPE AND FOR EACH STEP (S) THE BODY PART, THE TYPE OF ACTION AND THE NUMBER OF DEGREES OF FREEDOM USED (ABBREVIATIONS: DOM. (DOMINANT), L. (LEFT) AND R. (RIGHT))

	Display	S1	S2	S3
M1	DB	l.hand key press discrete 2DOF	r.hand translation continuous 2DOF	r.hand press button discrete
M2	VR	head rotation continuous 2DOF	dom.hand rotation continuous 2DOF	dom.hand press button discrete
M3	DB	r.hand translation continuous 2DOF	r.hand translation continuous 2DOF	l.hand press button discrete
M4	VR	head rotation continuous 2DOF	head rotation continuous 2DOF	dom.hand press button discrete
M5	DB	r.hand key press discrete 2DOF	eyes rotation continuous 2DOF	l.hand press button discrete
M6	VR	head rotation continuous 2DOF	eyes rotation continuous 2DOF	dom.hand press button discrete

The structure of these six modalities M1 to M6 is based upon the following:

- M1 and M2 use a dedicated device for target acquisition task and another dedicated device for the pointing task.
- M3 and M4 use one single device each for both, the target acquisition and pointing task.
- M5 and M6 use ET for the pointing task.

The setup allows to compare the performance of the DB user interface (M1, M3, M5) with the performance of the VR user interface (M2, M4, M6).

The purpose is to compare the efficiency of the modalities. Efficiency is measured as task completion time like it was done in [2], [6]–[8], [17].

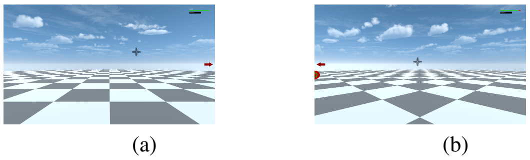
(a) (b)

Fig. 1. Arrow on the right side indicates the shortest distance to the sphere (a) and will disappear, when at least half the sphere is visible in the field of view (b).

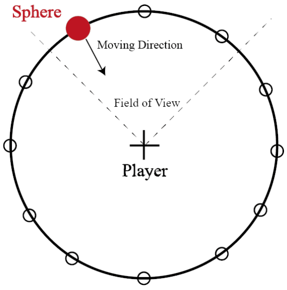

Fig. 2. Scenario set up with marked spawn position of the spheres. The subject is positioned in the center and may rotate, the field of view is marked with dashed lines.

IV. Evaluation Setup

A. Scenario Design and Task

A typical FPS scenario consists of a player who moves around the virtual world, searches for a target, aims and shoots at the target. Similar to [14] we have implemented a moving target, but in a 3D environment, not a 2D one. The authors of [5] use a 3D environment shown on a HMD, but static targets.

The subject finds himself in the center of a 3D world (Fig. 1).

Translational movement of the subject is prohibited, but rotational movement is allowed. According to the instructions for the subject, he searches for a sphere by rotating along the z-axis (yaw). He targets the sphere with a crosshair and shoots bullets at it.

The task of the subject is to find twelve spheres and hit each of them three times. The spheres appear one after the other and they move with constant speed towards the subject. Instead of vertical or horizontal movement this kind of movement was chosen, because it may simulate an approaching enemy in a fps game and also compensates for the lack of translational movement of the player. The first sphere appears two seconds after the start of the scenario. Once it is destroyed the next sphere appears after a delay of two seconds. The scenario is over when the subject has eliminated all spheres. There are twelve different spawning positions, at equal distance to the origin (Fig. 2). The angle distances between the positions are intervals of 30^o (0^o, $\pm30^o$, $\pm60^o$, $\pm90^o$, $\pm120^o$, $\pm150^o$, 180^o). For each new spawn position, the direction of the player is calculated and then the sphere spawns in one of the twelve spawning positions relative to the players gaze direction. The twelve trials also use each spawning position exactly once and in random order to avoid any learning effect. This assures that the sum of all required rotations in S1, which is 1080^o, is identical for each trial.

Every time a sphere appears, an acoustic signal is given. If the sphere is outside the field of view, an arrow at the middle left (Fig. 1 (a)) or middle right edge of the screen additionally indicates the shortest distance (smallest angle) to the current sphere. This is

Back to Contents

intended to reduce fluctuations in the search time that would result from rotation to the opposite direction (larger angle). The arrow disappears as soon as the sphere enters the field of view, that means when half of the sphere is visible (Fig. 1 (b)) .

The 3D world consists of a black and white checkerboard tiled floor on the x-y-plane and a fixed sky box with white clouds. According to pretests, both measures significantly facilitated rotational orientation. The spheres have all the same radius and are colored red for contrasting with the surrounding. So they have a signal effect to increase the subject's attention. A graphic in the form of a ring on the sphere indicates how often the sphere has been hit, per hit the color turns gradually from green to red. An additional graphic in the upper right corner shows the health of the subject and the number of already eliminated spheres. Health is only reduced, if a sphere couldn't be eliminated before it collides with the player (which only happened once in the test procedure). All graphics have been kept minimalistically in order to distract less from the task.

B. Input Devices

According to the steps described in Sec. III devices are required for controlling the rotation of the viewpoint in the target acquisition step, the movement of the crosshair in the pointing step and triggering a shot in the shooting step. Tab.II shows the devices used for each step and their input triggers. A keyboard is used in the target acquisition task (S1) in M1 and M5 using the keys W, A, S, D as input, a gyroscope in M2, M4 and M6 and a screen centered mouse in M3. The usage of the screen centered mouse in M3 and keyboard in M1 and M5 for target acquisiton at the DB modalities leads to unbalanced modalities in comparison to the usage of only one input device (head tracking) for target acquisiton in all three VR modalities. We accepted unbalanced modalities and did not use the keyboard in M3, but the centered mouse, because this interaction technique is commonly used in game based scenarios. The pointing task (S2) is done by a mouse in M1 (free movable on the screen) and M3 (fixed in the screen center), by a gyroscope in M4 and by a handheld controller using ray-casting as it is a easy and intuitive technique [6] in M2. Head-hand pointing or two hands pointing would have been input options, but we decided to use the rotations of a VR-controller to control ray-casting, because of its availability in many typical user interfaces. Most likely, jittering of the ray did not have a significant effect on precision, because the distance of the spawn positions of the spheres is in a comparably small distance, therefore the perceived radius of the sphere allows for easy targeting via ray-casting.

Unlike in [7] the controller provides hand tracking and trigger button in a single device.

M5 and M6 uses ET as input. M5 uses the pupil labs for ET, M6 the FOVE, as done in [17]. Shots are triggered by the mouse with a left mouse button in M1, by the keyboard pressing the space bar in M3 and M5 and by a hand held controller pressing a button in M2, M4 and M6.

C. Hardware and Software Implementation

DB modalities (M1, M3 and M5) have been performed on a desktop PC [1] and a 27" monitor (Apple CinemaDisplay, 2560x1440 pixels). Input devices have been a standard wired keyboard (Microsoft Wired 400) and a standard 3-button optical mouse (Microsoft Basic Optical v2.0). For ET (M 5) a portable Pupil Labs device with binocular eye trackers (infrared, 200MHz, tracking precision $< 1^o$) and a Worldview camera (640x480 pixels, 120Hz, 100^o FOV) have been used.

TABLE II
INPUT DEVICE WITH DETAILS PER MODALITY (M) AND STEP (S)

	Display	S1	S2	S3
M1	DB	keyboard (WASD)	mouse (free)	mouse (left button)
M2	VR	gyroscope	controller (ray-casting)	controller (button)
M3	DB	mouse (centered)	mouse (centered)	keyboard (space bar)
M4	VR	gyroscope	gyroscope	controller (button)
M5	DB	keyboard (WASD)	eye-tracking (pupil labs)	keyboard (space bar)
M6	VR	gyroscope	eye-tracking (FOVE)	controller (button)

VR modalities (M2, M4) have been performed on a Google Daydream View-Headset (first generation) and a Google Pixel 1 smart phone[2]. Input devices have been the integrated gyroscope of the smartphone and the additionally available Google Daydream Controller[3]. Due to the lack of one device, which meets all the desired requirements for the third VR-setup (M6) a gaming PC[4] in combination with VR goggles from FOVE[5] have been used. Input devices have been the FOVE orientation tracking IMU, the integrated FOVE ET system (binocular, infrared, 120MHz, tracking precision $< 1^o$) and a wireless game controller (SonyPlaystation, DualShock, Bluetooth). The PlayStation Controller had to be used instead of the Daydream controller because of incompatibility issues.

Both HMD, the Daydream View-Headset and the FOVE, lack support for ametropia correction and inter-pupillary distance (IPD) setting.

The test scenario was developed in Unity Game Engine (2017.3.1f1) and C# on the same hardware configuration as for the DB interaction environment. The Pupil Labs ET glasses were calibrated with the Pupil Labs software Pupil Capture using 5-point calibration mode. The FOVE ET calibration was done by the FOVE software.

D. Participants, Environmental Conditions and Procedure

A total of 16 subjects (14 male, aged 23-46, mean 28.4) participated in the study. All of them were in healthy physical and mental condition. None of the subjects had eye injuries or restricted field of vision. 11 subjects had normal vision, 5 needed vision correction. They wore contact lenses or glasses during the tests. Only in M5 and M6 no glasses could be worn because of the design of the ET device and the VR headset. 11 participants already had gaming experience. 13 had no VR experience, the others had only limited VR experience (once or twice). 13 participants were right-handed.

The test environment was chosen as a closed office room with daylight conditions because a comfortable gaming atmosphere was preferred to the controlled conditions of an international telecommunication union (ITU) conforming test setup. The DB modalities (M1, M3, M5) were performed sitting in a comfortable position on an adjustable swivel chair at a table. Desktop PC, monitor, keyboard and mouse were placed on the table. The distance from head to

[1] Apple MacMini, Intel Core i7-478U CPU, Intel Iris(TM) Graphics 5100 GPU, 16GB RAM, Windows 10 Education

[2] Qualcomm Snapdragon 821, 4GB LPDDR4, 5.5" AMOLED, 1080x1920 pixels

[3] wireless, one touch pad, two circular buttons, 9-axis IMU

[4] Intel Core i9-7900X CPU, NVIDIA GeForce RTX2080 GPU, 32 GB DDR4 2666MHz RAM

[5] WQHD OLED 2560x1440 pixels, 100^o FOV

monitor was 50-60cm. The VR modalities (M2, M4, M6) with HMD were experimented with the participants standing within an area of 2m x 2m. The corresponding wireless controller was held in the dominant hand. During the experiment, only one test participant and the experimenter interacted in the room.

First, the participant was informed about the purpose and procedure of the experiment. Then was asked about his personal data and health conditions. Personal data included age, gender, information if the subject is right or left handed, needs glasses, is an epileptic and has gaming or VR experience. After the introduction, each participant went through the conditions of the study from M1 to M6. There was no training session before each trial. Before each scenario starts, the experimenter set the according configuration for used modality and subject ID.

Afterwards it was ensured that hands and fingers were positioned on the input devices (keyboard, mouse, controller). This should prevent searching the keys while playing and allow for a faster response time. In M5 and M6 subjects first went through the calibration process, which lasted about 1 minute each. At the end of each trial, the subjects were interviewed about the quality of experience for this modality using a Likert scale. They were asked how intuitive target acquisition and pointing felt and whether they were able to keep orientation in the scenario. This questionnaire took about one to two minutes and also served as resting phase between testing of the various modalities. At the end of the study, the subjects were asked to rank the modalities with regard to the user experience. The total duration of the experiment was about 45 minutes per subject.

E. Data Recording and Pre-Processing

The net playing time as well as the target acquisition time (S1), pointing time (S2) and shooting time (S3) (Sec. III) per sphere were calculated as follows:

Playing time: Sum of all times when a sphere was present, visible and not visible. This excludes all delays of two seconds at the start and before the re-appearance of another sphere.

Target acquisition time (S1): Time from the appearance of the sphere to the time when the center of the sphere is visible, i.e. more than 50% of the sphere is inside the field of view. The sphere had to be at least 0.5 seconds in the field of view for the search time to end.

Pointing time (S2): Time duration that starts when search time ends until the time when the cursor is over the sphere and the first shot is triggered.

Shooting time (S3): Time duration from the end of S2 until the time when the sphere disappears. Please note: The sphere can either disappear because of three successful hits, or because the sphere has reached the subject's position.

V. Results and Analysis

The results were analyzed by an univariate analysis of variance (ANOVA) with repeated measurement. The experiment is a within-subjects design, the factor levels consist of the varying modalities. Per modality dependent variables are measured: target acquisition time (S1), pointing time (S2) and shooting time (S3) according to Sec. III. DB- and VR- modalities alternate in execution order to reduce a training effect within the user interface.

A. Playing Time

The average playing times of the DB modalities showed a statistically significant difference between measurements (Tab.III). There was a significant difference in gaming times between each DB modality.

The playing times of the VR modalities also differed significantly in their values (Tab.IV) . There was a significant difference in playing times between M6 and M4 as well as M2 and M4. Between M6 and M2 no significant difference could be observed.

The average playing time was significantly shorter in all groups for VR modalities (1,2 t(15) = -2.42, p<.05), (3,4 t(15) = -2.36, p<.05), (5,6 t(15) = -7.32, p<.001) (Fig. 3).

B. Target Acquisition Time

For the DB modalities, there was a significant effect on the target acquisition time in between the concerned modalities. (Tab. III). Using keyboard in combination with mouse (free) or ET required a much longer target acquisition time than mouse (centered) only. Keyboard in combination with ET shows a particularly high standard deviation and thus a high variation in performance.

Within the VR modalities, no significant influence on the average target acquisition time could be determined (Tab. IV). Using the head movement for target acquisition in all three cases took approximately the same time and the standard deviation is similar.

A comparison of target acquisition times between DB and VR modalities yields different results. M1 and M2 as well as M5 and M6 differ significantly in the average target acquisition times (1,2: t(15)=-11.98, p<.001) and (5,6: t(15)=-13.11,p<.001). There was no significant difference between M3 and M4 (3,4: t(15)=2.07, p=.056) (Fig. 4). That means target acquisition with head-tracking is as fast as target acquisition by movement of a mouse.

C. Pointing Time

A significant effect of the DB modalities on the pointing time was detected (Tab. III). A comparison between mouse (free) and mouse (fixed, centered) showed no significant difference. Both, screen centered mouse and free movable mouse requires adjusting the cursor, because the spheres are placed and moving along the environment's ground plane, while the point of view is placed two environment units above this ground plane. There were significant differences between using the mouse (free and centered) and ET as pointing device. The use of ET took longest and yielded the highest standard deviation.

Within the VR modalities there was a significant difference in pointing times (Tab. IV). ET and ray casting were not significantly different. There was a significant difference between the pointing times of ET and those of the head-tracker. There is also a significant difference between head-tracking and ray casting. Pointing with ET required the longest duration and yielded in the highest standard deviation, while head-tracking reached the shortest times, which is comparable to the results of [17].

The pointing times of DB and VR modalities (M1 and M2, M3 and M4, M5 and M6) differed significantly (1,2: t(15)=-3.06, p=.008; 3,4: t(15)=-4.22, p=.001; 5,6:t(15)=-6.94, p<.001)(Fig. 5).

VR modalities consistently had lower values in pointing times. For both DB and VR modalities, the times by ET were the highest and the standard deviation the largest which is similar to the results in [14].

D. Shooting Time

No significant effect of the DB modalities on the shooting time could be observed (Tab. III).

A significant influence of the VR modalities on the average shooting times could be determined (Tab. IV). There was a significant difference between M2 and M4. M4 and M6 also differ significantly. M2 and M6 showed no significant difference in shooting time.

In the paired comparison of the DB and VR modalities (M1 vs. M2, M3 vs. M4, M5 vs. M6), the shooting times of the DB

Back to Contents

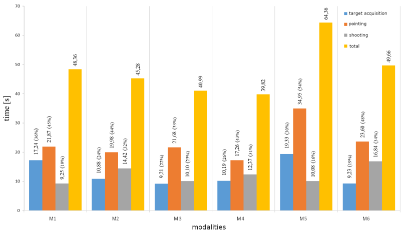

Fig. 3. Measured durations (target acquisition, pointing, shooting, total) of all six modalities.

TABLE III
RESULTS - ANOVA AND POST-HOC BONFERRONI TESTS OF DB MODALITIES, CONFIDENCE INTERVALS ARE PROVIDED ON 95% LEVEL

Time of steps	ANOVA	M5 and M1	M5 and M3	M1 and M3
Playing Time	F(1.14, 17.02)=46.74,p<.001	16.01,CI[8.49, 23.52] p<.001	23.38, CI[15.01, 31.75] p<.001	7.37, CI[4.84, 9.91] p<.001
Target Acquisition Time	F(1.44, 21.62)=195.53,p<.001	0.25,CI[0.08, 0.42] p<.001	1.14,CI[0.94, 1.34] p<.001	0.89,CI[0.79, 1.00] p<.001
Pointing Time	F(1.06, 15.93)=24.59,p<.001	1.14,CI[0.54, 1.74] p<.001	1.16,CI[0.52, 1.80] p<.001	not significant
Shooting Time	F(1.40, 20.93)=0.78,p=0.428			
Number of Shots	F(1.11, 16.60)=7.70,p<.05	1.54,CI[0.12, 2.95] p<.05	not significant	not significant

TABLE IV
RESULTS - ANOVA AND BONFERRONI POST-HOC TESTS OF VR MODALITIES, CONFIDENCE INTERVALS ARE PROVIDED ON 95% LEVEL

Time of steps	ANOVA	M6 and M4	M2 and M4	M6 and M2
Playing Time	F(1.39, 20.78) = 13.33,p<.001	9.33,CI[4.23, 14.43] p<.001	6.14,CI[3.03, 9.24] p<.001	7.37,CI[4.84, 9.91] p<.001
Target Acquisition Time	F(1.39, 20.80)=2.59,p=0.114			
Pointing Time	F(1.11, 16.71)=9.02,p<.05	0.58,CI[0.18, 0.98] p<.005	0.26,CI[0.12, 0.41] p<.005	not significant
Shooting Time	F(1.31, 19.58)=11.75,p<.05	0.39,CI[0.18, 0.60] p<.005	0.17, CI[0.04, 0.31] p<.005	not significant
Number of Shots	F(2, 30)=19.55,p<.001	1.41,CI[0.65, 2.17] p<.05	not significant	1.15,CI[0.52, 1.78] p<.001

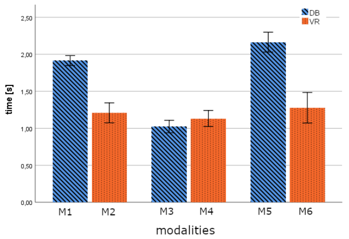

Fig. 4. Target acquisition times of all six modalities.

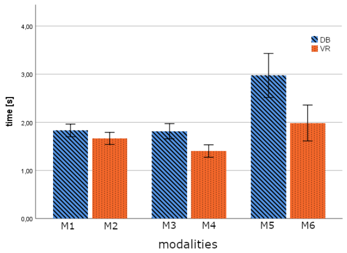

Fig. 5. Pointing times of all six modalities.

modalities are significantly shorter than that of the VR modalities (1,2: t(15)=6.98,p<.001; 3,4: t(15)=4.65, p<.001; 5,6: t(15)=4.45, p<.001).

E. Number of Shots

The average number of shots was significantly different within the DB modalities (Tab. III). There was a significantly higher value in the number of shots with the combination of keyboard and ET, also ET's standard deviation was the highest. The amount of shots did not differ significantly while using the mouse in free (M1) or centered mode (M3).

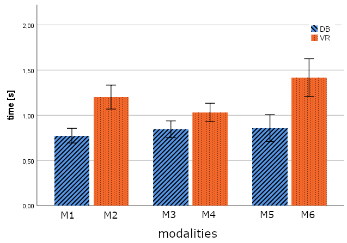

Fig. 6. Shooting times of all six modalities.

Back to Contents

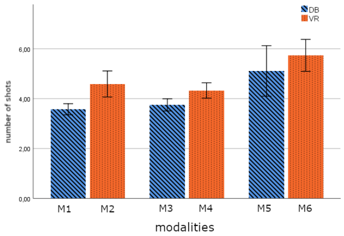

Fig. 7. Number of shots of all six modalities.

The number of shots was also significantly different in the VR modalities (Tab. IV). The highest shot rate had the device combination with ET (M6). Using a controller for shooting in combination with ray casting (M2) or head-tracking only (M4) did not differ in the shot rate.

The number of shots of the paired modalities (M1 vs. M2, M3 vs. M4) differed significantly (1,2: $t(15) = 3.65$, $p < .05$; 3,4: $t(15)=5.16$, $p<.001$). The VR modalities consistently had higher values in number of shots (Fig. 7). The paired modalities (M5 vs. M6) showed no significant difference (5,6: $t(15)=-1.14$, $p=.271$).

F. Qualitative Data

The evaluation of the questionnaire for the qualitative data acquisition resulted in the best ratings for the VR modalities. Ray casting (M2) was rated best, followed by ET (M6) and head-tracking (M4). The ratings of DB modalities are in this order: keyboard and mouse (M1), mouse only (M3) and finally keyboard with ET (M5).

VI. Discussion

First of all looking at the measured times for each step per modality, there may be some self-balancing effect due to closer and therefore larger perceived spheres. Longer times in target acquisition may lead to shorter time for the following steps, because the spheres are closer to the participant and therefore easier to target, however, the sphere is already sufficiently large for targeting when appearing at the spawning point. In general, the modalities that use ET (M5 and M6) require special consideration due to effects introduced by calibration. In particular the calibration of the Pupil Labs ET with the Pupil Capture software showed wide fluctuations in the quality of the calibration. In this study, the five-point method was used with no recalibration in order to avoid disturbance of the user. The calibration results were verified in the software but during the experiment issues occurred leading to discrepancies between the measured and the real gaze position. This led to difficulties in the pointing and shooting task. So both modalities had significantly higher values and variance in pointing time, which caused also higher net gaming duration compared to the other modalities. In [17] ET also showed high variance in performing selection and subjects rated ET as difficult to use and inaccurate.

ET used as pointing devices also affects the target acquisition times in M5 compared to M1. They used the keyboard in the same configuration for the target acquisition task. Therefore, similar results should be expected, but M5 yield higher values in target acquisition time. One possible explanation may be that the inaccurate calibration caused problems. The subjects needed a lot of concentration for pointing and leaving less attention for controlling the target acquisition task explaining the higher values.

This is different to the target acquisition times of the VR modalities using the head-tracking in all cases. Within them, there are no significant differences. That means ET does not seem to influence the target acquisition time to the same extent as it does in the DB modality. As the movement of the head is much more intuitive, we assume that this is the reason for the lack of influence on the target acquisition time with M6 or the better calibration of the FOVE compared to the pupil labs eye-tracker. Besides the calibration problems the fact that ET, as a novel input method, is not very familiar to the subjects and they had limited experience with it, may also cause higher pointing times.

All net game times of the VR modalities were shorter than the associated ones of the DB modalities. The main reason for this is the shorter target acquisition and pointing time. The slightly higher values for shooting were of less importance. The net gaming times were shortest in M3 and M4. In this cases, only one device was used to perform the target acquisition and pointing task yielding in less effort controlling two devices.

Concerning target acquisition times, the VR modalities were consistently faster. We assume that this is due to a lack of acceleration when entering the direction using the keyboard (constant rotation velocity while pressing the key) compared to head-tracking. This result was rather predictable, but we decided to use fixed speed rotation via input by keyboard as it is a commonly used input technique in current game based scenarios. There was an exception at mouse centered and head-tracking only. Both achieved equal target acquisition times, because both (compared to keyboard) allow similar intuitive acceleration of rotational movement.

The pointing times also were consistently lower for VR modalities. This may be explained by the fact that pointing on the spheres is not done by the pointing device solely, but also using the target acquisition device itself. Once the sphere appears in the subject's field of vision and the subject starts to aim to the sphere, the target acquisition process is still in progress. So the sphere will be placed in an optimal position (e.g. center of the screen) using the target acquisition device while the pointer simultaneously or subsequently approaches the desired position. The pointing process therefore mainly consists of the combination of target acquisition and pointing task. The faster rotational movement of head-tracking compared to the input of the keyboard probably leads to shorter pointing times in VR modalities. Using head-tracking only in M4 was also faster in pointing than using mouse only in M3 and so the most efficient modality.

The shooting times were smaller with DB modalities. Long-time experience in using the mouse might be the reason for the better results in M1 and M3. A reason for good shooting times in M5 may be the larger perceived shape of spheres. The higher target acquisition times and pointing times caused the sphere to be closer to the subject and therefore were easier to hit. Contrary to the VR modalities the spheres were further away during shooting and appeared smaller due to shorter target acquisition and pointing times, which made them more challenging for the subject to hit and, therefore, increased the shooting times. Also the number of shots tended to be higher for VR modalities than for DB modalities. The higher fire rates increased the probability of hitting the sphere successfully.

Overall, all six modalities were usable and effective in target acquisition, pointing and shooting tasks. The VR modalities were more efficient. They were faster in target acquisition and pointing, but slower in shooting and had more numbers of shots compared to DB modalities. Only target acquisition with mouse only (M3) is as fast as target acquisition with head-tracking (M2, M4, M6).

And pointing with the mouse (M1, M3) turned out to be as fast as pointing with an ET combined with head-tracking (M6). In [14] ET also turned out to be feasible as a pointing device in a shooting game, but it had a lower precision, than pointing with the mouse, however, only scores but no duration were measured. In [15] using the mouse also showed more accuracy than using ET, but ET was faster than mouse in selection task. In [5] and [12] ET was faster than using the mouse in a selection task, but the target objects did not move in their scenarios. Pointing with head-tracking was the fastest method and revealed best efficiency, although the subjects rated the combination of head-tracking and ray casting (M2) best. In [17] selecting with head movements only was also most effective and preferred by the subjects. Whereas in [6] head directed pointing was slower, than hand directed ray-casting, but the scenario used an image plane as projection and no VR-environment. In our study pointing with the mouse was slower than pointing with a controller using ray-casting, in [2] mouse pointing turned out to be faster. None of the participants reported a feeling of dizziness by using HMDs in the VR modalities. The reason for this might be the rather short sessions.

VII. Conclusions and Future Work

An interaction sequence consisting of target acquisition, pointing and shooting was performed in a DB user interface and a VR user interface using different modalities. For each modality, the time required for each step of the interaction sequence was measured and the effectiveness and efficiency of the modalities was evaluated from the obtained data and compared with each other. Using appropriate statistical tools, we searched for differences and similarities in the given dimensions. Overall, it turned out that the VR modalities were faster or as fast in their target acquisition and pointing steps as the corresponding DB modalities. The shooting step of the VR modalities, on the other hand, was slightly slower. Pointing time made up the largest share in the playing times. Head-tracking used for target acquisition and pointing was the most efficient of all the modalities observed, although the subjects rated the combination of head-tracking and ray casting (controller) best. Considering only the DB modalities, the best performing was the one using the screen centered mouse. In particular, for target acquisition, it reached similar efficiency compared to the best performing modality using head-tracking. ET yields longer times for pointing, probably due to calibration issues and resulting inaccuracy of the gaze position. In the next step, the influence of experience in VR and DB on efficiency should be analyzed in detail. This may either be done by a longer duration or a repetitive conduction of the experiment with the same observers. This seems important as an influence on the performance may be expected, in particular for the target acquisition and pointing step. In the next step, the influence of experience in VR and DB on efficiency should be analyzed in detail. This may either be done by a longer duration or a repetitive conduction of the experiment with the same observers. This seems important as an influence on the performance may be expected, in particular for the target acquisition and pointing step. Also extending the set up by the option of translational movement would be of interest, as this type of navigation is typically used in fps game based systems as well as replacing a method with a classical game controller as input device for comparison purposes. For example in [18] and [19] the authors obtain better results in pointing performance and throughput by using the mouse instead of a controller in fps scenarios. Furthermore enhancing the graphics should be considered in order to obtain a more realistic environment, which leads to enhanced feeling of immersion and could affect measured values.

References

[1] R. J. K. Jacob, "What you look at is what you get: eye movement-based interaction techniques," in *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*. Seattle, Washington, United States: ACM Press, 1990, pp. 11–18.

[2] R. J. Teather and W. Stuerzlinger, "Pointing at 3d target projections with one-eyed and stereo cursors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. Paris, France: ACM Press, 2013, p. 159.

[3] R. P. McMahan, A. J. D. Alon, S. Lazem, R. J. Beaton, D. Machaj, M. Schaefer, M. G. Silva, A. Leal, R. Hagan, and D. A. Bowman, "Evaluating natural interaction techniques in video games," in *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, Mar. 2010, pp. 11–14.

[4] R. Kopper, F. Bacim, and D. A. Bowman, "Rapid and accurate 3d selection by progressive refinement," in *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, Mar. 2011, pp. 67–74.

[5] J. Chun, B. Bae, and S. Jo, "BCI based hybrid interface for 3d object control in virtual reality," in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)*. Gangwon Province, South Korea: IEEE, Feb. 2016, pp. 1–4.

[6] S. Lee, J. Seo, G. J. Kim, and C.-M. Park, "Evaluation of pointing techniques for ray casting selection in virtual environments," in *Proceedings Volume 4756, Third International Conference on Virtual Reality and Its Application in Industry*, Z. Pan and J. Shi, Eds., Hangzhou, China, Apr. 2003, pp. 38–44.

[7] D. A. Bowman, D. B. Johnson, and L. F. Hodges, "Testbed Evaluation of Virtual Environment Interaction Techniques," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 1, pp. 75–95, Feb. 2001.

[8] D. A. Bowman, D. Koller, and L. F. Hodges, "Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques," in *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, Mar. 1997, pp. 45–52.

[9] A. Kaufman, A. Bandopadhay, and B. Shaviv, "An eye tracking computer user interface," in *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium*. San Jose, CA, USA: IEEE Comput. Soc. Press, 1993, pp. 120–121.

[10] J. Gips, P. Olivieri, and C. Hill, "EagleEyes: An Eye Control System for Persons with Disabilities," in *Proc. 11 th Int. Conf on Technology and Persons with Disabilities*, 1996, p. 15.

[11] Y. K. Meena, H. Cecotti, K. Wong-Lin, and G. Prasad, "A multimodal interface to resolve the Midas-Touch problem in gaze controlled wheelchair," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Seogwipo: IEEE, Jul. 2017, pp. 905–908.

[12] L. E. Sibert and R. J. K. Jacob, "Evaluation of eye gaze interaction," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*. The Hague, The Netherlands: ACM Press, 2000, pp. 281–288.

[13] D. Fono and R. Vertegaal, "EyeWindows: evaluation of eye-controlled zooming windows for focus selection," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*. Portland, Oregon, USA: ACM Press, 2005, p. 151.

[14] P. Kasprowski, K. Harezlak, and M. Niezabitowski, "Eye movement tracking as a new promising modality for human computer interaction," in *2016 17th International Carpathian Control Conference (ICCC)*. High Tatras, Slovakia: IEEE, May 2016, pp. 314–318.

[15] R. Vertegaal, "A Fitts Law comparison of eye tracking and manual input in the selection of visual targets," in *Proceedings of the 10th international conference on Multimodal interfaces - IMCI '08*. Chania, Crete, Greece: ACM Press, 2008, p. 241.

[16] E. D. Ragan, S. Scerbo, F. Bacim, and D. A. Bowman, "Amplified Head Rotation in Virtual Reality and the Effects on 3d Search, Training Transfer, and Spatial Orientation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 8, pp. 1880–1895, Aug. 2017.

[17] Y. Y. Qian and R. J. Teather, "The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality," in *Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17*. Brighton, United Kingdom: ACM Press, 2017, pp. 91–98.

[18] K. M. Gerling, M. Klauser, and J. Niesenhaus, "Measuring the impact of game controllers on player experience in FPS games," in *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11*. ACM Press, 2011, p. 83.

[19] D. Natapov, S. J. Castellucci, and I. S. MacKenzie, "ISO 9241-9 evaluation of video game controllers," p. 8.