Faster than in Real ime
-
Prediction of Human Motions in Virtual Reality

Andreas Pfaffelhuber
andreas.pfaffelhuber@stud.uni-regensburg.de
Universität Regensburg
Regensburg, Deutschland

Christoph Tögel
christoph.toegel@stud.uni-regensburg.de
Universität Regensburg
Regensburg, Deutschland

Julian Dietz
julian.dietz@stud.uni-regensburg.de
Universität Regensburg
Regensburg, Deutschland

ABSTRACT

Experiences in Virtual Reality (VR) strongly depend on the immersion of the user in the computer generated world. Hardware delays and a resulting asynchrony from real world motions to the virtual motions can induce motion sickness and overall a worse experience. Using machine learning on motion capture data, a user's motions can be predicted and shown in advance. In this study an apparatus to improve VR experiences by reducing delays or even present user's their future motions is proposed. Therefore experience in VR is evaluated using subjective and objective measures on a two-dimensional, multidirectional Fitts' law task. In the study 24 participants wearing full body motion capture suits were presented different motion alterations with past, present and multiple future versions of their motions. Performance was evaluated using a standardized Fitts' law task under six conditions as well as using three questionnaires for subjective measures. Initially expected improvements of performance as well as experience could not be found, but a worsening of both measures was observed.

KEYWORDS

virtual reality, neural networks, motion prediction, embodiment, presence, Fitts law

ACM Reference Format:
Andreas Pfaffelhuber, Christoph Tögel, and Julian Dietz. 2020. Faster than in Real ime - Prediction of Human Motions in Virtual Reality . In *Regensburg 20: Forschungsseminar, March, 2020, Regensburg, DE.* , 11 pages.

1 INTRODUCTION

When users move their limbs in Virtual Reality (VR) through active and voluntary motoric motions and the brains expected outcome positions match the received sensory afferent modalities like gaze and proprioception, a strong VR-Illusion is perceived [12]. Contrary, if the difference between the afferent sensory inputs and the expected outcome position is too big, the virtual limb-ownership illusion is often rejected by users [12]. Several studies in the field of VR experiences have demonstrated that body ownership illusion is induced with visuo-motor triggers when both real and fake bodies move their homologous body parts at the same time [4, 15]. Similarly, it has been observed that the body ownership illusion fails to be established when either temporal synchronicity [37] or spatial congruence [7] between the seen and the felt motions are not met. As shown above, the perceived plausibility of ones body perception and its congruency with motions in the virtual reality environment is of great importance for immersive experiences. To prevent a delay of one's virtual body motion through hardware-based factors, the current body motion can be used to predict the future position of limbs. Systems capable of predicting a user's motions are able to remove the hardware delay of the displayed virtual body, or even take the prediction one step further and present the user with a position that is still located in the future. This modified version of the motion allows for a faster task completion as it may better represent a users intention when interacting in the virtual space.

In an attempt to solve this problem, motion data has been recorded from 20 participants in a previous study performing prescribed and free form tasks . This motion data describes a skeleton consisting of 51 bones and positional data. Neural networks can be trained on this data to compute a different version of the skeleton, which is evaluated in this study. Depending on the input data this can be done for multiple time offsets in the past or future. The effect of the time offsets on user experience and motions can be measured objectively and subjectively. The final user performance will be measured as the throughput on a standardized Fitts' law tapping task [38] presented in VR, as already established in other related work [35]. This study aims to compare the user's performance, presence, avatar embodiment and acceptance of the proposed system at different PREDICTION TIME OFFSETS in virtual reality. This approach expects to find a certain PREDICTION TIME OFFSET which produces an observable maximum in user performance, presence as well as avatar embodiment and thus provides a better VR experience.

The following paper presents related work for virtual body alteration and limb ownership, motion prediction and Fitts' law performance evaluations. Building on previous works we use an objective, quantitative method to evaluate body alterations as well as the user's subjective feedback. The proposed Fitts' law task allows to compute throughput as a performance measure. Our findings show a worsening of performance depending on the strength of alteration of a user's motions contrary to initial expectations. While the results do not support the hypothesis, this study indicates a correlation between alteration and performance. The performance

Forschungsseminar 19/20, March 9, 2020, Regensburg, DE

never plateaus which might be caused by the apparatus and not by the motion prediction. As noted in later sections the used apparatus was limited by available hardware and constrained to a certain approach when defining the neural network architecture. Therefore the observed effect might be inverted to a performance increase when lifting these.

2 RELATED WORK

The related work covers findings on avatar embodiment and limb ownership as well as previous body alteration studies and their effect on the presence felt within a virtual environment. Since it is a main theme of this study, related work on the effects of differences between displayed and real limb positions is presented. Additionally, a quick overview on Fitts' law and its meaning for human computer interaction is given as it is later used to quantitively measure performance of conditions.

2.1 Virtual Limb Ownership

The concept of body ownership refers to the sense that one's own body is the source of sensations. It involves a strong afferent component that indicates the state of the body through various peripheral signals and is not only present while performing voluntary actions, but also during inactive experiences [40]. Similarly, the sense of embodiment refers to the ensemble of sensations that arise in conjunction with being inside, having, and controlling a body [16]. The concept of presence describes the sense of being in the virtual world rather than still consciously being in the real physical world [30]. Since body ownership refers to the form of identification with the body as a whole, limb-ownership refers to only certain parts of the body [6].

The "rubber-hand illusion" [5] shows that alien limbs can be perceived as being a part of the own body through perception and simultaneous tactile stimulation of a fake and real hand. Similar findings have been made concerning the adaptation of virtual limbs or bodies. Perez-Marcos et al. induced a virtual hand ownership illusion and showed that some aspects of the illusion occur through motor imagery used to control movements of the virtual hand. When the movements of the virtual hand followed motor imagery, the illusion of ownership of the virtual hand was evoked and measured muscle activity correlated with movements of the virtual arm [25].

Considering the topic of perceived limb ownership in a virtual environment, Rietzler et al. conducted a study that works with direct manipulation of limb positions [28]. Their goal was to simulate weight in the virtual reality by slightly displacing the subjects virtual arm position and thus for example forcing them to lift their arm higher than usual if an object should be heavier. Pure visual displacement is used instead of a physical actuator allowing for modification based on software only. While weights could only be simulated in a relative manner, the approach showed to have positive effects on the user experience, leading to much higher presence and immersion during the use of the apparatus [28]. Adjusting the virtually displayed limb positions generally suggests to be a viable approach since in VR, visual cues can overshadow haptic impressions under certain conditions. For example, Kohli [18] built an apparatus which mapped real world touches to visual feedback by snapping virtual limbs onto virtual object representations in order to align the haptic sensation. This improved the experience only up to a certain threshold as the visual cues did overshadow haptic impressions. Visual displacement of the virtual body position can also be used to simulate slow-motion in a VR environment, and has also shown to be accepted by the participants under this condition. Proving this, another one of Rietzler's works [27] explored virtual limb-transformation in the past time direction, slowing down motions and using previously captured data instead of predicted future data like this current study aims to. Using multiple Microsoft Kinect V2 cameras and a virtual environment, multiple filters were deployed to compute the slowed motion, allowing for a dynamic slow-motion experience. The strength of this effect is linked to the velocity of motions by the participant [27]. In a game of hitting bubbles no decrease in presence could be found. This may be caused by the user-controlled nature of the slow-motion effect, but also indicates once again that users accept adjustments of their virtual body position.

2.2 Motion prediction

Another study using limb-manipulation was conducted by Kasahara et al. [14], who developed a system that captures full-body motions and generates estimated past and future body positions using a time-based alteration model where the amount of temporal shift could be adjusted as a continuous value. They investigated how human movement and the user's observable behavior change according to the virtual body alterations. Results showed that spatial-temporal alteration of a virtual body results in perceivable changes of the physical feeling as well as physical movement. For example a virtual human body generated at approximately 25-100 milliseconds in the future induced a "lighter weight" sensation [14].

Other works in the context of predicting motions in virtual reality have already been conducted, many of them centering around predicting the user's head movement. While some approaches include creating and examining algorithmic methods towards this goal [3], others follow a similar approach as this study and use neural networks. In one case, both recurrent and time delayed neural networks have been used to predict the future head position, and both prediction systems have shown to be usable and provide adequate results [29] in the virtual reality context.

The state-of-the-art approach of using machine learning to predict movement has already shown to be able to enhance the user's performance in multiple scenarios, for example when using a touchscreen in a non-virtual, two-dimensional paradigm [19].

Considering the successful application in such a scenario as well as the shown applicability of the same method within virtual reality environments [29], we decided to also use machine learning to predict the user's motions and alter their virtual body position accordingly. The presented related work already tried similar fashioned approaches and managed to enable the user to have a better virtual immersion or to create new experiences in the virtual world. The now conducted study and its used apparatus try to enable the motion prediction approach for the entire virtual body in a three-dimensional, virtual environment and aim to improve performance in selection tasks as well as the immersive experience.

2.3 Fitts' Law

To objectively compare the overall performance of selection tasks during the study, a standardized Fitts' law task is used. Fitts' law [10] itself is an information-theoretic view of human motor behavior which expresses the time to complete a movement task (MT) in terms of distance or amplitude of the move (A) and the width of the region within which the move must terminate (W). With the intercept (a) and the slope (b), the performance of a pointing device can be evaluated [21]. Fitts' law over the years has been verified over a wide range of conditions and has been applied by HCI researchers in primarily two ways. Firstly, it can be used as a predictive model to e.g. calculate the estimated time for the user of a graphical interface to move the mouse tracker to a button and click on it. Secondly, it is used for the comparison and evaluation of novel pointing devices or conditions [38]. Instead of predicting movement times, researchers measure several of them and then determine how the different conditions and devices affect the coefficients within the Fitts' law relation. Fitts' law offers the utility to compress several movement time measurements into a single statistic, the throughput. Throughput combines both speed and accuracy, and its value in describing input devices and conditions has been both academically and industrially recognized [38].

2.4 Summary

The presented work shows an overall interest in modifying human motions in virtual environments in order to significantly alter or improve the user experience. The related work on virtual limb ownership demonstrates that displacements influence the user experience and might be used to enhance the virtual experience due to often being accepted by the user, as long as the displacements are not making too drastic changes to the body position. This study now focuses on evaluating the use of deep learning to create prediction models for future motions. It also focuses on quantifying the results and experience of the displacements with standardized measures. As mentioned above, neural networks have already shown that they are capable of improving user performance in a two dimensional scenario, so applying them in a virtual, three-dimensional environment might yield similar results. To test this the conducted study evaluates how participants perform a standardized Fitts' law task, which is widely recognized and used by the scientific community.

3 GENERAL METHOD

In this paper an evaluation study of different PREDICTION TIME OFFSETS for predicting human motions using a horizontal, two-dimensional, multidirectional Fitts' law task in a virtual environment is conducted. In order to record user motion data we use a full-body motion capture system (OptiTrack) with its proprietary "Motive" recording software at version 2.1.1. During the study participants wear a full-body motion capture suit with 49 passive markers. As even small motions can be dependent on bigger bone and muscle structures it is necessary to capture the whole body. Due to imprecision of a finer marker setup, only an abstract marker configuration for hands is used. With this only the thumb, index finger and pinkie are captured with the remainders interpolated.

Table 1: Time prediction offsets used in the study.

Name	Frames	Prediction Time [ms]
Past model	-12	-48
Base system	0	0
Zero Latency model	+12	+48
Future model 2	+24	+96
Future model 3	+36	+144
Future model 4	+48	+192

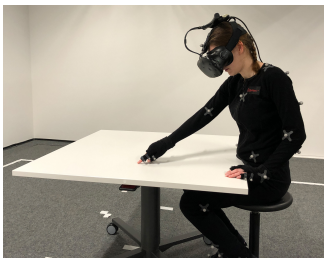

Figure 1: Participant performing the Fitts' law task wearing a full-body motion capturing suit.

3.1 Prediction Times Determination

The tested PREDICTION TIME OFFSETS in this experiment have been determined from pervious experiences with the apparatus. The latency of the base system caused by hardware was determined, which is used as the first offset towards predicting a future version. Apart from this condition that eliminates the hardware delay through the used PREDICTION TIME OFFSET, three additional ones are located in the future. All PREDICTION TIME OFFSETS then use the same time interval between them, so that the resulting models can be linearly compared. Additionally the *base system* without any prediction is evaluated as a condition. As the sixth condition a *past time model* is used. This condition differs from the future ones, as it does not use an algorithm to determine motions but rather presents buffered data from the user. It has the same offset as the *zero latency model* but in the other time direction. This way it can be evaluated how performance changes according to the different PREDICTION TIME OFFSETS compared to the *base system*. The *past time model* allows to calculate a regression model along the different evaluations and serves as a plausibility check. The different conditions can be seen in table 1.

Four future prediction models have been chosen so that some of the models should over-predict the user's motions up to a point where performance starts to decline again. This leads to the expectation to model a user's performance more detailed.

3.2 Evaluation Study

The hypothesis of the evaluation study is that different PREDICTION TIME OFFSETS will have an effect on user performance, presence and avatar embodiment. These are expected to increase from the past model until a slightly-in-the-future model and then decrease again. To test this hypothesis, the evaluation study will be conducted using a within-subjects design that consists of six different PREDICTION TIME OFFSETS. For each condition participants first perform the

Fitts' law task and then answer three questionnaires, as well as some qualitative questions orally. The six PREDICTION TIME OFFSETS are evaluated using a 6×6 counterbalanced latin square design. The system is tested on 24 participants, which is similar to other studies in the same context [14, 27].

Using G*Power the size of the sample was evaluated. This study conducted a within-factors repeated measures design. Therefore, assuming a significance level of 5 percent ($\alpha = .05$) and a power of 95 percent , even a small effect size of Cohen's d (.3) could be observed. This results in a minimum of 20 required participants [8, 9]. In order to align with the latin squared design the sample was extended to 24 participants.

3.3 Fitts' Law Design

The user performance will be measured quantitatively using the throughput of a standardized, two-dimensional multidirectional Fitts' law tapping task for each condition. The entire task is laid out according to the established recommendations by Soukoreff and Mackenzie [38], which go in accordance with and supplement the ISO 9241-9:2000. This norm has been revised by ISO 9241-400:2007 and ISO 9241-411:2012. The task consists of a circular arrangement of multiple round targets, which then have to be tapped by the participant (See figure 2). These are displayed on a horizontal virtual desk. The procedure always starts with the upper middle target, then the target at the opposite side of the circle is chosen next. This way the participant taps clockwise through the circle.

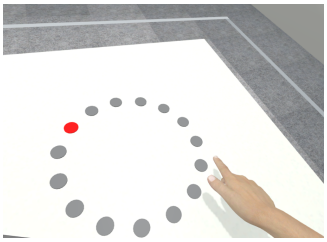

Figure 2: Participant performing the Fitts' law task in the virtual environment.

In our study two repetitions of nine different Fitts' law circle configurations (3 target sizes(W) $\times$ 3 target distances(D)) are performed. Every circle consists of 15 clickable targets, and participants thus perform 288 taps per PREDICTION TIME OFFSET condition. According to acknowledged recommendations [38] and a similar setup by Schwind et al. [34], this setup captures sufficient data to compare the different conditions. The different configurations appear randomly ordered for each condition. The targets are big enough to be precisely activated using the virtual index finger (W), the distances cover the whole table (D). To compute the initial index of difficulty (ID) for the different circle configurations the Shannon formulation is used, which offers an improved link with information theory as well as better correlations compared to other models proposed by Welford or Fitts [21, 22].

The employed setup covers indexes of difficulty for the different configurations that range from 2.26 to 4.83 bits, which roughly compares to general recommendations [38] as well as previous related work [34]. While participants perform the Fitts' law task, multiple quantitative measures are captured to evaluate their performance during the different movement conditions. Movement times as well as a measure of the scatter of the participants' movement is gathered. For this, we log the physical end-points of each movement task.

3.4 Questionnaires

Presence as well as embodiment are evaluated using both the igroup presence questionnaire (iPQ) [31] and an avatar embodiment questionnaire [13]. The iPQ is a widely used tool in the field of VR evaluation and was developed and refined over a multitude of studies [26, 32, 33].

The avatar embodiment questionnaire was created by Gonzalez-Franco and Peck [13] in an attempt to provide a standardized questionnaire to measure embodiment. Since it has multiple sub-scales the questionnaire was shortened down to three, namely "body ownership", "agency and motor control" and "location of the body". These scales are also weighted highest by the original authors [13] and are the only applicable ones for this setup. For a shallow investigation of motion sickness another questionnaire is presented, though only the four most significant questions named by the authors [11] are used. For the study all questionnaires are shown after each PREDICTION TIME OFFSET in randomized ordering of their respective questions using an online survey tool. The participants leave the virtual environment in order to fill them out on a laptop. Furthermore six question which can be answered in a free-from manner are asked after the questionnaires:

- Wie hat sich dein Körper während dieses Durchlaufs angefühlt? (How did your body feel during this iteration?)
- Was ist dir allgemein hierbei aufgefallen? (What did you notice during this iteration?)
- War diese Erfahrung positiv oder negativ? (Was this experience positive or negative for you?)
- Hast du einen Unterschied zu den bisherigen Läufen bemerkt? (Did you notice a difference in regards to the other iterations?)
- Konntest du deine Auswahl präzise treffen? (Were you able to chose your targets precisely)
- Wie hat sich der digitale Avatar angefühlt ? (How did the digital avatar feel to you?)

These questions are presented verbally in German as all participants were native speakers.

3.5 Procedure

Participants are first welcomed and asked to fill out an informed consent form as well as a demographics survey. They are told that they will be testing different motion tracking configurations while performing the Fitts' law task described in section 3.3 followed by multiple questionnaires and free-form questions. They are instructed to use their right hand index finger for this. Participants don't know about the motion prediction and are thus expected to not be biased in their impressions. Participants are additionally informed that their participation is on a purely free basis and that they can abort the experiment at any given time if they wish to.

The volunteer then take on their motion suit and head-mounted display (HMD) to enter an VR replica of the laboratory. They then are presented the first of the PREDICTION TIME OFFSETS and perform the Fitts' law task followed by the used questionnaires and the free-from questions. They are allowed to take a short break after each condition, repeating the process until all six conditions are completed. At the end the participant is thanked for taking part in the study and is compensated.

3.6 Sample

The study consisted of 24 participants (13 Males, 11 Females). All participants were right-handed with the exception of one. The mean age was 23.21 years-old (*SD* = 3.24 years). Participants were either wearing contact lenses, took their glasses off or were comfortable with wearing them underneath the HMD. Students were compensated with credit points, which resembled most of the test subjects. Remaining participants got sweets. Additionally the height of the students was recorded with the average being 1.77 meters tall (*SD* = .08 m). Neither of the participants reported to have much previous experience using virtual reality devices.

4 APPARATUS

4.1 Motion capturing system

In order to record user motion data we use the motion capture system (OptiTrack[1]) with the proprietary "Motive" application (version 2.1.1) as recording software. During the study a whole skeleton is captured using 49 passive markers. Participants wear a full-body motion capture suit recorded by 12 OptiTrack "Prime 13W" infrared cameras[2]. These cameras capture at 240 frames per second with a latency of 4.2 milliseconds for this study. Due to limitations the study only used passive markers for fingers. As participants are instructed to tap the targets with their index finger, precision is unlikely to be affected by the motion tracking during the Fitts' law task. This finger has a dedicated marker on the tracking suit's gloves and is not interpolated.

Motive provides functionality for streaming the captured data over network, based on the proprietary "NatNet-Stream" protocol[3]. The streamed data can be received in Unity[4] with the use of the OptiTrack plug-in (version 1.2) provided by OptiTrack[5] themselves. The data of individual markers, rigid bodies and skeletons from Motive can then be used to animate Unity bones or objects. Since the stream can be received by any machine within the local network, it is possible to modify the data, repackage it and then send it to the unity engine running on a a second PC. This setup is provided by the laboratory. Using Python, a client was developed which allows to toggle the modification of the incoming skeleton data based on the desired PREDICTION TIME OFFSET. Multiple TensorFlow models are loaded and predict the data accordingly to the selection in the UI. The given setup allows for mean computational time of 1.42 ms (*SD* = .36 ms) of the motion alterations using neural networks. The *base model* just passes the incoming data on without modifying it.

[1]https://optitrack.com
[2]https://optitrack.com/products/prime-13w/
[3]https://optitrack.com/products/natnet-sdk/
[4]https://unity.com
[5]https://optitrack.com/unity-integration/

For the *past time model* an array is used as a buffer for the incoming data. Once enough data is loaded (12 frames resembling 48 ms) each subsequent incoming frame leads to the oldest frame being streamed.

4.2 Neural Network

In order to train a regression model for the prediction, deep learning is used. Related work shows the effectiveness of this method [19]. A neural network was trained on the quaternion data of previously recorded data outside of this study. The data set consists of 20 participants performing various motions over roughly 45 minutes of recording time. During this duration three types of task were presented and performed. For this particular study only data of motions from general motions was used. This equals to roughly 15 minutes of recorded data per participant. Keras with TensorFlow as the backend is used as the underlying architecture. The model is trained on the rotation of the bones in the quaternion data format.

The data consisting of 207 columns can be sectioned into three parts. In the first the hip position is transmitted. Three columns describe the XYZ-coordinates within the scene. This bone is used to position all others, using local rotations. The second section, up to column 87, describes the rotation of all bones as quaternions. Due to imprecision the third section describing data of the finger bones is excluded. Using local rotation data, the machine learning framework is only exposed to a certain value range on each particular bone, resembling natural human ranges of motions, resulting in a smaller range of plausible values. This is also the main assumption behind training a regression model.

The neural network for the future predictions is then trained on 87 input dimensions per frame (first section of the data), using a time shifted version of the data set as the output targets. This is done for each PREDICTION TIME OFFSET describing a future version of the movements, creating different models for each setting. The *past model* and *base model* do not use neural networks.

The network has 87 input neurons resembling the orientation of the bones as well as the hip position. The first dense layer contains 8096 units and is fully connected to the second dense layer also consisting of 8096 units. These layers use the built-in ReLu activation for activation [1]. For training an additional dropout layer to combat overfitting [39] (droput-rate = .1) is added. The last dense layers reduces the data down to the dimensionality of the skeleton. This apparatus has been adopted from another study in order to validate results.

The OptiTrack stream delivers new data at 240 Hz and the training data was recorded at this frequency. As an example, shifting the data by 12 frames corresponds to a prediction of about 48 milliseconds. For stochastic optimization the built in ADAM optimizer[17] of the TensorFlow framework[6] is used. The training process was started at a learning rate of .001 and batch size of 1024 samples. By using Keras callback functions[7], the learning rate is dynamically adjusted during the training. Depending on the validation accuracy, the learning rate is adjusted by the training algorithm, optimizing

[6]https://keras.io/optimizers/#adam
[7]https://keras.io/callbacks/

Table 2: Prediction models

Name	Shifted frames	Validation Accuracy [%]
Zero Latency model	+12	94.2
Future model 2	+24	91.4
Future model 3	+36	89.5
Future model 4	+48	86.6

runtime. The Mean Squared Error (MSE), is used as the loss function. The final validation accuracy of each model can be seen in table 2.

During the experiment only the modified data for the shoulder, upper arm, lower arm and hand is used. No data for the head will be used, as accelerated and dissimilar head motions can induce motion sickness in VR users [2].

4.3 Testing environment

The participants perform the task while sitting close to a square table (1m×1m). The table plate functions as haptic feedback when hitting the targets of the Fitts' law configuration. Before each run the table gets repositioned with its height set to 76 cm above the floor, resembling common desk heights. The virtual task is then displayed on a virtual representation of the table. Haptics are mostly congruent from the real world to the HMD. Participants are told to sit upright in order to keep an equal distance to the targets throughout the studies duration. They also remain seated on a turnable stool for the whole experiment. Additionally the Unity engine renders real time soft shadows allowing for easier spatial orientation when trying to hit the targets.

5 EVALUATION

5.1 Data Analysis

To analyze the captured data and to compare the different PREDICTION TIME OFFSETS, a multitude of commonly used performance measures regarding Fitts' law tasks were employed. This includes the comparison of the effective throughput, mean movement time and accuracy, as well as an analysis of the questionnaire data described in chapter 3.4. In order to determine the effective throughput (TP_e), the effective index of difficulty (ID_e) was used instead of the default formulation. The effective index of difficulty adjusts the specified target width according to the spatial variability in the human operator's output responses over a sequence of trials. Using this adjustment for accuracy, the effective index of difficulty better represents the actual difficulty of the task for the individual participant who performed it [23]. The effective index of difficulty is calculated using equation 1:

$$ID_e = log_2(\frac{A_e}{W_e} + 1) \tag{1}$$

To compute the effective target width (W_e), the standard-deviation method was used [23]. The endpoint coordinates of every target selection made by the participant were recorded. The standard deviation of the offset from a targets center point regarding the direction of the movement (SD_x) for each circle configuration is then used to calculate the effective width:

$$W_e = SD_x * 4.133 \tag{2}$$

To compute the effective target amplitude (A_e), the mean of the actual hit-to-hit movement distances for each circle configuration, as projected on the movement axis, was used.

$$ID_e = log_2(\frac{A_e}{SD_x * 4.133} + 1) \tag{3}$$

Using the mean movement time (MT) for a single goal selection task of the circle configuration, the effective throughput was calculated using Equation 4:

$$TP_e = \frac{ID_e}{MT} \tag{4}$$

In order to compare the different PREDICTION TIME OFFSETS, the mean of all circle configurations is used to get a single value for each participant and PREDICTION TIME OFFSET. This method is applied when analyzing the effective throughput, accuracy and movement time.

5.2 Results

On average, participants spent 41.7 minutes (SD = 24.2 min) completing the Fitts' law part of the experiment in VR. Completing some PREDICTION TIME OFFSETS took longer than others, mostly on the *+192 ms model*. The average completion time of the whole experiment was roughly 75 minutes. Every participant performed six different PREDICTION TIME OFFSETS, each of them consisting of 2×9 different circle configurations with 16 target selections. Thus, participants performed a total of 1728 target selections. The study was conducted using a counter-balanced, repeated measures design with one independent variable (6 PREDICTION TIME OFFSETS). Thus the six groups can be tested on differences in their mean values using a one-way repeated measures ANOVA. When the assumptions for the ANOVA are not met the Friedmann-Test, a non-parametric alternative, can be employed. Since none of the evaluated datasets could completely fulfill all requirements for the ANOVA, the Friedmann-Test was always used. All tests are computed using a significance level of 5 percent (α = .05).

Throughput

The throughput describes the information transmitted per task. Since the Shapiro-Wilk test showed that not all of the throughput values could be described by a gaussian distribution, the assumptions for a repeated-measures ANOVA were not met and the Friedmann-Test was used instead. The effective throughput values were statistically significantly different at the different PREDICTION TIME OFFSETS (Friedman-Test: $\chi^2(5)$ =84.79, $p < .001$, n = 24, F). Post-hoc pairwise Wilcoxon signed-rank tests were computed for all possible PREDICTION TIME OFFSET pairs. The Bonferroni-adjustment was used to counteract the skewing effects of conducting multiple tests on the same dataset. The resulting Bonferroni-adjusted p-values for the multiple comparisons can be seen in table 3.

The pairwise comparisons showed a significant difference on all of the models except for the *-48 ms model* compared to the *+48 ms model* (p = 1), the *+96 ms model* compared to the *+144 ms model* (p = .131) and the *+144 ms model* compared to the *+192 ms model* (p = .436). The mean effective throughput as well as variance and median for each PREDICTION TIME OFFSET can be seen in

Table 3: Bonferroni-adjusted *p*-values of pairwise Wilcoxon signed-rank tests computed on the effective throughput.

	Base	+48 ms	+96 ms	+144 ms	+192 ms
-48 ms	.007**	1	<.001****	<.001****	<.001****
Base		.004**	<.001****	<.001****	<.001****
+48 ms			.002**	<.001***	<.001****
+96 ms				.131	.001**
+144 ms					.436

figure 3. The mean effective throughput over all PREDICTION TIME OFFSETS was 3.09 bit/s (*SD* = .97 bit/s). The statistical evaluation confirmed that there were significant differences between all not previously mentioned groups, and that the PREDICTION TIME OFFSETS had an effect on the achieved effective throughput. Despite showing an effect of the PREDICTION TIME OFFSETS, it is not the initially expected one. The initial expectation was that some future prediction models would produce a higher throughput than the baseline system. Especially the *+48 ms model* was supposed to combat the hardware delay and offer a real-time user experience. Instead the findings indicate that any modification of user movements using the employed apparatus was worse than the baseline system.

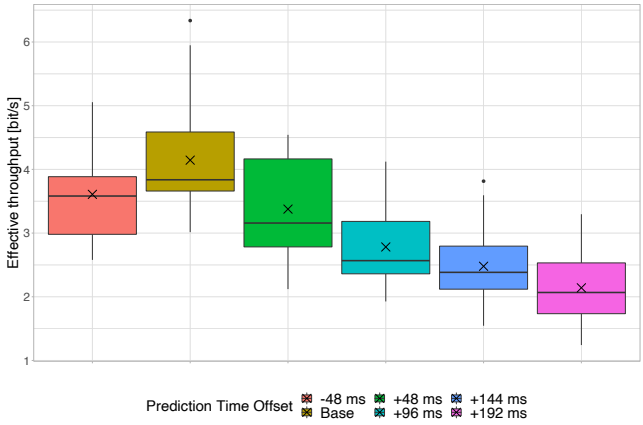

Figure 3: Boxplot of the effective throughput (TP_e) for each PREDICTION TIME OFFSETS.

Overall, harsher modifications lead to worse results. Interestingly, the pairwise Wilcoxon signed-rank test showed no significant difference between the past modification and the *+48 ms model*.

It has to be noted that the throughput values observed in this study differ from other related studies, which have been conducted using a similar setup and indexes of difficulty [35]. Throughput values are lower here, which might indicate that participants were not successfully motivated to conduct the task as fast as possible. Another explanation might be that they were intimidated by the negative effects of the prediction models which when once observed caused them to perform slower under all conditions.

Table 4: Bonferroni-adjusted *p*-values of pairwise Wilcoxon signed-rank tests computed on the mean movement time.

	Base	+48 ms	+96 ms	+144 ms	+192 ms
-48 ms	.08	1	<.001****	<.001****	<.001****
Base		.003**	<.001****	<.001****	<.001****
+48 ms			.002**	.001**	<.001***
+96 ms				.736	.001**
+144 ms					.182

Movement Time

The mean movement time for each participant performing a single target selection task was analyzed for the different PREDICTION TIME OFFSETS. Conducting Shapiro-Wilk tests on the participants mean movement times once again showed that those were not normally distributed in all of the conditions. Since prerequisites for a one way repeated-measures ANOVA were not met, the Friedmann-Test was used. Concerning movement times, there was again a significant difference between the tested groups (Friedman-Test: $\chi^2(5) = 78.41$, $p < .001$, n = 24).

Similar to the throughput, post-hot pairwise Wilcoxon signed-rank tests showed that there were significant differences for all conducted comparisons on movement time, except for the *-48 ms model* compared to the *+48 ms model* ($p = 1$), the *+96 ms model* compared to the *+144 ms model*($p = .131$) and the *+144 ms model* compared to the *+192 ms model* ($p = .436$), as can be seen in table 4.

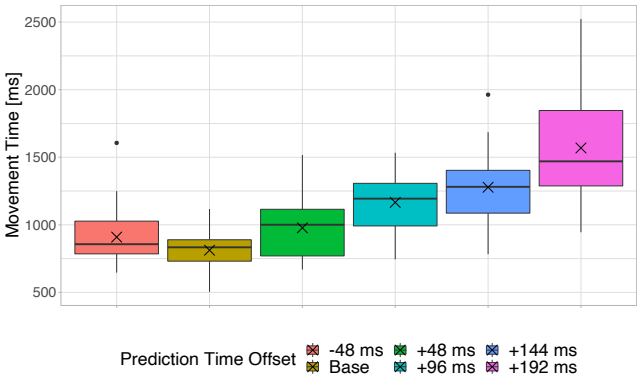

Figure 4: Boxplot of the mean movement times (*MT*) for each PREDICTION TIME OFFSETS.

Additionally, the comparison between the *base model* and the *-48 ms model* showed no significant difference this time. Using the given apparatus, the *base model* had the lowest mean movement time for a single target selection task with 812 ms (*SD* = 156 ms), followed by the *-48 ms model* with 909 ms (*SD* = 218 ms). The worst performing condition was the *+192 ms model* with a mean movement time of 1569 ms (*SD* = 444 ms). The average movement time over all conditions combined was 1119 ms (*SD* = 370 ms). A visualization of the movement times for each condition can be seen in figure 4.

Mean movement times were the best for the *base model* and got worse for each further PREDICTION TIME OFFSET. As mentioned previously, the *-48 ms model* produced mean movement times closest to the *base model* and even outperformed the *+48 ms model*.

For every PREDICTION TIME OFFSET, the characteristic Fitts' law linear relationship between ID and movement time can be established using linear regression models. The parameters of the linear equation indicate better performance by smaller intercepts ($a = \beta_1$) and shallower slopes ($b = \beta_2$) [23]. Figure 5 shows worse performance on harsher modifications using the given apparatus. Given the stable R^2_{adj}-values and the increase in intercept as well as slope throughout the PREDICTION TIME OFFSETS, an overall worsening in performance can be measured.

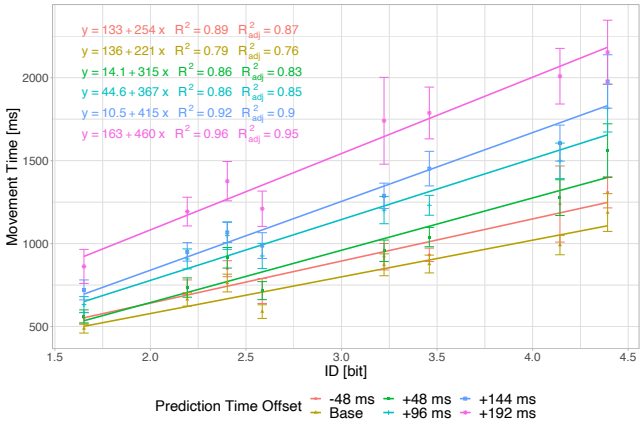

Figure 5: Linear regression models for mean movement time on all PREDICTION TIME OFFSETS and IDs.

Accuracy

As a measurement of accuracy, the mean distance of the target center to the exact point where the participant hit the targets was used.

Table 5: Bonferroni-adjusted *p*-values of pairwise Wilcoxon signed-rank tests computed on the accuracy.

	Base	+48 ms	+96 ms	+144 ms	+192 ms
-48 ms	1	.207	.034*	<.001****	.016*
Base		.291	.002**	<.001****	.005**
+48 ms			1	.004**	.144
+96 ms				.158	1
+144 ms					1

The Shapiro-Wilk tests showed the data was not normally distributed in all of the conditions (compare table 5). Again, a significant difference existed between the conditions (Friedman-Test: $\chi^2(5) = 44.90$, $p < .001$, n = 24). Conducting the same post-hoc tests as before, significant differences were found for the *-48 ms model* compared with the *+96 ms model* ($p = .034$), the *+144 ms model* ($p < .001$) and the *+192 ms model* ($p = .016$). The *Base model* also showed significant differences compared to the *+96 ms model* ($p = .002$), the *+144 ms model* (p < .001) and the *+192 ms model* ($p = .005$). The last significant difference was found between the *+48 ms model* and the *+144 ms model* ($p = .004$). Mean distance to target center was .265 unity measures ($SD = .028$).

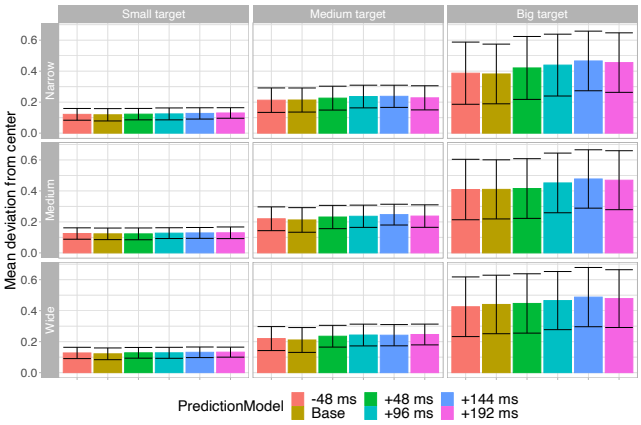

Figure 6: Mean distance from selection point to target center for all featured circle configurations with their respective standard deviation.

Figure 6 displays the accuracy more precisely over all possible circle configurations. The small targets show nearly no difference in accuracy as they are as big as the avatars fingers and don't allow for much variance in selection points. The big targets as expected show the highest variance in accuracy. Although the PREDICTION TIME OFFSETS show similar standard deviations, different means can be observed. Accuracy seems to decline for harsher prediction modifications, suggesting similar interactions as the previous measures. The apparatus, data and evaluation can be found on Github [8]

Questionnaires

During the study, the participants were asked to fill out the questionnaires described in chapter 3.4 after each condition. To analyze the questionnaire data, non-parametric tests were used. This is suggested by Gonzalez-Franco & Peck [13], who created the avatar embodiment questionnaire used in this study, as well as common practice in other related work [20, 36].

Values of the iPQ are computed by using the mean of all involved items. Results for the avatar embodiment and the used sub-scales in "body ownership", "agency and motor control" as well as "location of the body" are calculated using the equations presented by its authors [13].The motion sickness is computed using the mean of all involved items.

When evaluating the iPQ on presence a significant difference existed between the conditions (Friedman-Test: $\chi^2(5) = 20.1$, $p < .05$, n = 24). The post-hoc tests resulted in no significant differences between the conditions except for the *base model* and the *+192 ms model* ($p = .03$)

Evaluating the avatar embodiment resulted in significant differences between multiple conditions (Friedman-Test: $\chi^2(5) = 51.3$, $p < .05$, n = 24).

[8] https://github.com/andreasPfaffelhuber/Faster-than-in-Real-Time

Table 6: Bonferroni-adjusted *p*-values of pairwise Wilcoxon signed-rank tests for all questionnaires.
(presence , motion sickness , avatar embodiment , "agency and motor control" , "location of the body" , "body ownership")

	-48 ms	Base	+48 ms	+96 ms	+144 ms	+ 192 ms
-48 ms		1 / 1	1 / 1	.021* / 1	.024* / .046	<.001*** / .005**	.066 / .022*	<.001*** / .003**	.003** / <.001***
Base	1 / 1		.153 / 1	.117 / .464	.009** / 1	.016* / .06	<.002** / .018*	.011* / .018*	<.001*** / .005**	<.001*** / <.001***
+48 ms	1 / 1	1 / 1		.465 / 1	.255 / .87	.102* / .005**	.926 / 1	.07** / .005**	.001* / .026*
+96 ms	1 / 1	1 / 1	1 / .128		.92 / .003**	1 / 1	.002** / .006**	.113 / <.001***
+144 ms	.859 / .784	.321 / .274	1 / .109	1 / 1		.032* / 1	.348 / .008**
+192 ms	.071 / .149	.03* / .081	.146 / .022*	.095 / .594	.61 / 1	

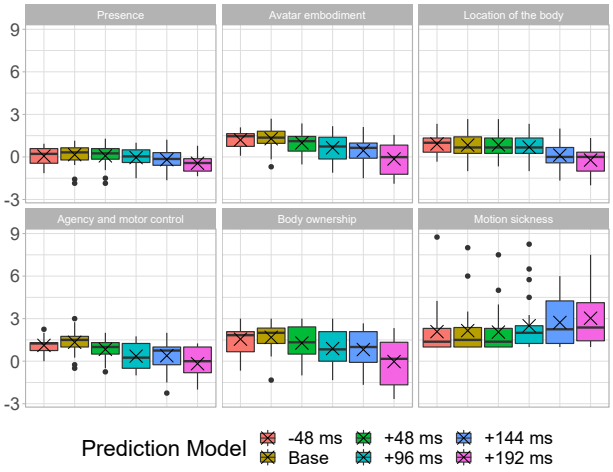

Figure 7: Function fit and mean movement time for all PREDICTION TIME OFFSETS and IDs.

When evaluating the Avatar Embodiment sub-scales significant differences could be found for "agency and motor control" (Friedman-Test: $\chi^2(5) = 47.8$, $p < .05$, n = 24), "body ownership" (Friedman-Test: $\chi^2(5) = 49.7$, $p < .05$, n = 24). as well as "location of the body" (Friedman-Test: $\chi^2(5) = 41.4$, $p < .05$, n = 24). Overall the sub-scales resemble the avatar embodiment scale. More drastic changes report worse avatar embodiment scores.

The motion sickness questions found significant differences between some groups (Friedman-Test: $\chi^2(5) = 25.7$, $p < .005$, n = 24), though only between the *+48 ms model* and *+192 ms model* ($p = .022$) was reported by the Wilcoxon signed-rank test. Overall participants didn't report effects of motion sickness regardless of the PREDICTION TIME OFFSET, though a slight increase of the means towards the harsher PREDICTION TIME OFFSETS indicates that those overstepped the boundary of subjectively pleasant motion alteration.

Qualitative Feedback

The qualitative feedback given after each condition was analyzed in order to gain further insight on how participants perceived the different PREDICTION TIME OFFSETS. Participants had only been told they were testing different tracking configurations during the study, and had answered the questions described in chapter 3.4 in a thinking aloud manner. Considering the *+96 ms model* and the even more intense prediction models, nearly all participants reported that they did not find those to function particularly well. They reported that the hand of the digital avatar was shaking as well as jumping around most of the time, that moving the avatar felt a bit uncontrollable, and that precise selections were nearly impossible to do. Those factors seemed to generally get worse using stronger PREDICTION TIME OFFSETS. As a result, most participants reported to feel that the currently tested condition was working pretty bad, and that they did not feel as if the avatar resembled their real body and its motions. The *+48 ms model*, the *base model* and the *-48 ms model* were rated much better compared to the previously described ones. Participants did usually not perceive the negative effects listed for the other models, but reported much more smooth and congruent motions. Interestingly, participants rated the *+48 ms model*, *base model* and the *-48 ms model* very similar. Many participants noted that their motions felt either somewhat slow or a little bit sped up when compared to their experience with the baseline system or the other mentioned model, but didn't agree on which model caused which effect. While some participants reported the *+48 ms model* was speeding up their motions and the past model was slowing them down, others had the exact opposite impressions. We suspect that a possible explanation for this might be that reverting the motion's direction could take longer using future prediction, since the movement would continue in the original direction for a bit even if the participants hand never did, maybe even longer than when buffering the past frames.

The base system without any prediction was the one most participants rated as their favorite condition, often stating that it offered the best direct translation of their real motions to the avatar's motions and lacked the negative effects of the future prediction.

6 DISCUSSION

In this study, employing neural networks to predict user motions using different PREDICTION TIME OFFSETS has been investigated.

Additionally, the baseline model as well as a past time model using buffered frames were used to further validate the findings and to quantitatively model the occurred prediction effect. A standardized two-dimensional Fitts' law task was conducted to compare user performance in the virtual environment. The initial expectations were that the smaller future prediction models would result in a better user performance and presence than the past and the baseline model, before those would drop off again due to the stronger prediction models over-predicting the participants motions. Statistical evaluations concerning the performance measures show the overall existence of significant differences between the different PREDICTION TIME OFFSETS. Throughput, movement time and accuracy all confirm that the used model makes a difference, yet the observed differences are not at all like initially expected. The expected improvement using future and therefore faster versions of the motions could not be observed. Instead we found a negative influence using stronger changes, with even the smallest model already negatively impacting all of the performance measures compared to the *base system*. Questionnaire data collected during the study confirms that the avatar embodiment as well as as the connected sub-scales show similar trends as the performance measures. Presence as computed by the iPQ on the other hand was not affected by the described setup and conditions. No significant difference was observed except once for the harshest motion alteration.

One possible explanation for the observed results might be that future motions generally make participants perform worse since they notice that their real-world motions are not congruent with the avatars motions observed in the virtual environment. Participants could thus intuitively slow down their motions simply because of the decrease in avatar embodiment and limb-ownership, due to the translation not being entirely correct and resulting in a difference between their expected virtual limb position and the out-coming one. Therefore, even a theoretically perfect prediction of a user's movement might not increase performance as users could be too cautious or possibly overshoot due to the prediction continuing in movement direction for a little while longer.

Another possible explanation could be that the trained prediction models did not reach the necessary amount of accuracy to really speed up user motions correctly. Shaking limbs or too inaccurate predictions of their motions also would require the participants to spend time correcting their avatar positions and could result in lower performance in the given Fitts' law task. Qualitative data captured during the experiment would at least confirm this suspicion for the worst prediction models, where participants often complained about these negative effects.

Limitations and Future Work

Given the hardware setup and the high frame-rate of the motion tracking system, the used neural network architecture was limited by computation time. This resulted in the use of a shallow architecture, which does not work on time sequences. Using more capable hardware, network architectures like recurrent neural networks (RNN) which allow for time sequences [24] might lead to a more stable motion prediction and thus better results.

Additionally, in the current setup only general prediction models were used. Training data was not specifically captured for the sitting Fitts' law task and thus the models are not optimized for the given task.

Future work might want to explore whether specifically captured data and better hardware allow more accurate models resulting in better user performance and better embodiment. Additionally, the study only tested one experimental setup where all participants conducted the task on an horizontal axis sitting at a table. Since virtual reality environments and the used motion capture system would theoretically allow for motions along other directions as well, different free standing or moving configurations not only focusing on a table setup could be evaluated as well. This might provide a better well-rounded insight instead of limiting the findings to a rather specific set of tasks.

REFERENCES

[1] Abien Fred Agarap. 2018. Deep Learning using Rectified Linear Units (ReLU). 1 (2018), 2–8. arXiv:1803.08375 http://arxiv.org/abs/1803.08375

[2] Hironori Akiduki, Suetaka Nishiike, Hiroshi Watanabe, Katsunori Matsuoka, Takeshi Kubo, and Noriaki Takeda. 2003. Visual-vestibular conflict induced by virtual reality in humans. *Neuroscience Letters* 340, 3 (2003), 197–200. https://doi.org/10.1016/S0304-3940(03)00098-3

[3] A. Deniz Aladagli, Erhan Ekmekcioglu, Ahmet Kondoz, and Dmitri Jarnikov. 2018. Predicting Head Trajectories in 360 Virtual Reality Videos. *2017 International Conference on 3D Immersion, IC3D 2017 - Proceedings* 7 (2018), 1–6.

[4] Domna Banakou, Raphaela Groten, and Mel Slater. 2013. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences of the United States of America* 110, 31 (2013), 12846–51. https://doi.org/10.1073/pnas.1306779110

[5] Matthew Botvinick and Jonathan Cohen. 2006. Rubber hands 'feel' touch that eyes see. *Nature* 391, 1 (2006), 756. https://doi.org/10.1016/j.jalz.2006.05.522

[6] Niclas Braun, Stefan Debener, Nadine Spychala, Edith Bongartz, Peter Sörös, Helge H.O. Müller, and Alexandra Philipsen. 2018. The senses of agency and ownership: A review. *Frontiers in Psychology* 9, APR (2018), 1–17. https://doi.org/10.3389/fpsyg.2018.00535

[7] Timothy Dummer, Alexandra Picot-Annand, Tristan Neal, and Chris Moore. 2009. Movement and the rubber hand illusion. *Perception* 38, 2 (2009), 271–280. https://doi.org/10.1068/p5921

[8] Edgar Erdfelder, Franz FAul, Axel Buchner, and Albert Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 4 (2009), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

[9] Franz Faul, Edgar Erdfelder, Albert Georg Lang, and Axel Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 2 (2007), 175–191. https://doi.org/10.3758/BF03193146

[10] Paul M. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[11] Peter J. Gianaros, Muth Eric R., Jonathan Toby Mordkoff, Max E Levine, and Robert M. Stem. 2001. A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness. *Aviation Space and Environmental Medicine* 72, 2 (2001), 115–119. https://doi.org/10.1038/jid.2014.371 arXiv:NIHMS150003

[12] Mar Gonzalez-Franco and Jaron Lanier. 2017. Model of illusions and virtual reality. *Frontiers in Psychology* 8, JUN (2017), 1–8. https://doi.org/10.3389/fpsyg.2017.01125

[13] Mar Gonzalez-Franco and Tabitha C. Peck. 2018. Avatar embodiment. Towards a standardized questionnaire. *Frontiers Robotics AI* 5, JUN (2018), 1–9. https://doi.org/10.3389/frobt.2018.00074

[14] Shunichi Kasahara, Keina Konno, Richi Owaki, Tsubasa Nishi, Akiko Takeshita, Takayuki Ito, Shoko Kasuga, and Junichi Ushiba. 2017. Malleable Embodiment: Changing Sense of Embodiment by Spatial-Temporal Deformation of Virtual Human Body. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* (2017), 6438–6448. https://doi.org/10.1145/3025453.3025962

[15] Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. 2013. Drumming in immersive virtual reality: the body shapes the way we play. *IEEE transactions on visualization and computer graphics* 19, 4 (2013), 597–605. http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6479188

[16] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387. https://doi.org/10.1162/PRES_a_00124

[17] Diederik P. Kingma and Jimmy Lei Ba. 2015. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015), 1–15. arXiv:1412.6980

[18] Luv Kohli. 2010. Redirected touching: Warping space to remap passive haptics. *3DUI 2010 - IEEE Symposium on 3D User Interfaces 2010, Proceedings* (2010), 129–130. https://doi.org/10.1109/3DUI.2010.5444703

[19] Huy Viet Le, Valentin Schwind, Philipp Göttlich, and Niels Henze. 2017. PredicTouch: A System to Reduce Touchscreen Latency using Neural Networks and Inertial Measurement Units. *Proceedings of the Interactive Surfaces and Spaces on ZZZ - ISS '17* (2017), 230–239. https://doi.org/10.1145/3132272.3134138

[20] Lorraine Lin and Sophie Jörg. 2016. Need a hand? How appearance affects the virtual hand illusion. *Proceedings of the ACM Symposium on Applied Perception, SAP 2016* (2016), 69–76. https://doi.org/10.1145/2931002.2931006

[21] I. Scott MacKenzie. 1989. A note on the information-theoretic basis for fitts' law. *Journal of Motor Behavior* 21, 3 (1989), 323–330. https://doi.org/10.1080/00222895.1989.10735486

[22] I. Scott MacKenzie. 1992. *Fitts' Law as a Performance Model in Human-Computer Interaction.* Ph.D. Dissertation. CAN. UMI Order No. GAXNN-65985.

[23] I. Scott MacKenzie. 2018. Fitts' Law. *Handbook of human-computer interaction* (2018), 349–370.

[24] Julieta Martinez, Michael J. Black, and Javier Romero. 2017. On human motion prediction using recurrent neural networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-Janua (2017), 4674–4683. https://doi.org/10.1109/CVPR.2017.497 arXiv:1705.02445

[25] Daniel Perez-Marcos, Mel Slater, and Maria V. Sanchez-Vives. 2009. Inducing a virtual hand ownership illusion through a brain-computer interface. *NeuroReport* 20, 6 (2009), 589–594. https://doi.org/10.1097/WNR.0b013e32832a0a2a

[26] Holger Regenbrecht and Thomas Schubert. 2002. Real and Illusory Interactions Enhance Presence. *Presence: Teleoperators and Virtual Environments* 11, 4 (2002), 425–434.

[27] Michael Rietzler, Florian Geiselhart, Julia Brich, and Enrico Rukzio. 2018. Demo of the Matrix Has You: Realizing Slow Motion in Full-Body Virtual Reality. *25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings* (2018), 773–774. https://doi.org/10.1109/VR.2018.8446136

[28] Michael Rietzler, Florian Geiselhart, Jan Gugenheimer, and Enrico Rukzio. 2018. Breaking the Tracking: Enabling Weight Perception using Perceivable Tracking Offsets. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (2018), 1–12. https://doi.org/10.1145/3173574.3173702

[29] Emad W. Saad, Thomas P. Caudell, and Donald C. Wunsch. 1999. Predictive head tracking for virtual reality. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)*, Vol. 6.

[30] Maria V. Sanchez-vives and Mel Slater. 2005. From Presence Towards Consciousness. *Nature Reviews Neuroscience* 6, 10 (2005), 332. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.5596{&}rep=rep1{&}type=pdf

[31] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 1999. Embodied Presence in Virtual Environments. *Visual Representations and Interpretations* (1999), 269–278. https://doi.org/10.1007/978-1-4471-0563-3_30

[32] Thomas Schubert and Holger Regenbrecht. 2002. Wer hat Angst vor virtueller Realität ? Angst , Therapie und Präsenz in virtuellen Welten Angst , Therapie und Präsenz in VR. *Virtuelle Realitäten* (2002), 225–274.

[33] Thomas Schubert, Holger Regenbrecht, and Frank Friedmann. 2001. The experience of presence: Factor analytic insights. Presence: Teleoperators and Virtual Environments. *Presence: Teleoperators and Virtual Environments* 10, 3 (2001), 266—-281. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.69.3630

[34] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using presence questionnaires in virtual reality. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–12. https://doi.org/10.1145/3290605.3300590

[35] Valentin Schwind, Jan Leusmann, and Niels Henze. 2019. Understanding visual-haptic integration of avatar hands using a Fitts' law task in virtual reality. *ACM International Conference Proceeding Series* (2019), 211–222. https://doi.org/10.1145/3340764.3340769

[36] Valentin Schwind, Lorraine Lin, Massimiliano Di Luca, Sophie Jörg, and James Hillis. 2018. Touch with foreign hands: The effect of virtual hand appearance on visual-haptic integration. *Proceedings - SAP 2018: ACM Symposium on Applied Perception* (2018). https://doi.org/10.1145/3225153.3225158

[37] Sotaro Shimada, Yuan Qi, and Kazuo Hiraki. 2010. Detection of visual feedback delay in active and passive self-body movements. *Experimental Brain Research* 201, 2 (2010), 359–364. https://doi.org/10.1007/s00221-009-2028-6

[38] R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a standard for pointing device evaluation , perspectives on 27 years of Fitts ' law research in HCI. *International journal of human-computer studies* 61, 6 (2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001

[39] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15, 1 (2014), 1929–1958. https://doi.org/10.1016/0010-4361(73)90803-3

[40] Manos Tsakiris, Gita Prabhu, and Patrick Haggard. 2006. Having a body versus moving your body: How agency structures body-ownership. *Consciousness and Cognition* 15, 2 (2006), 423–432. https://doi.org/10.1016/j.concog.2005.09.004