The Effects of Full-Body Avatar Movement Predictions in Virtual Reality using Neural Networks

Valentin Schwind
Frankfurt University of Applied Sciences
Frankfurt, Germany
valentin.schwind@acm.org

David Halbhuber
Jakob Fehle
Jonathan Sasse
Andreas Pfaffelhuber
Christoph Tögel
Julian Dietz
University of Regensburg
Regensburg, Germany
{firstname.lastname}@stud.uni-regensburg.de

Niels Henze
Media Informatics Group
University of Regensburg
Regensburg, Germany
niels.henze@ur.de

ABSTRACT

Motion tracking technologies and avatars in virtual reality (VR) showing the movements of the own body enable high levels of presence and a strong illusion of body ownership (IBO) – key features of immersive systems and gaming experiences in virtual environments. Previous work suggests using software-based algorithms that can not only compensate system latency but also predict future movements of the user to increase input performance. However, the effects of movement prediction in VR on input performance are largely unknown. In this paper, we investigate neural network-based predictions of full-body avatar movements in two scenarios: In the first study, we used a standardized 2D Fitts' Law task to examine the information throughput in VR. In the second study, we utilized a full-body VR game to determine the users' performance. We found that both performance and subjective measures in a standardized 2D Fitts' law task could not benefit from the predicted avatar movements. In an immersive gaming scenario, however, the perceived accuracy of the own body location improved. Presence and body assessments remained more stable and were higher than during the Fitts' task. We conclude that machine-learning-based predictions could be used to compensate system-related latency but participants only subjectively benefit under certain conditions.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; **Virtual reality**; *User studies.*

KEYWORDS

Virtual reality; presence; questionnaire; break in presence.

VRST '20, November 1–4, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-7619-8/20/11…$15.00
https://doi.org/10.1145/3385956.3418941

ACM Reference Format:
Valentin Schwind, David Halbhuber, Jakob Fehle, Jonathan Sasse, Andreas Pfaffelhuber, Christoph Tögel, Julian Dietz, and Niels Henze. 2020. The Effects of Full-Body Avatar Movement Predictions in Virtual Reality using Neural Networks. In *26th ACM Symposium on Virtual Reality Software and Technology (VRST '20), November 1–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3385956.3418941

1 INTRODUCTION

Motion tracking technologies and avatars in virtual reality (VR) that show the movements of the own body enable high levels of presence and the illusion of body ownership (IBO) – key features of immersive systems and gaming experiences in virtual environments. Strong presence and the IBO is perceived when a sense of full-body ownership [14, 43, 45] and agency [23, 35, 48] allow users to correctly locate their body pose within the virtual environment. When users move their own limbs through active motor control, the brain's expected positions must match the perceived sensory afferent modalities such as gaze, haptics, or body proprioception [7, 14, 41, 51]. Consequently, lacking temporal synchronicity [44] or spatial congruence [8] between the real and the virtual movements can cause conflicting cues from the visual and vestibular afferent and, thus, VR motion sickness or postural instability [1].

To prevent delays between the real and virtual body movements sophisticated tracking systems seek for low to zero latencies. Despite technological advances in hardware that detects movements precisely and quickly, a residual latency remains that can only be compensated by software to provide real-time tracking. However, software-based prediction algorithms are not only capable of estimating a user's recent pose [30] and to remove delays caused by the system [39] but also to go beyond the latency of the hardware and to predict where one's own body will be in an upcoming time step beyond the recent position [15, 27]. However, the effects of machine learning (ML)-based algorithms predicting real-time or future movements of the own body particularly in the context of time critical input performance or immersive experiences are currently unknown. Previous research of linearly extrapolated movements indicates that induced a "lighter weight" sensation [22], however, objective findings about using neural networks able to predict the own movements in VR or input measures quantifying the input performance in a gaming scenario are currently unknown.

An investigation of how movement predictions in VR change the perception of people is particularly important for the design of systems that are on the one hand designed to improve the users' input performance, but on the other hand not at the expense of an immersive experience hindered by the asynchronicity between the real and virtual body. Thus, the research question is, which prediction time is the best tradeoff between the increased performance due to e.g. accelerated body parts and the negative effects caused by a reduced IBO . The use of ML-based algorithms to calculate future movements seems feasible and to be particularly promising to answer the research question. Using ML-based algorithms to compute future movements is widely acknowledged in human-computer interaction (HCI) research [15, 16, 27] and particularly promising to answer that question.

In this paper, we investigate the effects of ML-based prediction of avatar body movements in VR on human input performance in two studies: In the first study, we used a standardized 2D Fitts' Law task to examine predicted as well as delayed (from −50 *ms* to +200 *ms*) human body movements in VR on information throughput. In the second study, we utilized a realistic full-body VR game to determine the gaming performance when avatar body movements are being predicted using the same conditions as in our first study. We found that both performance and subjective measurements of the participants in a standardized 2D Fitts' law task could not benefit from the predicted avatar movements. In an immersive gaming scenario, however, the perceived accuracy of the own body location improved. Presence and body assessments remained more stable and were higher than during the Fitts' task even at higher prediction times. We contribute with an approach for avatar-based full-body movement prediction using ML and conclude that predicted movements could be used to compensate system-related latency, but only subjectively benefit in certain scenarios with future predictions.

2 RELATED WORK

Algorithms able to predict the users' movements are often used to compensate system-based latencies. Thus, our work is related to previous work compensating latencies and predicting movements on real devices first. We then discuss related approaches to predict body movements in VR.

2.1 Latency and Software-Based Predictions

Latency denotes an inevitable delay between a stimulus and the corresponding response of a system. Thus, latency also describes the responsiveness of a system that actually is supposed to react in real-time [29, 36, 49]. A widely recognized test to measure the effects of system-based latency on human motor control goes back to Paul Fitts [10]. Using Fitts' law paradigm [28] previous work repeatedly showed that very small amounts of latency are noticeable and can reduce a user's performance [20, 29]. For example, MacKenzie and Ware [29] found how delays negatively affect human input performance and measured a continuous performance degradation from 8.33 *ms* to 225 *ms* to a total of 63.9% increased movement time and 214% higher error rate compared to the zero lag condition.

However, the perception and the objective effects of latencies depend on the task. For example, Pavlovych and Stuerzlinger [36] showed that latency of real-time systems is able to negatively affect user performance in a 2D target following task. The authors found that errors increased for latencies of over 110 *ms*, for latency jitters above 40 *ms*, and for dropout rates of more than 10%. Moreover, Jota et al. [20] found lower thresholds for detecting latency while dragging than for tapping. Using a stylus humans are even able to discriminate latency differences of ca. 1 *ms* [33]. Researchers have also shown that perception of latencies changes other properties of the system. For example, users perceive buttons with longer delays as being heavier, with a need for greater force when pressing [21].

As even very low latencies are noticeable and negatively affect the performance, researchers proposed different approaches to compensate latency. Such approaches can be categorized into hardware-based [5, 33, 34] and software-based [15, 16, 33] ones. While hardware-based approaches heavily depend on the technology, software-based approaches can be readily deployed into off-the-shelf applications and adapt to delays caused by the hardware [27]. While early application use linear extrapolation [50] recent work predicts movements using artificial neural networks (ANNs) and achieve higher throughputs and faster reaction times than classical approaches [15, 16].

Movement Predictions in VR

End-to-end tracking delays between the rendering of a 3D environment and one's own head movement have dramatic consequences on the users' perception and their well-being. They produce vestibular conflicts that result in VR motion (or cyber) sickness, degraded or oscillating vision, and reduced input performance [3, 25, 26]. For example, Meehan et al. [32] compared end-to-end latencies using a head-mounted display (HMD) with delays between 50 ms and 90 ms. The authors showed that higher latency results in a lower sense of presence and weaker physiological fear/stress responses. Findings by Ellis et al. [9] indicate that perceptual stability across virtual environments requires latency of an HMD less than 16 *ms*.

Not only head movements, but also avatars are subject of research investigating the effects of delayed movements on the IBO . A strong IBO is perceived when full-body ownership [14, 43, 45] allow users to correctly locate their whole body pose within the virtual environment. For example, Waltemate et al. [49] examined the impact of latency on perceptual judgments and motor performance in closed-loop interaction in VR using avatars and found that more complex movements rather induce a strong sense of ownership "than a simple button press" [49], even despite high latencies of the visual feedback. The authors assume that participants rather rely on correlation between the temporal structure of the motor and visual signals to infer a common cause. Interestingly, findings by Imaizumi and Asai [4, 18] indicate that the sense of ownership might be more affected by latencies than the sense of agency.

Due to the immense consequences of VR motion sickness, researchers and developers seek to reduce delays and to predict head [2, 39, 50] and body [22, 31, 42] movements to improve subjective experiences and objective measures. Wu and Ming demonstrated that compensating delays using head movement extrapolation techniques in VR can improve human performance in spatial tasks [50]. The authors used linear extrapolation functions to predict HMD movements and found that humans performed significantly better than without latency compensation. Based on their work, Kasahara et al. [22] estimated past and future body positions using a time-based linear extrapolation model where the amount of temporal shift could be adjusted using a continuous time value. The virtual human avatar movement model generated ranged from 25 to 100 ms in the future and subjectively induced a "lighter weight" sensation [22]. ANNs were used to predict head movements and to calculate the translation of a virtual camera in an upcoming frame [39] as well as for inverse kinematics (IK) end-effector movements [37].

2.2 Summary

Real-time systems often suffer from latency [20, 28, 29, 36, 49]. While hardware-based latencies [5, 33, 34] are difficult to realize and cannot predict future movements, software-based algorithms are being used to improve the interaction with real devices and to increase input performance [15, 16, 33]. Considering VR, delays can cause forms of sickness [3, 9, 25, 26, 32] and the prediction of body movements is being used to overcome such delays [39, 50]. Previous work also uses classical software-based approaches to estimate future movements [22]. While ANNs are used to predict head and end-effector translation [37], the effects of full-body-related predictions using ANNs and evaluations of system with embodied users, gauging how latency compensation and future predictions are actually perceived by the user, are currently unknown.

3 DATA COLLECTION AND MODEL

Similar to previous work [15, 27], this research follows a data-driven approach to explore the effects of machine full-body avatar movements prediction in VR using ANN:

(1) Gathering the data set: A preliminary data acquisition study was conducted to collect the data necessary for training an ANN. Participants were instructed to perform natural motions. Motions of the participants were recorded using common motion capturing technologies.
(2) Model and system development: ML-based algorithms were trained based on the data set and tested for real-time capability to investigate the feasibility of motion prediction of movement in real-time. We optimized the model parameters to achieve the highest accuracy on the test set and speed of the prediction of the client software.
(3) Evaluation of the models in two user studies: a standardized task according to Fitts' law and a fully-body VR game. As optimizing purely for the test set increases the probability of introducing model overfitting, we evaluate the generalization of our model with a validation set.

To enable future work to improve our results based on steady advances in machine learning research and specialized models, we publicly released our data set (see end of this paper).

3.1 Full-Body Motion Data Collection

A preliminary data acquisition study was conducted to collect the data necessary for training and understanding of an ANN capable of predicting motions in real-time.

3.1.1 Apparatus. To track full-body motion, we used an OptiTrack motion tracking system with twelve cameras (8 PRIME 13 and 4 PRIME 13W) mounted on a traverse and covering a 4.2 m x 3.9 m tracking volume. We calibrated the OptiTrack system according to the manufacturer's specification and achieved an "exceptionally precise" calibration result with an overall reprojection error of .853 mm. Participants wore a marker-based full-body suit with 49 markers. The motion tracking software was running on a dedicated PC with Windows 10, Intel i7-8700, 26GB RAM, and a NVIDIA GeForce GTX 1080 with 8 GB RAM graphics card. The boundary of the tracking volume was highlighted with white stripes on the floor. To provide the best motion tracking using our hardware, takes were recorded using OptiTrack Motive 2.2 with the maximal framerate of 240 Hz. Thus, 1 frame corresponds to 4 ms.

3.1.2 Procedure and Tasks. After signing a consent form, participants put on one of three motion capturing suits in their corresponding body size (S, M, L). Skeleton of each participant was created and calibrated in T-pose. Participants were orally instructed to perform 20 short tasks for ca. 30 seconds using varying directions, speeds, and limbs. The tasks were designed to detail a broad range of motion of the human body including natural and wide variation of movements. The tasks were: (1) waving with different hands, (2) miming a sport they like, (3) looking at hands, (4) putting hands on the back of the head, (5) walking comfortably in a circle, (6) turning left, (7) turning right, (8) walking forward and back, (9) walking back and forward, (10) circling the arms, (11) throw something from below, (12) throw something from above, (13) hands on hips, (14) stretching hands, (15) one step forward and back again (forwards, backwards, left, right) (16) boxing (both hands), (17) jumping, (18) small leap forward, (19) walk comfortably in a circle (3 × ccw), (20) walking on a line (foot by foot). Task sequence was randomized for each participant.

3.1.3 Participants. We recruited 20 participants (9 female, 11 male) via mailing lists and compensated them with one credit point for their study course. Their average age was 24.4 years ($SD = 4.0$) and ranging from 19 to 30 years. All of them were right-handed. Participants were not informed about the exact purpose of the study (development of motion prediction models) that participants move as naturally as possible.

3.2 Model and System Development

We present our data set and describe three steps towards developing finger identification models: (1) pre-processing the data set, (2) exploring and using the data to train deep learning models for predicting full-body motions, and (3) development of a client software.

3.2.1 Data Pre-Processing. We recorded a total of 4233916 frames (294.02 mins) with a mean of 211695 (14.7 mins) frames per participant. Data of the skeleton movements were visually inspected to find potential errors. In some cases (ca. 15 of 100,000 frames), no values for bone skeleton position or rotation data were recorded and skipped by the OptiTrack streaming protocol. Missing frames in the recording were reconstructed using linear interpolation between two samples. Data in the recordings (meta-data, frame sequence, header data, ...), which were not processed by the OptiTrack Client implementation, were removed. As the individual rotation of the finger limbs have been interpolated by the OptiTrack software, finger movement were not considered in further training. Rotation of the remaining 21 joints of the human skeleton provided by OpitTrack were processed using quaternions. As position data, except the translation of the animation skeleton root, were omitted, the training of the ANNs was, in sum, performed on 87 input values.

3.2.2 Stream Interceptor Software. To test the real-time capabilities of the ANNs and to externally validate the prediction models, we developed a native Python application to intercept the OptiTrack stream and to replace the data with the output of the ANN. The software performed several time-critical tasks: (1) unpacking the UDP stream provided by OptiTrack, (2) feeding the raw data into the loaded TensorFlow model, (3) waiting for model interference and accepting the prediction, and (4) piping the output of the model into the UDP stream. The software ran on the motion capturing server to relieve the client running the real-time application in Unity.

As each packet in the stream contained a single frame, the intercepting client had to process each packet individually. The motion capturing system ran at a frequency of 240 Hz, which means that every 4.167 ms a client had to process the packet completely and to forward it to the VR application before the frame would have been discarded by the OptiTrack plugin due to buffer overflows. Discarded packages would have caused the target application to tremble limbs of the avatar representation. We evaluated the prediction time of various network architectures *using* the intercepting client. The PCs were connected via 1 GBit network connection.

3.2.3 System Latency Analysis. The marker-based OptiTrack motion tracking system Motive captures the movement of an object or the human skeleton and broadcasts the data to one or more external computers. Motive's functionality of broadcasting motion data is based on the NatNet streaming protocol[1]. The data is being received in Unity[2] using the OptiTrack plug-in (version 1.2) provided by the manufacturer[3]. The data stream can be received by any machine within a local network and it is possible to modify the data, repackage it, and forward it again to the network and to the Unity VR render engine running on an external PC.

However, the entire data transfer of the system, including computation, processing, and display of full-body motion data also suffers from internal latency. In order to develop an ANN able to predict motions upon the system-related latency, it is necessary to determine the time it took for the system to loop physical information through the entire system. We used a latency test framework (LTF) based on an Arduino microcontroller coupled to a vibration and a photosensitive sensor determining the delay between a physical in- and output. The initial timestamp of system delay was determined after registering the vibration of the sensor triggered by a tracked rigid body object. The virtual collision via OptiTrack has been determined by the Unity client receiving the intercepted stream. Further, the Unity application registered the object collision and light up the rendered view within the HMD by HTC Vive to indicate signal change. Increasing light intensity has then been registered by the photo sensor and a second timestamp. Using the LTF, the difference between both timestamp and the delay of the system infrastructure has been determined multiple times ($N = 120$) with a mean of 51 ms ($SD = 7$) which corresponds to a prediction of 12 frames into the "future".

3.2.4 ANN Training and Model Development. Even when the problem of motion prediction cannot be considered as a real classification problem, we used a customized Deep Neural Network (DNN). A crucial factor in choosing the right ANN architecture was the time required for predicting the next frame. As the prediction had to be computed in real-time, it was necessary to minimize the interference time. To achieve real-time computation, we designed the network to be as simple as possible. In contrast to classical and conventional implementation of DNNs, the network in this paper does not use the SoftMax function, since a categorical representation of the probability distribution of the output yields no benefit.

A network consists of 87 input neurons from the recorded raw motion data were passed to the first of two hidden layers. The first hidden layer contains 4096 units and is fully connected to the second hidden layer with 8192 units. A dropout function has been added to the second hidden layer [17, 46] to prevent the network from overfitting. The dropout rate was chosen to be deliberately low (20 %) to avoid slipping in the under-learning range. The output layer follows subsequently with built-in ReLu activation function [12]. For stochastic optimization we used a variant of the ADAM [24] optimizer, which is included in TensorFlow.

The training process was initiated with a learning rate of 0.001 and a batch size of 512 samples. By using Keras callback functions [6], the learning rate could be dynamically adjusted during the training process. Depending on the validation accuracy, the learning rate was either increased or decreased up to a fixed threshold value. This process is comparable to classical learning rate decay in stochastic gradient descent method implemented in TensorFlow [47], but the herein presented implementation showed to be much more effective in dealing with optimization plateaus. The loss has been fitted via computing the mean square error (MSE) as shown in Formula 1.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2. \quad (1)$$

First tests with the model showed that the neural network was able to solve the given problem, however, the accuracy did not diverge clearly against a value, which was caused by the implemented dropout function in the last hidden layer. To fix this problem, we implemented a Keras callback function which allowed the model to cancel the training process if performance deteriorated. In such cases, the previous version of the model was loaded and saved. This procedure is called *Early Stopping* (ES).

3.2.5 Model Selection. Considering the computation time of the client alone (1.42 ms), the time for model interference had to be below a duration of 2.747 ms (cf. *Stream Interceptor Software*). We evaluated the prediction time of common network architectures: a classical DNN structure ($M = 2.543$, $SD = .941$), a classical recurrent neural network (RNN) structure ($M = 5.136$, $SD = 2.107$), an RNN based on CudnnLSTM units ($M = 4.055$, $SD = 1.846$), and an RNN

[1] https://optitrack.com/products/natnet-sdk/
[2] https://unity.com
[3] https://optitrack.com/unity-integration/

based on CudnnGRU units (M = 3.492, SD = 1.466). Only the classical DNN structure could further be considered as solver, as all other architectures did not reliably remain below the 2.747 ms interference time limit within our system.

Using that DNN structure, we built various models that differ in their prediction time. The accuracy of the prediction decreases the further the motion forecast lies in the future. Inaccuracies in the prediction are noticeable through tremor of the own limbs, which would – transferred to fully-body motions in VR – no longer enable any interaction or illusion of limb ownership. Thus, we initially explored different prediction times using a VR prototype starting from 100 frames (416 ms) with an accuracy of 37.2% while using 90% of the data for training and 10% for internal validation. As the accuracy was too low to enable any interaction with the VR prototype, we successively reduced the prediction time to achieve an adequate tradeoff between prediction and accuracy. Acceptable accuracies for interaction were found to be above 85%.

To determine the effects of avatar motion prediction in VR, we used four time intervals starting with 12 frames (zero latency) to be tested in our user studies. Thus, final prediction times were 12 frames (50 ms) with an accuracy of 94.2%, 24 frames (100 ms) with an accuracy of 91.4%, 34 frames (150 ms) with an accuarcy of 89.5%, and 48 frames (200 ms) with an accuracy of 86.6%. Position and orientation of the virtual head (or headset) were not predicted to avoid any motion sickness. All studies received ethics clearance according to the ethics and privacy regulations of our institution and, thus, follow the policies of our country and funding body.

4 STUDY 1: FITTS' LAW TASK IN VR

Subject of the herein research is to determine the effects of predicting a user's motion using avatars in VR. As already indicated, a widely acknowledged model of human motor control was proposed by Paul Fitts [10]. We determined in our first study the effects of motion prediction using ANN on the impact of motion prediction and human motor control in a 2D Fitts' law task [28] in VR.

4.1 Study Design

We conducted a VR user study to test the hypothesis that different Prediction Times will have an effect on user performance (throughput), presence, and body ownership. We used Prediction Time as within-subject variable. We expected to increase those measures while compensating the system latency and then to decrease them with higher Prediction Times. In addition to the four ANN models (*+50 ms*,*+100 ms*,*+150 ms*,*+200 ms*), we also tested a *0 ms* condition without prediction as baseline. To understand how the measures behave while adding latency, we added a *-50 ms* condition, streaming the skeleton data from 12 frames in the past. The six Prediction Times have been evaluated using a 6×6 balanced Latin square design.

We collected data about input performance, presence, and body ownership. Effective throughput as measure of input performance has been computed based on the duration between two targets and the position data of the target selection (see Data Analysis). To measure presence, we used the igroup presence questionnaire (IPQ) [38, 40] with the subscales general presence (GP), involvement (INV), realism (REAL), and spatial presence (SP). The IBO has

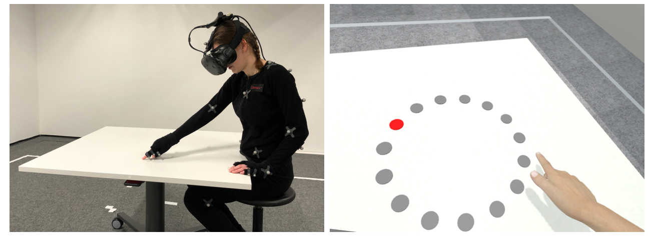

Figure 1: Participant in a full-body motion capturing suit performing the Fitts' law task.

been quantified using the avatar embodiment questionnaire by Gonzalez-Franco and Peck [13] with the subscales *body ownership*, *agency/motor control*, and *location* of the own body. To learn if prediction of one's own motions can cause motion sickness, we used the motion sickness assessment questionnaire (MSAQ) scale [11]. To gain more insights about the subjective experience, we conducted then a semi-structured interview based on six questions: (1) How did your body feel during this condition? (2) What did you notice during this condition? (3) Was this experience positive or negative for you? (4) Did you notice any differences in regards to the other iterations? (5) Were you able to chose your targets precisely? If not why. (6) How did the digital avatar feel to you? Thematic analysis was used to identify common themes within the feedback.

4.2 Apparatus and Procedure

The participants performed the task according to Fitts' law while sitting close to a squared table (1m×1m). The table functions as haptic feedback when hitting targets. Table was repositioned with its height to 76 cm above the floor, resembling common desk heights. The virtual task was displayed on a virtual representation of the table. Participants were told to sit upright to keep an equal distance to the targets throughout the study's duration. We used the HTC Vive HMD with wireless adapters to provide the VR experience. Target frame rate was set to 60 frames per second (fps). The Unity (2019.1.15f) application rendered real-time soft shadows allowing easier spatial orientation when selecting targets. We used an androgynous human-like virtual avatar created in DAZ3D to render the own body.

Participants were informed about the purpose (testing different motion tracking algorithms) but were blind to the conditions of the study. After signing the consent form, participants put on the motion capturing suit according to their size. We calibrated the participant's skeleton using the OptiTrack software. After taking a seat, the participants put on the HMD. Then, we started the Unity application with one of the six conditions and participants were asked to perform the Fitts' task. Each of the 16 trials of the Fitts' law task has been repeated twice. After each condition, participants were asked to remove their headset and to answer the questionnaires. The questionnaire items were presented in randomized order. Subjective experience items were asked verbally.

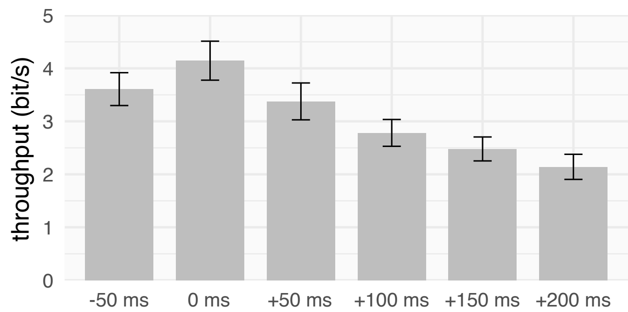

Figure 2: Mean effective throughput (TP_e) measures during the Fitts' law task for each condition. Error bars show 95% confidence interval (CI95).

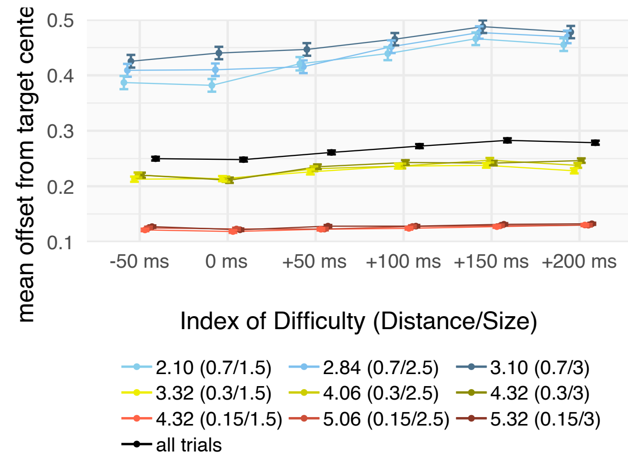

Figure 3: Mean distances from each selection point to the target center for all featured IDs indicate a decreased precision with increased prediction times. Distance and target size are in *cm*. Error bars show CI95.

4.3 Participants

We invited 24 participants (11 females, 13 males) via mailing lists. The participants mean age was 23.2 year ($SD = 3.24$) ranging from 19 to 31 years. Two participants were wearing contact lenses, 4 took their glasses off or were comfortable with wearing them underneath the HMD. All participants were students and compensated with credit points for their study course. None of the participants reported having previous experience using virtual reality devices and none of them had taken part in our data collection study.

4.4 Results

We determined the effective throughput (TP_e) using the model proposed by MacKenzie and Buxton [28]. The model provides an improved link to information theory, better fits, and index of difficultys (IDs) that cannot be negative (*cf.* ISO 9241-411 [19]). To compare the different Prediction Times we used the mean of the trials to get a single value for each participant per condition. No filtering of the data has been applied.

On average, participants spent 41.7 *mins* ($SD = 24.2$) in completing the Fitts' law part of the experiment in VR. The average completion time of the whole experiment was about 75 *mins*. Each participant performed a total of 1728 target selections. Effects of the six conditions were tested using one-way repeated measures analysis of variances (RM-ANOVAs). All tests were employed at a significance level of 5 percent ($\alpha = .05$). As Shapiro-Wilk tests on condition level indicated that the assumption of normal distribution ($p > .05$) has been violated for all measures, we performed Friedman tests for nonparametric data.

4.4.1 Throughput. The average throughputs as performance measures between the conditions are shown in Figure 2. Friedman tests showed an effect of Prediction Time, which could be confirmed, $\chi^2(5) = 84.79$, $p < .001$. Bonferroni corrected pairwise Wilcoxon signed-rank tests were performed to determine between which conditions significant differences occurred. Significant differences ($p < .007$) were found between all pairs, except for *-50 ms* and *+50 ms* ($p = 1$), *+100 ms* and *+150 ms* ($p = .131$), and *+150 ms* and *+200 ms* ($p = .436$).

4.4.2 Movement Time. A Friedman analysis of variance (ANOVA) revealed a significant effect, $\chi^2(5) = 78.41$, $p < .001$. Pairwise post-hoc comparisons showed that there were significant differences between all measures ($p > .05$), except for *-50 ms* and *+50 ms* ($p = 1$), *+100 ms* and *+150 ms* ($p = .131$), as well as between *+150 ms* and *+200 ms* ($p = .436$).

4.4.3 Movement Precision. To learn more about the quality of the predictions, we analyzed the mean distance from the target center as precision measure. We found significant effects effects, $\chi^2(5) = 44.90$, $p < .001$, and post-hoc tests revealed significant differences between *-50 ms* and *+100 ms* ($p = .034$), *+150 ms* ($p < .001$) and (+200 ms) ($p = .016$). The *0 ms* base line showed significant differences compared to the *+100 ms* ($p = .002$), the *+150 ms* ($p < .001$), and the *+200 ms* condition ($p = .005$). Statistically significant differences were also found between *+50 ms* and *+150 ms* ($p = .004$). Means of the precision measures considering the initial IDs given by target distance and target size are shown in Figure 3.

4.4.4 Presence. Multiple ANOVAs on main and subscales of the IPQ were performed and revealed significant effects for the main score, $\chi^2(5) = 13.701$, $p = .018$, general presence subscale, $\chi^2(5) = 12.323$, $p = .03$, spatial presence subscale, $\chi^2(5) = 18.566$, $p < .001$, and realism subscale, $\chi^2(5) = 12.323$, $p = .03$. No effects were found on the involvement subscale, $\chi^2(5) = 4.814$, $p = .438$. Pairwise post-hoc comparisons of the main score, the realism subscale, and the spatial presence subscale could not reveal between which conditions significant the effects on those scales occurred (all with $p > .05$). However, we found significant differences between *-50 ms* and *+200 ms* ($p = .03$), between *0 ms* and *+200 ms* ($p < .001$), as well as between *+50 ms* and *+200 ms* ($p = .016$) on the general presence subscale. Means and 95% confidence interval (CI95) of all presence measures can be found in Figure 6 (blue line).

4.4.5 Illusion of Body Ownership. The illusion of body ownership has been measured using three subscales suggested by Gonzalez-Franco and Peck [13]. Significant effects were found for all scales: *body ownership* $\chi^2(5) = 48.883$, $p < .001$, *agency/motor control*, $\chi^2(5) = 53.981$, $p < .001$, and *body location*, $\chi^2(5) = 51.515$, $p < .001$. Pairwise post-hoc comparisons revealed significant differences between *-50 ms* and *+200 ms* ($p < .001$) *0 ms* and *+200 ms* ($p < .001$), as well as *+50 ms* and *+200 ms* ($p = .001$) on the *body ownership* scale. For *agency/motor control* the mean significantly differed between *-50 ms* and *+100 ms* ($p = .02$), *-50 ms* and *+36 ms* ($p = .03$), *-50 ms* and *+200 ms* ($p < .001$), *0 ms* and *+100 ms* ($p < .001$), *0 ms* and *+150 ms* ($p < .001$), *0 ms* and *+200 ms* ($p < .001$), as well as *+50 ms* and *+200 ms* ($p < .001$). Body Location significantly differed between *-50 ms* and *+150 ms* ($p = .028$), *-50 ms* and *+200 ms* ($p < .001$), *0 ms* and *+200 ms* ($p < .001$), *+50 ms* and *+200 ms* ($p < .001$), as well as *+100 ms* and *+200 ms* ($p = .007$). Means and CI95 of body ownership measures can be found in Figure 6. The negative effects of predictions of one's own body movements on performance and presence increase with the uncertainty of ANN, which becomes apparent through the trembling of one's own limbs.

4.4.6 Motion Sickness. Effects on the perceived motion sickness were found, $\chi^2(5) = 31.553$, $p < .001$, post-hoc tests, however, were not able to identify between which of the conditions significant differences occurred (all with $p > .61$). The increasing mean scores of motion sickness indicate (see Figure 6) that motion sickness increases with increased prediction times.

4.4.7 Qualitative Feedback. The participants were able to provide subjective comments after each condition. We went through the comments to identify attributes that have been associated with the conditions. While the *0 ms* baseline condition was described to be "fluid" (P10), "accurate" (P13), "normal" (P14), or "precise" (P16), participants found that the *-50 ms* condition was more likely "lagging" (P8, P19), "slower" (P16, P20), or "hard" (P3). With increased prediction times, the participants reported of "jittering" (P20 at *+50 ms*, P13 at *+150 ms*), "flipping" (P22 at *+150 ms*), "confusing" (P23 at *+150 ms*), or "annoying" (P7 at *+200 ms*). Nevertheless, a low prediction time (*+50 ms*) also seemed to have positive effects on the experience such as to be "faster" (P19), "more fluid" (P14, P20), and "well" (P12, P23). Participants had the impression that their movements were "faster" (P5) and followed their real movements "better" (P7).

4.5 Discussion

In our first study, we investigated the effects of ML-based movement prediction of an avatar on the input performance in a Fitt's law task in VR. Predicting the limb rotation up to 200 ms such as adding delay (-50 ms) negatively affects the throughput, presence, and illusion of limb ownership. Higher prediction times even cause motion sickness. Predicting one's body movements inhibits any congruence between the physical and the virtual body, which is required for visuomotor integration and still takes place in the present. The negative effects of predictions of one's own body movements on performance and presence increase with the uncertainty of ANN, which becomes apparent through the trembling limbs.

Figure 4: Screenshot of the full-body VR game from the experimenter's perspective with a player's avatar, the arena cage, and a hostile oversize wasp approaching from one of four cage entries.

Precision measurements during the Fitts' Law tasks show that rather inaccuracies during the prediction than unplanned virtual movements that precede the physical ones influence the results. Nevertheless, the results still allow no generalization as the task is mainly limited to the movement of the right arm and requires physical feedback, which is rarely available in VR applications. Full-body movements or an immersive real-time application that does not require physical feedback have not been considered so far. Due to the limitations that result from the Fitts' task, we developed a game that more closely simulates the requirements of an immersive experience, does not happen while being seated, and involves all limbs of the body in the interaction.

5 STUDY 2: VR GAME

We developed a VR game that utilizes full-body movement as basic gaming mechanism. Thus, free avatar movements of the player should provide an immersive experience and a mean to fight hostile entities. This enables the introduction of a score that can be operationalized as an objective instrument to measure the performance of users whose virtual movements are being predicted. Aim is to determine the generalizability of findings from the first study and to create a use case that directly implements whole-body movement predictions as gaming mechanism.

5.1 Study Design

Similar to our first validation study, a one-factorial within-subject study design has been carried out to determine the effects of PREDICTION TIMES. We measured the gaming performance of our subjects indicated by the players' score reflecting their ability to hit hostile entities. Sequence of the conditions was ordered via balanced Latin square. We asked the same subjective measures (IPQ, IBO , MSAQ) as in the previous study.

5.2 VR Game and Procedure

In the Unity-based VR game, the player is situated in a virtual cage and attacked by hostile entities spawned in a surrounding jungle environment. The cage indicates the tracking volume of our

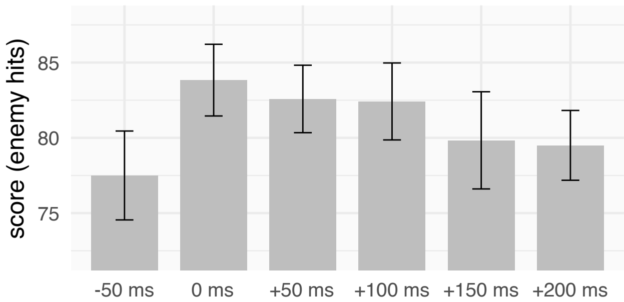

Figure 5: Mean values of the VR game scores for each prediction time. Error bars show CI95.

laboratory, in which the player can move freely around. The cage has four open doors that allow hostile oversized wasps to enter the cage and to attack the player. The players' objective during the game was to hit attacking wasps using their hands before a wasp can reach the avatar's torso. A sting hit was indicated by a short flashing red vignette in the field of view. To keep the player during each condition engaged, we increased the pace by decreasing the enemy spawn time by 90% every 3 seconds within a range from 4 sec to 1 sec. In total, 100 hostile wasps flew directly to the player so that the target play time (140 sec) could only vary slightly due to changes of the own position. If a player successfully hit a wasp, she or he were getting rewarded with one point and the wasp disappeared. Avatar appearance of the player was the same as in our first study. Figure 4 shows a screenshot of the game from the experimenter's perspective. The player needs to actively watch out for wasps coming from each direction.

After giving informed consent, participants put on one of the three motion capturing suits and the HMD. Similar to our first study, the skeleton configuration was calibrated in OptiTrack Motive. All participants were unaware of the conditions and instructed to hit the wasps and to score as many points as possible. Since the wasps came from different directions, the participants had to duck, turn their bodies, and use other parts of their body (e.g. their feet) to get points. After each condition, the participants had to complete the questionnaires. To gain more subjective insights, we asked them if they had any further remarks regarding the system.

5.3 Participants

We recruited 24 students (9 females, 15 males) via mailing lists of our institution. Mean age of the participants was 25.4 years (SD = 4.1) ranging from 19 to 33 years. Two participants were left handed. The participants were compensated with credit points for their study course and snacks. None of them was involved in our data collection study.

5.4 Results

The overall play time per condition was 140 sec on average (SD = 0.78) ranging from 139 to 141 secs. The timing show that play time remained constant and ensured an unbiased gaming experience between all Prediction Time conditions.

5.4.1 Game Score. Friedman ANOVA showed a significant effect of Prediction Time, $\chi^2(5) = 18.559$, $p = .002$, on the game score of the players. Pairwise comparisons using Wilcoxon signed rank tests showed a significant difference between *-50 ms* and *0 ms* (p = .033) as well as between *-50 ms* and *+100 ms* (p = .028). Mean values of the gaming scores are shown in Figure 5. Adding prediction did not decrease the gaming performance to the extent to which latency decreased the score.

5.4.2 Presence. There was no significant effect on the IPQ main score, $\chi^2(5) = 6.735$, $p = .241$, however, on the general presence subscale, $\chi^2(5) = 12.975$, $p = .023$. Subscales for involvement, $\chi^2(5) = 5.870$, $p = .319$. realism, $\chi^2(5) = 8.134$, $p = .375$, and spatial presence, $\chi^2(5) = 5.349$, $p = .374$, were not significant. Pairwise comparisons of the general presence subscale revealed significant differences between *-50 ms* and *+50 ms* (p = .013), as well as between *+50 ms* and *+200 ms* (p = .003). Subscale means are shown in Figure 6 (red lines).

5.4.3 Illusion of Body Ownership. There were no effect of Prediction Time on body ownership, $\chi^2(5) = 5.289$, $p = .381$, and agency, $\chi^2(5) = 6.932$, $p = .225$. Significant effects were found for location, $\chi^2(5) = 27.418$, $p < .001$, with significant differences between *-50 ms* and *+50 ms* (p = .013), as well as between *+50 ms* and *+200 ms* (p = .004). Mean values (Figure 6 show that the highest perceived body location ratings were achieved during the *+50 ms* condition.

5.4.4 Motion Sickness. No effects were found on the measures of the MSAQ scale, $\chi^2(5) = 8.191$, $p = .146$.

5.5 Discussion

In our second study, we used a full-body motion capturing VR gaming environment to investigate the effects of predicted movements of the own avatar. Gaming performance indicated by the score (hostile entity hits) after each round indicated significant higher mean scores during the baseline condition and using the +50 ms prediction model as used in the delayed condition with -50 ms. Using our movement predicting models, all participants performed higher gaming scores than in the delayed condition. Highest scores were achieved during the baseline condition tough, however, the results indicate that predicted body movements do not decrease a user's performance to the same extent as in the -50 ms delayed condition. We even found that ratings of the perceived correct body location were the highest using the +50 ms model. This indicates that games can benefit from overcoming the system-related latency using motion prediction of body movements using ANN. We assume that predicting one's own body movements enable players in an immersive environment to believe that they respond faster and more accurately.

6 GENERAL DISCUSSION

In two studies, we explored the effects of full-body avatar movement predictions in VR using ANN. In our first study, we found that both performance and subjective measures in a standardized 2D Fitts' law task could not benefit from the predicted avatar movements. Interestingly, the prediction leads to similar interferences in one's own motor control in terms of performance and task performance

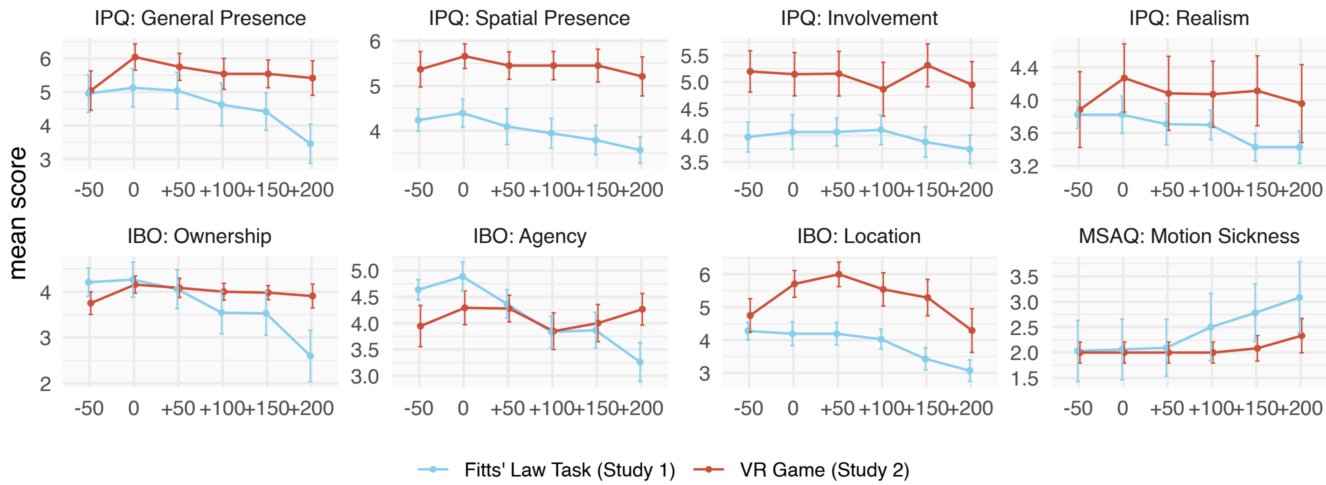

Figure 6: Mean ratings of presence and presence subscales of the IPQ, illusion of body ownership (IBO) subscales, and motion sickness during the Fitts' Law task (Study 1) and the VR Game study (Study 2). Error bars show CI95.

as delays of those movements. An analysis of the precision and the qualitative feedback show that the problems did not occurred due to lacking body control cause by the time difference, but due to inaccuracies caused by the models. In our second study, an immersive gaming scenario, however, the perceived accuracy of the own body location improved for the 50 ms condition. Presence and body assessments remained more stable and were higher than during the Fitts' task. We assume that games can utilize motion predictions of one's own body not only to overcome latency of the system but also to use optimized models to increase the users' performance in VR.

In both studies, future motions generally make participants perform worse due to lacking congruency between real and rendered movements. We assume that participants then intuitively slowed down their movements to get back in that congruency and to reduce the visual difference between expected and rendered virtual limb position. Using predicted body movements while playing the VR game, we found not only that ratings of presence, involvement, and body location were generally higher than during the Fitts' task but also that the likelihood of VR motion sickness caused by predictions of the own body movements occurred "later". This means that the participants were more likely to ignore negative effects of the predictions while being involved in an immersive application than during the monotonous and repetitive movement such as in a Fitts' law task.

Based on the game evaluation the participants experienced higher levels of avatar body location during the 50 ms predictions than in any other condition. This could be due to the fact that this condition offered the subjectively best body pose experience as the system latency was within this time window (*c.f.* System Latency Analysis subsection). Thus, we conclude that players could be more likely to benefit from motion predictions than users of other applications if speed and immersion can ensure that possible inaccuracies in the prediction are being ignored.

6.1 Limitations and Future Work

Predicting the own body movement in VR was not able to improve the performance of users and players, but in line with related work [15, 16, 27], we see future potential to reduce latency. Higher precision and accuarcy of model predictions can be potentially achieved through improved hardware and more complex ANN model architectures. More complex model architectures such as recurrent neural networks RNN, however, require significantly more time for computation in real time and considering that duration counter-effects a higher precision of the models as fewer pose information from previous frames are available. Better computation hardware might overcome this issue. However, even a theoretically perfect prediction of the users' body movement might not increase the performance when they cannot ignore the deviation between own body scheme and the prediction. Source code, data, and assets to replicate our work and to further explore the effects of full-body movement predictions are available on github[4].

We recommend to explore whether specifically captured data and better hardware allow more accurate models resulting in better user performance and better embodiment. In our Fitts' task, we only tested one experimental setup where all participants conducted the task on an horizontal axis sitting at a table. Since VR environments and the used motion capture system would theoretically allow other movements, free standing configurations could be evaluated as well. Additionally, providing and considering finer time steps and more limbs such as fingers and toes could allow the system to predict the users' body movements with higher precision. We would also like to point out the possibility of cheating in games using movement prediction if the output signal is intercepted before it reaches the other players in an online multiplayer game. The technology could also be used to reduce lags in real-time video streaming of games when graphics are being rendered on a remote server.

[4]https://github.com/Slimboy-90/motionprediction

REFERENCES

[1] 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

[2] A. D. Aladagli, E. Ekmekcioglu, D. Jarnikov, and A. Kondoz. 2017. Predicting head trajectories in 360° virtual reality videos. In *2017 International Conference on 3D Immersion (IC3D)*. 1–6.

[3] Robert S. Allison, Laurence R. Harris, Michael Jenkin, Urszula Jasiobedzka, and James E. Zacher. 2001. Tolerance of Temporal Delay in Virtual Environments. In *Proceedings of the Virtual Reality 2001 Conference (VR'01) (VR '01)*. IEEE Computer Society, USA, 247.

[4] Tomohisa Asai. 2015. Agency elicits body-ownership: proprioceptive drift toward a synchronously acting external proxy. *Experimental brain research* 234 (02 2015). https://doi.org/10.1007/s00221-015-4231-y

[5] Elie Cattan, Amélie Rochet-Capellan, Pascal Perrier, and François Bérard. 2015. Reducing Latency with a Continuous Prediction: Effects on Users' Performance in Direct-Touch Target Acquisitions. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces* (Madeira, Portugal) *(ITS '15)*. Association for Computing Machinery, New York, NY, USA, 205–214. https://doi.org/10.1145/2817721.2817736

[6] Francois Chollet. 2019. Keras. https://keras.io/callbacks/. Accessed on 2019-08-12.

[7] Marcello Costantini and Patrick Haggard. 2007. The rubber hand illusion: Sensitivity and reference frame for body ownership. *Consciousness and Cognition* 16, 2 (2007), 229 – 240. https://doi.org/10.1016/j.concog.2007.01.001

[8] 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

[9] Stephen R. Ellis, Katerina Mania, Bernard D. Adelstein, and Michael I. Hill. 2004. Generalizeability of Latency Detection in a Variety of Virtual Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 48, 23 (2004), 2632–2636. https://doi.org/10.1177/154193120404802306 arXiv:https://doi.org/10.1177/154193120404802306

[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] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Geoffrey Gordon, David Dunson, and Miroslav Dudík (Eds.), Vol. 15. PMLR, Fort Lauderdale, FL, USA, 315–323. http://proceedings.mlr.press/v15/glorot11a.html

[13] Mar Gonzalez-Franco and Tabitha C. Peck. 2018. Avatar Embodiment. Towards a Standardized Questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74. https://doi.org/10.3389/frobt.2018.00074

[14] M S Graziano. 1999. Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Sciences of the United States of America* 96, 18 (aug 1999), 10418–21. http://www.ncbi.nlm.nih.gov/pubmed/10468623http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC17903

[15] Niels Henze, Markus Funk, and Alireza Sahami Shirazi. 2016. Software-Reduced Touchscreen Latency. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 434–441. https://doi.org/10.1145/2935334.2935381

[16] Niels Henze, Sven Mayer, Huy Viet Le, and Valentin Schwind. 2017. Improving Software-Reduced Touchscreen Latency. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 107, 8 pages. https://doi.org/10.1145/3098279.3122150

[17] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580 [cs.NE]

[18] Shu Imaizumi and Tomohisa Asai. 2015. Dissociation of agency and body ownership following visuomotor temporal recalibration. *Frontiers in Integrative Neuroscience* 9 (05 2015). https://doi.org/10.3389/fnint.2015.00035

[19] ISO. 2012. 9241–411 Ergonomics of human-system interaction–Part 411: Evaluation methods for the design of physical input devices. *International Organization for Standardization* (2012).

[20] Ricardo Jota, Albert Ng, Paul Dietz, and Daniel Wigdor. 2013. How Fast is Fast Enough? A Study of the Effects of Latency in Direct-Touch Pointing Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2291–2300. https://doi.org/10.1145/2470654.2481317

[21] Topi Kaaresoja, Eve Hoggan, and Emilia Anttila. 2011. Playing with Tactile Feedback Latency in Touchscreen Interaction: Two Approaches, Vol. 6947. 554–571. https://doi.org/10.1007/978-3-642-23771-3_42

[22] 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

[23] 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

[24] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG]

[25] Jr. L. James Smart, Thomas A. Stoffregen, and Benoît G. Bardy. 2002. Visually Induced Motion Sickness Predicted by Postural Instability. *Human Factors* 44, 3 (2002), 451–465. https://doi.org/10.1518/0018720024497745 arXiv:https://doi.org/10.1518/0018720024497745 PMID: 12502162.

[26] Joseph J. LaViola. 2000. A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bull.* 32, 1 (Jan. 2000), 47–56. https://doi.org/10.1145/333329.333344

[27] 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. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (Brighton, United Kingdom) *(ISS '17)*. Association for Computing Machinery, New York, NY, USA, 230–239. https://doi.org/10.1145/3132272.3134138

[28] I. Scott MacKenzie and William Buxton. 1992. Extending Fitts' Law to Two-dimensional Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. ACM, New York, NY, USA, 219–226. https://doi.org/10.1145/142750.142794

[29] I. Scott MacKenzie and Colin Ware. 1993. Lag as a Determinant of Human Performance in Interactive Systems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 488–493. https://doi.org/10.1145/169059.169431

[30] Julieta Martinez, Rayat Hossain, Javier Romero, and James J. Little. 2017. A Simple yet Effective Baseline for 3D Human Pose Estimation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

[31] Sven Mayer, Jens Reinhardt, Robin Schweigert, Brighten Jelke, Valentin Schwind, Katrin Wolf, and Niels Henze. 2020. Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020-04-25) *(CHI '20)*. ACM, Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376340

[32] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks. 2003. Effect of latency on presence in stressful virtual environments. In *IEEE Virtual Reality, 2003. Proceedings.* 141–148. https://doi.org/10.1109/VR.2003.1191132

[33] Albert Ng, Michelle Annett, Paul Dietz, Anoop Gupta, and Walter F. Bischof. 2014. In the Blink of an Eye: Investigating Latency Perception during Stylus Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1103–1112. https://doi.org/10.1145/2556288.2557037

[34] Albert Ng and Paul H. Dietz. 2014. The effects of latency and motion blur on touch screen user experience. *Journal of the Society for Information Display* 22, 9 (2014), 449–456. https://doi.org/10.1002/jsid.243 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jsid.243

[35] Kristine L. Nowak and Frank Biocca. 2003. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 12, 5 (2003), 481–494. https://doi.org/10.1162/105474603322761289 arXiv:https://doi.org/10.1162/105474603322761289

[36] Andriy Pavlovych and Wolfgang Stuerzlinger. 2011. Target Following Performance in the Presence of Latency, Jitter, and Signal Dropouts. In *Proceedings of Graphics Interface 2011* (St. John's, Newfoundland, Canada) *(GI '11)*. Canadian Human-Computer Communications Society, Waterloo, CAN, 33–40.

[37] Sebastian Pohl, Armin Becher, Thomas Grauschopf, and Cristian Axenie. 2019. *Neural Network 3D Body Pose Tracking and Prediction for Motion-to-Photon Latency Compensation in Distributed Virtual Reality*. 429–442. https://doi.org/10.1007/978-3-030-30508-6_35

[38] Holger Regenbrecht and Thomas Schubert. 2002. Real and Illusory Interactions Enhance Presence in Virtual Environments. *Presence: Teleoper. Virtual Environ.* 11, 4 (Aug. 2002), 425–434. https://doi.org/10.1162/105474602760204318

[39] E.W. Saad, Thomas Caudell, and Donald Wunsch. 1999. Predictive head tracking for virtual reality, Vol. 6. 3933 – 3936 vol.6. https://doi.org/10.1109/IJCNN.1999.830785

[40] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoper. Virtual Environ.* 10, 3 (June 2001), 266–281. https://doi.org/10.1162/105474601300343603

[41] 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. In *Proceedings of the ACM Symposium on Applied Perception (SAP '18)* (2018-08-03). ACM, New York, NY. https://doi.org/10.1145/3225153.3225158

[42] Valentin Schwind, Sven Mayer, Alexandre Comeau-Vermeersch, Robin Schweigert, and Niels Henze. 2018. Up to the Finger Tip: The Effect of Avatars on Mid-Air Pointing Accuracy in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 477–488. https://doi.org/10.1145/3242671.3242675

[43] Andrea Serino, Adrian Alsmith, Marcello Costantini, Alisa Mandrigin, Ana Tajadura-Jimenez, and Christophe Lopez. 2013. Bodily ownership and self-location: Components of bodily self-consciousness. *Consciousness and Cognition* 22, 4 (2013), 1239 – 1252. https://doi.org/10.1016/j.concog.2013.08.013

[44] 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

[45] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria Sanchez-Vives. 2009. Inducing illusory ownership of a virtual body. *Frontiers in Neuroscience* 3 (2009), 29. https://doi.org/10.3389/neuro.01.029.2009

[46] 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, 56 (2014), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

[47] Google Brain Team. 2015. TensorFlow. https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/SGD. Accessed on 2019-05-21.

[48] 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

[49] Thomas Waltemate, Irene Senna, Felix Hülsmann, Marieke Rohde, Stefan Kopp, Marc Ernst, and Mario Botsch. 2016. The Impact of Latency on Perceptual Judgments and Motor Performance in Closed-Loop Interaction in Virtual Reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich, Germany) *(VRST '16)*. Association for Computing Machinery, New York, NY, USA, 27–35. https://doi.org/10.1145/2993369.2993381

[50] Jiann-Rong Wu and Ming Ouhyoung. 2000. On Latency Compensation and Its Effects for Head Motion Trajectories in Virtual Environments,. *The Visual Computer* 16 (03 2000), 79–90. https://doi.org/10.1007/s003710050198

[51] Ye Yuan and Anthony Steed. 2010. Is the rubber hand illusion induced by immersive virtual reality?. In *2010 IEEE Virtual Reality Conference (VR)*. IEEE, 95–102. https://doi.org/10.1109/VR.2010.5444807