

# Sick Moves! Motion Parameters as Indicators of Simulator Sickness

Tobias Feigl<sup>\*,§</sup>, Daniel Roth<sup>†</sup>, Stefan Gradl<sup>‡</sup>, Markus Wirth<sup>‡</sup>, Marc Erich Latoschik<sup>†</sup>, Bjoern M. Eskofier<sup>‡</sup>,  
Michael Philippsen<sup>\*</sup> and Christopher Mutschler<sup>§</sup>

**Abstract**— We explore motion parameters, more specifically gait parameters, as an objective indicator to assess simulator sickness in Virtual Reality (VR). We discuss the potential relationships between simulator sickness, immersion, and presence. We used two different camera pose (position and orientation) estimation methods for the evaluation of motion tasks in a large-scale VR environment: a simple model and an optimized model that allows for a more accurate and natural mapping of human senses. Participants performed multiple motion tasks (walking, balancing, running) in three conditions: a physical reality baseline condition, a VR condition with the simple model, and a VR condition with the optimized model. We compared these conditions with regard to the resulting sickness and gait, as well as the perceived presence in the VR conditions. The subjective measures confirmed that the optimized pose estimation model reduces simulator sickness and increases the perceived presence. The results further show that both models affect the gait parameters and simulator sickness, which is why we further investigated a classification approach that deals with non-linear correlation dependencies between gait parameters and simulator sickness. We argue that our approach could be used to assess and predict simulator sickness based on human gait parameters and we provide implications for future research.

**Index Terms**—Human-centered computing, virtual reality, user studies, computing methodologies, perception, machine learning.

## 1 INTRODUCTION

Today's VR systems support learning, narrative experiences, gaming, physical and psychological therapy, enabling users to move and navigate in the simulations using various interaction metaphors [61, 78], etc. For a convincing experience of the simulation, users have to feel comfortable, become involved, and experience *presence*, the sense of "being there" [20, 58, 88] or feeling a "place illusion" [69].

While presence is a phenomenon of everyday life [13] it lays at the center of a virtual experience and can be further distinguished by physical, social and self-related aspects [39], or assessment factors that build on involvement, sensory fidelity, realism, interface quality, adaptation, and immersion [87]. Skarbez et al. [65] provide a survey on self-reported, behavioral, psychological, and physiological measures of presence. *Immersion* was identified as an important factor that facilitates the perception of presence [67, 71]. In contrast to presence, immersion can be described as a technologically-driven concept: "The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it is 'immersive'." [67]. A perfect immersion would hence require a one-to-one mapping of all senses from the real to the virtual world [4]. Mapping inaccuracies and simulation errors, e.g., position or orientation tracking errors, latency, jitter, and flicker disturb the user and cause *simulator sickness* [37, 77, 91], the analogon of motion-sickness [29]. Vice versa, Slater argues that the place illusion strongly depends on sensorimotor contingencies in all sensory modalities, i.e., an accurate and consistent mapping between physical actions and virtual feedback [69]. Fig. 1 summarizes our interpretation of the literature: A better mapping accuracy, e.g., highly accurate sensors and simulations that capture and map motion more

realistically, can yield a higher immersion. In turn, a more accurate mapping can reduce cue conflicts and thus simulator sickness [27, 72]. Higher immersion and lower sickness can foster the user's perception of presence, and further support the suspension of disbelief. However, a high level of simulator sickness seems to hinder and suppress presence [47]. Note that a high immersion is not the only boundary condition for presence, e.g., in desktop systems, users may perceive presence, but they are also bound to longer exposure, deliberate attention, and learning [69]. For example, users have to learn and process mappings between (abstracted) physical action and virtual feedback.

While immersion can be objectively measured [67], for example by assessing the degrees of freedom of movement, or the field of view, presence is a phenomenon of subjective perception and is thus difficult to assess objectively. Despite debates [66, 68], questionnaires such as the Slater-Usch-Steed or the Presence Questionnaire (PQ) [71, 88, 89] are widely adopted. Similarly, questionnaires are the current practice to assess simulator sickness. It has been argued to differentiate *cybersickness* from simulator sickness [76]. Yet, assessing simulator sickness is widely adopted in VR research. Most commonly, the simulator sickness questionnaire (SSQ) [29], the fast SSQ [31], or nausea rating scales [43, 57] are employed. According to Meehan et al. [45], valuable measurement instruments should be reliable (produce repeatable results within and across subjects), valid (measuring the underlying construct), sensitive (discriminate amongst multiple outcome levels), and objective (well shielded from the subject and experimenter bias). In general, questionnaires are time-consuming for the participant and suffer from structural disadvantages. For example, they are difficult to administer during the stimulus presentation, so that often shorter oral (in-situ) questionnaires are used. Post-exposure measures are time delayed and thus may not capture the maximal sickness level. Further disadvantages are a potential bias or reliability issues (e.g., inconsistencies due to structures that assess latent variables) [31, 68].

We argue that if simulator sickness results in similar symptoms than motion sickness, such as sweating, nausea, and losing balance [27, 30], it results in similar physiological as well as motoric reflexes and restrictions. Hence, it should be objectively measurable. Researchers and practitioners may have observed users that restrict their movement pace, stumble, or have difficulties to execute precise movements because of latency and sickness, as users suffer from postural disturbance (also known as postural instability, postural disequilibrium, or ataxia) [30, 35, 81]. An objective and implicit measure to assess simulator sickness through motion parameters may be of great benefit to VR researchers and applicants, as a real-time simulator sickness indicator may allow for a controlled exposure or a dynamic content adaptation in response to sickness. We propose to use gait motion parameters for an objective sickness assessment, as gait is most likely to suffer

<sup>\*</sup>Is with the Programming Systems Group, Friedrich-Alexander University Erlangen-Nürnberg (FAU), {tobias.feigl | michael.philippsen}@fau.de.

<sup>†</sup>Is with the Human Computer Interaction Group, University of Würzburg (JMU), {daniel.roth | marc.latoschik}@uni-wuerzburg.de.

<sup>‡</sup>Is with the Machine Learning and Data Analytics Lab, FAU, {stefan.gradl | markus.wirth | bjoern.eskofier}@fau.de.

<sup>§</sup>Is with the Machine Learning and Information Fusion Dept., Fraunhofer Institute of Integrated Circuits (IIS) Nürnberg, christopher.mutschler@iis.fraunhofer.de.

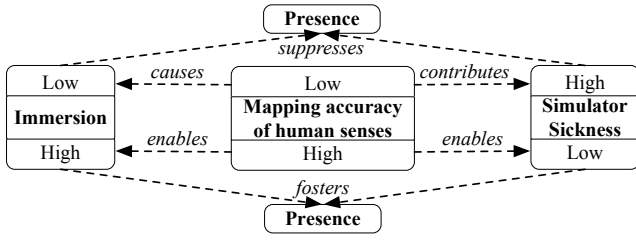


Fig. 1. Relationships between technical concepts.

from simulator sickness and as gait motion is most likely a motion that is present in many VR simulations. While previous work measured or manipulated gait variables in VR [1, 12, 23–25, 41, 46, 49, 53] or suggested pose measures to assess sickness [17, 30, 81], there is no work yet on the impact of simulation characteristics on gait parameters and simulator sickness, and their relationship with presence.

### 1.1 Contributions

We present novel findings on the impact of simulation characteristics on gait, simulator sickness, and presence, and argue for motion parameters as a potential measure to assess simulator sickness. There are four main contributions. (1) We used two pose estimation models (i.e., methods for camera pose tracking in a large-scale VR environment that sense and map a user’s camera pose from the real to the virtual world) with different characteristics: a simple approach that induces more (varying) latency and results in a lower accuracy of the mapping of human senses, and an optimized approach that preserves the fidelity of sensory modalities and allows for a higher immersion. (2) We compared both models and a non-VR baseline in a user study with 34 participants in repeated measures. We find that the optimized model significantly decreases simulator sickness, significantly increases presence, and significantly impacts gait parameters. (3) We discuss the stability of the gait parameters as a comparison measure as well as a direct (singular) measure of sickness. (4) We classify the models of origin as well as the sickness level based on different parameter sets which implies the potential for a real-time sickness assessment via gait parameters.

### 1.2 Structure

Sec. 2 reviews related work. Sec. 3 discusses our apparatus and pose estimation models along with the description of the evaluative study. Sec. 4 presents the results, further investigating gait parameters as sickness indicator. Sec. 5 describes our classification approach. After a discussion in Sec. 6, we conclude.

## 2 RELATED WORK

### 2.1 Relations of Simulator Sickness and Presence

With regard to presence, Borrego et al. [6] evaluate both presence and motion-sickness in their small-scale VR system ( $5 \times 5$  m) but they do not consider gait parameters. Similar to simulator sickness, the refresh rate and latency seem to affect the perceived presence. Meehan et al. [47] show that a frame rate per second  $> 15$  yields good presence scores [47]. A motion-to-photon latency  $> 100$  ms decreases presence [48]. Other sickness affecting characteristics are also found to affect presence. Recent studies [55, 91] do not find sickness differences between 18 ms and 90 ms of latency, but between constant and varying latency. Although we make use of these findings in our models, results are not directly comparable, as the human senses that are stimulated throughout all experiments vary. Our users walk freely on a large space, thus perception and its effects may differ.

### 2.2 Objective Measures for Simulator Sickness

Early works use so-called postural equilibrium tests [17, 30, 81] to objectively assess sickness symptoms, i.e., dynamic and static pose assessment such as standing on one leg or walking on a line with eyes open or close. Hamilton et al. [17] criticize the reliability of these tests. Closest to our approach, Kemeny et al. [26] show that fast head

rotations in VR induce postural instabilities and increase simulator sickness. Supporting these findings, Villard et al. and Li et al. [41] find that visual oscillations induce both postural instability (e.g., changes in the body sway) and simulator sickness over time. While we are inspired by these results, our proposed approach uses an assessment of gait parameters and automated analyses. Furthermore, we do not need a phase for the participant to train certain poses compared to postural equilibrium tests.

Previous works also investigate visually induced self-motion between frames to estimate motion sickness by analyzing the optical flow [38], assess scene movement [74], or predict motion-sickness by applying deep convolutional autoencoders (trained on video streams with and without motion-sickness) [33]. In contrast to our work, these works rely on visually induced motion sickness (movement produced by purely visual stimulation, i.e.,vection) which by itself causes simulator sickness [32].

Bertin et al. [3] discover relationships between simulator sickness and skin resistance as well as skin temperature, both of which were lower in sick than in non-sick participants of a driving simulator. When seated subjects are perturbed by virtual visual stimuli this causes changes in physiological parameters (e.g., heartbeat, skin conductivity or temperature, breathes per epoch) [8] which can be used to determine simulator sickness [47]. While such parameters can be a reliable measure, they are complicated to assess in movement tasks or physical activity as in our situation.

### 2.3 Gait in VR

Spatio-temporal gait parameters such as the stance or swing time, step length, or angular relations can reliably describe human gait performance [23]. As such, gait analysis in VR can be used as a means to investigate gait performance and to assess functional mobility [78] and to inspect impacts of simulation characteristics. Boone et al. [5] explore gait parameters in VR as a measure of rehabilitation. Thompson et al. [82] technically evaluate the effect of perturbed optical flow on the gait balance and the regulation of walking control, but they do not link their findings to simulator sickness or presence. Hollman et al. [21, 22] use a VE with oscillating visuals to induce gait instability in treadmill-based walking (compensatory efforts to control the body’s mass). Lansink et al. [36] use a treadmill and compare VR with the physical world. They find that gait parameters reflect instability (stride length or width, velocity), suggesting that walking in a VE induces instability even to healthy subjects.

Even though unnatural navigation techniques (e.g., walking in place, flying, teleportation, or joystick movement) impact presence, affect gait, and add discomfort in VR, while more natural walking techniques increase presence and decrease sickness [70, 84], related work that considers movement often makes such restrictions [23]. Thus, the regularized gait and its characteristics [75] render the results only partially applicable to redirected walking or free walking. For example, Rieser et al. [59] find that humans constrain their self-motion to become unnatural or tensed in unknown and limited environments. To avoid this bias, we evaluate gait parameters, sickness, and presence with natural walking. Yet, differences to the physical world can occur. Mohler et al. [50] find that participants had a shorter stride length, slower walking velocity, and lower head-trunk angle in VR compared to real-world walking. Therefore, we also compare our pose estimation models to a real-world baseline.

With regard to mapping accuracy and superficial mapping, Janek et al. [23] define the terms isometric (one-to-one mapping of real and virtual movement) and non-isometric (e.g., added translational gain) [23, 24]. Non-isometric mapping causes a divergence in spatio-temporal gait parameters for participants of all ages.

In summary, these studies show distinct effects of navigation techniques and the simulation characteristics on gait parameters.

### 2.4 Gait Classification

Preliminary research [2, 18, 64, 79] detects alterations in gait parameters by various methods that examine spatial measures [2] or temporal dynamics of movement [79]. From large amounts of input data, deep

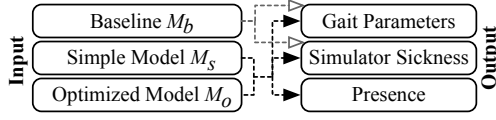


Fig. 2. Relationships between models and dependent measures.

neural networks can extract the information that is necessary to classify gait [18]. According to Smith et al. [73], the walking speed affects the spatio-temporal gait parameters in a non-linear way. However, there are also non-linearities in sensor measurements and hence, to interpret the sensor data and the gait parameters correctly, at least a second-order polynomial function is needed. This is in line with findings in [44, 54]. They use spatio-temporal features (from time and frequency domain) and apply a non-linear solver (e.g., Support Vector Machine) to derive the characteristics of gait and to provide an accurate classification of typical gait parameters. Our classification approach adopts their findings, classifies (non-linear) gait parameters, and predicts simulator sickness levels.

## 2.5 Hypotheses

Previous works separately investigated the effects of simulation characteristics on gait parameters, simulator sickness, and presence. As they review simulation latency and cue conflicts as a strong cause of simulator sickness, we hypothesize **H1**: *A more natural pose estimation model results in less sickness.*

As according to the literature, gait parameters seem to be affected by the characteristics of the simulation and the respective mapping accuracy we study **H2**: *A more natural pose estimation model impacts the gait in VR exposure.*

As the perceived presence seems to be higher in simulations that have a higher one-to-one mapping and accuracy we hypothesize **H3**: *A more natural pose estimation model results in a higher presence.*

While the related work has studied the presumed effects, the interconnection between sickness and gait is mostly unstudied. Hence, our research question is **RQ**: *Are gait parameters an objective measure of simulator sickness?*

## 3 METHOD

We use two different pose estimation models for a large-scale ( $40\text{ m} \times 35\text{ m}$ ) environment: a simple but unnatural model  $M_s$  and an optimized model  $M_o$  (we used a reduced portion ( $15\text{ m} \times 7\text{ m}$ ) as the running task implies that participants need space to be able to decelerate and stop). After pre-testing these models, we evaluate them using a within-subjects design (as simulator sickness is individual per subject). We compare a non-VR baseline  $M_b$ ,  $M_s$ , and  $M_o$ . For the baseline and for both models, participants perform multiple motion tasks. By measuring simulator sickness and presence as well as the objective outcomes for the gait parameters, we assess the relationships between the model parameters and the dependent measures, see Fig. 2. Furthermore, we evaluate a classifier.

### 3.1 Apparatus

We used the Unity3D engine (version 2017.4.1) to create our VR simulation. We displayed it with a Samsung Galaxy Note 4 smartphone (Android 6.0.1, Qualcomm Snapdragon 805 CPU, and 3 GB RAM) mounted to a Samsung GearVR HMD (version SM-R320, and 6 DOF Bosch BMI055 inertial measurement unit, IMU) which senses the user orientation. An InvenSense (MPU-6500) 6 DOF IMU was also attached to the HMD to classify the movement for the pose estimation models. We tracked user positions with RedFIR, an RF-based real-time location system (RTLS) using a single radio-frequency sensor, also attached to the HMD, see Fig. 3(a), that covers a space of up to  $300\text{ m} \times 300\text{ m}$  and is adaptable to 5G.

RedFIR implements a Kalman filter algorithm [15] that provides smooth positions with a circular error probable (i.e., horizontal accuracy) in 95% ( $CEP_{95}$ ) of  $< 10.2\text{ cm}$  and a precision of  $< 2.6\text{ cm}$  in

both static and dynamic motion situations. The accuracy is the absolute accuracy of a measurement to match a corresponding reference point. In contrast, precision describes how accurate several repeated measures match the corresponding reference position. For our pose estimation models, the accuracy does not matter, as the absolute point accuracy in relation to physical space is irrelevant for our experiment. The precision, however, is important. The worse the precision, the more *jumpy*, *jittery*, and *unnatural* the position is. Precision can be considered as a Gaussian distribution that *jumps* around a global optimum, i.e., the reference position. RedFIR's loss rate is  $0.0046\text{ Hz}$  on avg. Other open-source tracking systems may be used to reproduce our experiments [83].

As the HMD-estimated head orientations suffer from long-term yaw drift, we (re-)calibrated the absolute head orientation (and the corresponding VR-image) for every participant before a new task [9, 10]. To measure the gait parameters, we used eGait and two Shimmer IMU sensors [2], see Fig. 3(b), mounted to the same type of shoe (varying sizes to fit) for every participant. The eGait application on an Apple iPad Pro tablet used Bluetooth to communicate with the sensors to record their data streams, to perform the stride segmentation, the feature extraction, and to derive the gait parameters (details in Sec. 3.4.3).

Fig. 4 shows the real world and its virtual counterpart.  $W$ ,  $B$ , and  $R$  represent the trajectories for the Walking, Balancing, and Running tasks.  $S$  is a slalom path that was only used in the VR (details in Sec. 3.3). The test environment was bright and free of noise. To make users more sensitive to small movements and to provoke sickness reactions, the outer floor of the VR scene was textured with a ( $1.0\text{ m} \times 1.0\text{ m}$ ) black-and-white checkerboard pattern, the inner floor with a ( $10\text{ cm} \times 10\text{ cm}$ ) green checkerboard pattern, colored pillars, and a plank for balancing, see Fig. 4(b,c).

#### 3.1.1 Simple Pose Estimation Model $M_s$

For the experiments with the simple pose estimation model  $M_s$  we use both the absolute positions provided by the RTLS [15] and the calibrated (absolute) orientations from the Samsung GearVR (official Oculus orientation estimates) to render the pose, and thus the estimated camera viewpoint for the participant. While the orientation stream constantly delivers close to 60 orientations per second, the position stream only provides 20 updates per second. As we render the virtual world at a quasi-constant rate of  $60\text{ Hz}$ , the orientation update frequency suffices [10, 77]. But the low update frequency of the positions (to render the VR view) introduces a varying motion-to-photon delay (MTP) of  $246.34\text{ ms}$  on avg. ( $N = 87$ , min.  $198.7\text{ ms}$ , max.  $319.6\text{ ms}$ , SD  $76.4\text{ ms}$ ). We assessed MTP as the latency between physical translational (head) movement and the *according* visual feedback (display) [19, 42]. Whereas the rendering frame rate/frequency describes the simulation/display refresh rate, the MTP describes the latency (delay) of the position updates. We measured the positional MTP by frame counting [19]. Therefore, we simultaneously recorded sensor movements and the resulting screen response in repeated measures with a high-speed camera ( $1000\text{ Hz}$ ) and then assessed the number of frames between physical motion apexes and the corresponding screen response

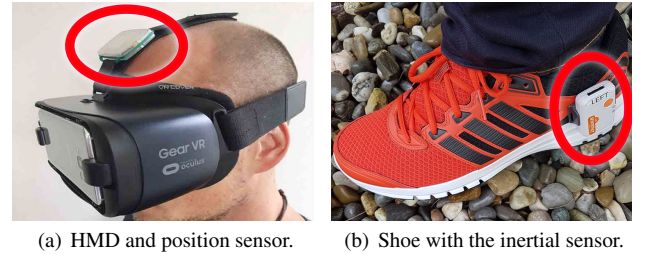


Fig. 3. (a) Samsung GearVR HMD (provides relative head orientation) and RedFIR position sensor (in the circle, provides absolute head position); (b) eGait system (in the circle, Shimmer Inertial-Navigation-Sensor (INS) that delivers gait information).

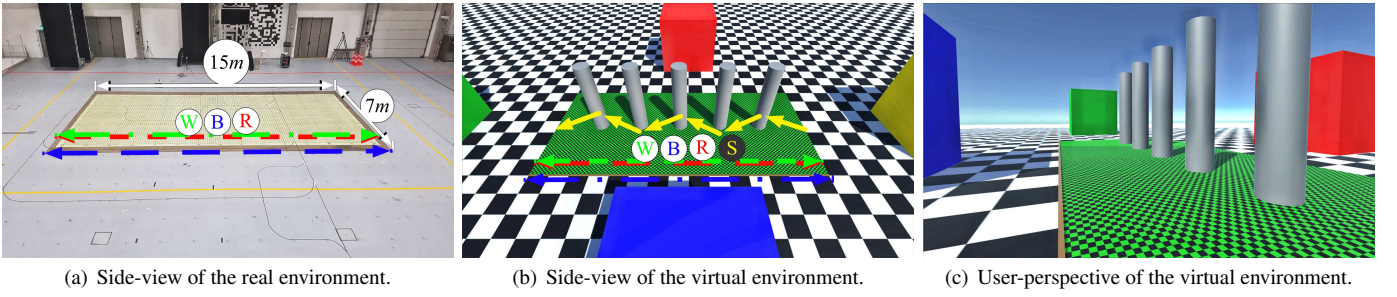


Fig. 4. Real and virtual environment of our experiments. Paths on which users (W)alk, (B)alance, (R)un, and perform a (S)alom walk.

(note: the HMD renders at 60 Hz, hence the screen response varies, SD 16.6 ms). The RTLS yields a precision error of  $< 2.6$  cm. The Kalman filter (KF) that is used in the positioning system suffers from sudden movement changes which result in over- and undershooting of the estimated trajectory. Hence, the mapping between real and virtual motion is incorrect, which facilitates simulator sickness. Thus, our simple model does *move* while standing still. This *jitter* is noticeable when standing still (subjective judgment), i.e., while moving it feels a bit like being held.

### 3.1.2 Optimized Pose Estimation Model $M_o$

Our optimized model  $M_o$  uses the same orientations and positions as  $M_s$ . But this time, we use the absolute positions in combination with the head orientations (provided by the Samsung GearVR) to calculate control points of a Kochanek-Bartels-Spline with parameters  $Tension=1$ ,  $Bias=-1$ , and  $Continuity=0$  [34]. This yields a precision error of unnoticeable  $< 0.01$  mm. Thus, our optimized model does not *move* while standing still.

Technically, the magnitude (velocity) of the spline's tangent represents the motion displacement, which we determine based on distances of consecutive absolute positions. As the KF of the tracking system [15] is parameterized to yield responsive but inaccurate and varying positions, we smooth the position with a head motion state (square root magnitude of the accelerometer and gyroscope sensors from the HMD). By correctly classifying the current motion state of a user's head we cap over- and undershooting of position changes. We exploit the fact that humans tend to move towards their viewing direction more frequently [9, 10]. Hence, we extrapolate the future control points (that steer the tangents and thus the view of the HMD) based on the orientation of the position vector (provided by the last positions by about 25%) and the absolute yaw head orientation (by about 75%). Thus, we can extrapolate the future position. Since we know that humans either tend to adapt to insignificant motion errors [60] or tolerate yaw orientation offsets below  $20^\circ$  while walking [10], our algorithm can make small mistakes. The high-rated motion state and its local processing on the HMD thus yields an accurate mapping of human senses and a smaller and constant MTP delay of 116.3 ms on avg. ( $N = 91$ , min. 98.5 ms, max. 148.3 ms, SD 18.7 ms), see Fig. 5.

For  $M_o$ , the tcb-spline enables the camera to move on fast cornering rides, i.g., when users turn abruptly, the tcb-spline can directly switch the direction. Moreover, t-clipping helps to avoid damping or overshooting and reduces the number of mispredictions. The combination of a weighted motion state classification and view-direction-based weighted extrapolation (to dip the movement along the tangent per delta time) yields a responsive, more natural pose estimation model. Finally, by adjusting the weights, i.e., adding/subtracting to/from each fragment per frame, we can unnoticeably compensate for accumulating errors.

### 3.1.3 Rendering

We always render the simulation at a fixed frame rate of 60 Hz (loss rate  $< 0.1$  Hz). We update the camera position at 20 Hz (RTLS raw positions) for  $M_s$ . For  $M_o$ , we extrapolate in-between at 60 Hz (a Kalman filter fuses IMU data at 100 Hz and RTLS raw positions at

20 Hz) with the tcb-spline. In both models, we always update the camera orientation at 60 Hz (calibrated GearVR orientation estimates). Note, the camera position and orientation are updated independently according to the pipeline in Fig. 5.

### 3.1.4 Pretest

In a preliminary experiment, we evaluated the configuration parameters of  $M_s$  and  $M_o$  and determined user perception. The pretest revealed that subjects ( $N = 46$ , 24 male, 22 female,  $M_{age} = 25.03$ ) felt (based on self-reported measures) less sick, more natural, more comfortable, and more involved when we render the virtual view with  $M_o$ . We thus employed these models for the main study.

## 3.2 Motion Tasks

We designed the tasks according to the related findings [16, 35, 51, 52]. According to [35, 51, 52] sickness is not only a function of motion-to-photon latency or inaccurate mapping of human senses, but also depends on the task. The more complex the task is and the longer it takes, the more sickness it causes. The tasks reflect what we believe is typical for large-scale VR simulations [63]. We gave instructions orally based on an experimenter protocol.

**Walking (W).** We asked the participants to naturally walk forward three times on a 15 m long straight path while we recorded their gait. For the baseline task  $T_{M_b}$ , we asked the participants to walk at a natural speed with either free or rigid head orientation. We introduced secondary-level objectives to account for a controlled exposure and focus in the VR assessments. In the first iteration of the VR assessments  $T_{M_s}$  and  $T_{M_o}$ , we asked participants to walk at a natural speed with non-rigid head orientation. In the second iteration, we encourage the participants to focus on far-away objects (green or yellow cubes) or near objects (e.g., the floor) while they walked along the path W. In the third iteration, we asked them to look to the left or right and to focus on far-away objects (red or blue cubes) or near objects (e.g., gray pylons).

**Balancing (B).** We asked the users to balance three times (as precisely as they can, i.e., with as few sidesteps as possible) naturally on a 15 m long and 10 cm wide virtual plank, see Fig. 4(c).

**Running (R).** We asked the participants to run three times (as fast as they can) naturally along the 15 m long path with free head orientation. We asked the participants to only stop running after leaving the measurement area (hence, in total they ran about 20 m but we also recorded the gait for 15 m). For the tasks  $T_{M_s}$  and  $T_{M_o}$  we encouraged

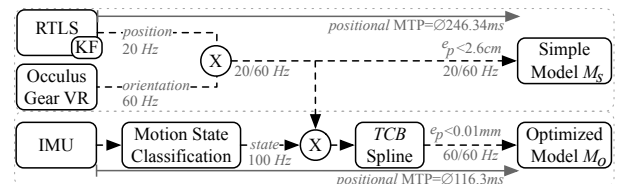


Fig. 5. Pose estimation processing pipeline (precision error  $e_p$ ).

the participants to specifically focus on the moment when they stopped running, i.e., when most likely artifacts occur due to  $M_s$  and  $M_o$ .

**Slalom (S).** Since in a preliminary study we found that walking a slalom disturbs gait behavior and measurements (such as sidesteps, missteps, and strafing), we did not record the gait during the slalom walk. On the other hand, as slalom is typical in sports applications, its impact on simulator sickness is relevant. Hence, we asked the participants to walk once on a slalom path  $S$  for the virtual tasks  $T_{M_s}$  and  $T_{M_o}$ , see Fig. 4(b), with quasi-free head orientations.

### 3.3 Procedure

Fig. 6 holds a detailed overview of the study procedure. The complete experiment lasted for about 45 minutes. After greeting the participants we presented the study information before they agreed to participate. We then assessed demographics with a pre-study questionnaire, followed by the baseline assessment  $T_{M_b}$  with the motion tasks Walk ( $W$ ), Balance ( $B$ ), and Run ( $R$ ). After assessing a visual test, the participant continued with the motion tasks while either being exposed to  $M_s$  or to  $M_o$ , followed by the assessment of the respective other model. The first model after the baseline was varied and balanced throughout the sample. Again, we asked participants to perform  $W$ ,  $B$ , and  $R$  three times. We also asked them to walk a slalom path ( $S$ ) in  $T_{M_o}$  and  $T_{M_s}$ . After  $T_{M_b}$ ,  $T_{M_s}$ , and  $T_{M_o}$ , we assessed simulator sickness. In  $T_{M_s}$  and  $T_{M_o}$ , after  $W$  and  $R$ , we additionally assessed presence (see Sec. 3.4.1). We asked the participants to stand still immediately before and after  $W$ ,  $B$ , and  $R$ , to start and stop the gait recording with a clean alignment. We recorded the gait information of both feet for each motion task in each of the three assessments. On average, participants spent 10.71 minutes in  $M_s$  and 11.79 minutes in  $M_o$  to fulfill the tasks.

### 3.4 Measures

#### 3.4.1 Control Measures

We asked the subjects to gauge their *sensitivity to sickness* (e.g., on a car, a boat, an amusement ride, or a plane) on a 0-10 scale.

To ensure the best possible virtual view (direction, sharpness, and brightness) for all participants, we carried out a visual test (both in reality and in the VR) before assigning tasks. The VR visual test mimicked the usual visual tests known from reality (users determined the alignment of apertures of circles of varying size) [40]. The users performed the visual tests first without the HMD and then with the HMD. To successfully pass the virtual visual tests, users had to identify all visual tasks clearly. Therefore, they had to take their time to adjust the HMD device perfectly. All users had a normal or corrected-to-normal vision (i.e., contact lenses or glasses) and passed both visual tests.

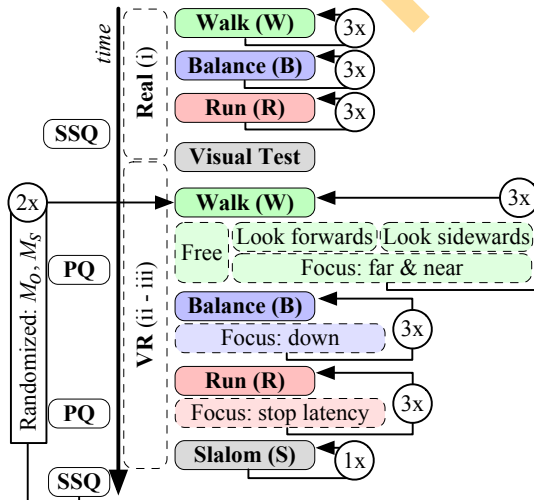


Fig. 6. Study procedure.

#### 3.4.2 Subjective Measures

After each assessment ( $T_{M_b}$ ,  $T_{M_s}$ ,  $T_{M_o}$ ) we measured the simulator sickness using the SSQ questionnaire [29] and analyzed the factors Nausea, Oculomotor, Disorientation, and Total score. All subjects had to indicate their current state on a 0-3 scale. In addition, we assessed the perceived presence after  $W$  and  $R$  in both  $T_{M_s}$  and  $T_{M_o}$ , using the revised PQ questionnaire from [62, 88, 89]. Subjects had to indicate their perception on a 1-7 scale. We assessed these presence measures to investigate the robustness of the construct and its influence across the tasks. Because of the complexity of the experiment, we only assessed presence after the exposure to  $W$  and  $R$  since subjects had a hard time focusing on  $B$  and  $S$  and thus were unable to also provide self-reports on presence.

#### 3.4.3 Gait Parameter Extraction

We used typical spatio-temporal gait parameters from the literature [49, 53] that are known to show relationships between different movement behaviors, see Table 1. Hence, we expected these parameters to show (i) a close relation between the baseline model  $M_b$  and the optimized model  $M_o$ , and (ii) to show a clear divergence between  $M_b$  and the simple/unnatural model  $M_s$  and – according to **H2** – to show a clear divergence between  $M_o$  and  $M_s$ .

We extracted the data from the IMU recordings, i.e., 3 DoF acceleration  $a_{xyz}$  and 3 DoF gyroscope  $g_{xyz}$  signal streams. The extraction worked as follows. In the recorded streams ( $a_{xyz}$ ,  $g_{xyz}$ ) we detected and identified the footsteps at times  $t$ . To do so, we found the moments when a foot touches the ground (stance event) by detecting the peaks  $P$  in the sum of the square root magnitude of the sensor data:  $steps(t) = \sqrt{a_x^2(t) + a_y^2(t) + a_z^2(t)} + \sqrt{g_x^2(t) + g_y^2(t) + g_z^2(t)}$ . This way of step detection is invariant of the placement and the orientation of the sensor. We used a threshold  $\geq 13.0$  that we determined in a preliminary study. With a step found in the sensor data, we selected a window around its peak that holds all information belonging to this step. In this window of raw sensor data, we determined the gait parameters of the step according to [64].

### 3.5 Participants

We recruited 34 participants (8 female,  $M_{age} = 27.4$ ,  $SD_{age} = 4.98$ ) using a mailing list and recruiting system of our institution. All were blind to the goal of the experiment. All had normal or corrected to normal vision. Thirteen wore contact lenses or glasses. None had any motoric impairments. 26 participants had previous experience with VR environments; of those, 12 indicated they had less than three previous VR experiences and 14 indicated they had more than three previous experiences. Most of the pre-experiences were with Oculus Rift, HTC Vive, and Samsung GearVR headsets. 22 participants indicated that they do sports regularly. 24 participants indicated that they played video games, 17 on a daily basis. We had to exclude gait data of 9 participants due to corrupt data (the Bluetooth connection between the Shimmer sensors was disturbed for a full day). Hence, we extracted

Table 1. Gait parameters.

Type [unit]	Description
Heel strike angle [deg]	Angle between heel and ground, when a heel touches the ground (during movement)
Toe out angle [deg]	Angle between mid heel to 2nd toe (long foot axis)
Gait velocity [m/s]	Stride length divided by stride time
Max. lateral excursion [cm]	Max. foot movement to the right/left side
Max. toe clearance [cm]	Max. distance between toe and ground, when the foot swings in the air (forward movement)
Stance time [%]	Percentage of the time when a foot is on the floor (not on the floor)
Stance time [s]	Time when a foot is on the floor
Stride length [cm]	Sum of two consecutive step lengths
Stride time [s]	Time of a full gait cycle
Swing time [%]	Percentage of the time when a foot is in the air
Swing time [s]	Time when a foot is in the air

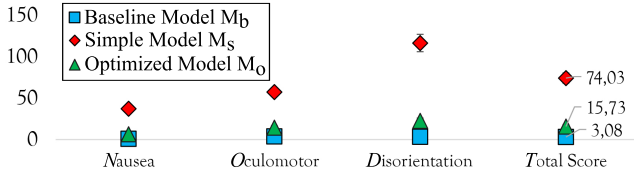


Fig. 7. Mean values of the sickness measure [29] for the subscores and the total score. Error bars denote the standard error.

valid gait parameters from 25 participants (5 female,  $M_{age} = 26.96$ ,  $SD_{age} = 4.58$ ). Of those, 12 wore contact lenses or glasses, and 21 had previous experience with VR.

## 4 RESULTS

### 4.1 Simulator Sickness

To be comparable to other published results, the sickness scores, including the sum of the related items, were calculated and weighted according to the procedure and recommendations by Kennedy et al. [29]. The total score is the sum of the unweighted subscores, multiplied by 3.74 (conversion formula), see Fig. 7 for the results. To compare the sickness scores between each condition, we first conducted a Friedman two-way analysis of variance by ranks for related samples with a follow-up pairwise comparison (Dunn test for non-parametric pairwise comparisons). We chose a Friedman test as we considered the 4-point scale of the SSQ (ranging from “none” to “severe”, ordinal categories in a logical order [11]) and the pre-exposure measurements to typically have little variance. We also report parametric ANOVAs and pairwise t-tests to support the results. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons. Statistical significance was accepted at the  $p < .0167$  level. Adjusted values are reported.

The **Total sickness score** was significantly different for the non-parametric comparison of  $M_b$ ,  $M_o$ , and  $M_s$ ,  $\chi^2(2) = 63.75$ ,  $p < .001$ . Non-parametric pairwise comparisons showed that the total sickness was statistically significantly different between  $M_b$  ( $Mdn = 0.0$ ) and  $M_s$  ( $Mdn = 61.71$ ,  $p_{adj} < .001$ ), between  $M_b$  and  $M_o$  ( $Mdn = 11.22$ ,  $p_{adj} = .005$ ) and between  $M_s$  and  $M_o$  ( $p_{adj} < .001$ ). The parametric ANOVA confirmed these results,  $F(1.075, 35.474) = 147.68$ ,  $p < .001$ ,  $\eta_p^2 = .817$ . All pairwise t-tests confirmed statistically significant differences ( $p's < .001$ ). An exploratory parametric analysis of covariance (ANCOVA) calculated by using trait sickness and gender as covariates showed that neither trait sickness (i.e., the subjective pre-assessed sensitivity of each participant) nor gender significantly affected these results.

Assessed by non-parametric testing, the **Nausea score** was statistically significantly different for the assessments of  $M_b$ ,  $M_o$ , and  $M_s$ ,  $\chi^2(2) = 63.25$ ,  $p < .001$ . Pairwise non-parametric comparisons showed that nausea scores were statistically significantly different between  $M_b$  ( $Mdn = 0.0$ ) and  $M_s$  ( $Mdn = 28.62$ ,  $p_{adj} < .001$ ), between  $M_s$  and  $M_o$  ( $Mdn = 0.0$ ,  $p_{adj} < .001$ ), but not between  $M_b$  and  $M_o$  ( $p_{adj} = .345$ ). Assessed by non-parametric testing, the **Disorientation score** was statistically significantly different for the assessments of  $M_b$ ,  $M_o$ , and  $M_s$ ,  $\chi^2(2) = 64.19$ ,  $p < .001$ . Pairwise non-parametric comparisons showed that disorientation scores were statistically significantly different between  $M_b$  ( $Mdn = 0.0$ ) and  $M_s$  ( $Mdn = 97.44$ ,  $p_{adj} < .001$ ), between  $M_s$  and  $M_o$  ( $Mdn = 13.92$ ,  $p_{adj} < .001$ ) and between  $M_b$  and  $M_o$  ( $p_{adj} = .011$ ). Assessed by non-parametric testing, the **Oculomotor score** was statistically significantly different for the assessments of  $M_b$ ,  $M_o$ , and  $M_s$ ,  $\chi^2(2) = 62.54$ ,  $p < .001$ . Pairwise non-parametric comparisons showed that oculomotor scores were statistically significantly different between  $M_b$  ( $Mdn = 0.0$ ) and  $M_s$  ( $Mdn = 56.85$ ,  $p_{adj} < .001$ ), between  $M_s$  and  $M_o$  ( $p_{adj} < .001$ ), but not between  $M_b$  and  $M_o$  ( $Mdn = 7.58$ ,  $p_{adj} = .033$ ).

Results of the parametric ANOVAs and consecutive pairwise comparisons (t-tests) for nausea, disorientation, and oculomotor scores corroborated with the non-parametric results.

### 4.1.1 Reliabilities

A reliability analysis of the SSQ measure revealed that multiple items had zero variance and that only the Oculomotor sub-scale yielded a reliable measure across all models, see Table 2. Items with zero variance are salivation, sweating, nausea, stomach, burping, and dizziness (open eyes) in the  $M_b$  assessment; burping in the  $M_s$  assessment; salivation, sweating, and burping in the  $M_o$  assessment. Typically, we exclude items from the analysis to increase the reliability of a construct. Here, excluding items did not yield a significant improvement in the overall reliability.

In summary, supporting **H1**, the optimized pose estimation model  $M_o$  successfully decreased simulator sickness, see Fig. 7. The baseline  $M_b$  seemed to have created the least or almost no sickness, the optimized model  $M_o$  resulted in negligible sickness, while the simple model  $M_s$  caused much sickness.

### 4.2 Presence

Normality was violated for some measures of presence as assessed by a Shapiro-Wilk test. We therefore conducted Wilcoxon signed-rank tests for the presence measures comparing the simple model  $M_s$  and the optimized model  $M_o$  for the assessment after the tasks *W* and *R*. Fig. 9 holds the presence results.

In the walking task *W*, all 34 participants perceived a higher total presence with  $M_o$ , than with  $M_s$ . The distribution of the differences was symmetrical.  $M_o$  elicited a statistically significant median increase in total presence compared to  $M_s$ ,  $z = 5.090$ ,  $p < .001$ . Similarly,  $M_o$  elicited a statistically significant median increase in total presence compared to  $M_s$ ,  $z = 5.094$ ,  $p < .001$ , in the running task. In addition, participants perceived significantly higher presence with  $M_o$  than with  $M_s$ , with regard to all subscores ( $p's < .001$ ) in both tasks. Similar to the sickness assessment, the reliabilities of the presence factors were low. For example, the total score reliabilities ranged from  $\alpha = .241$  to  $\alpha = .476$ .

Supporting **H3**, the optimized model  $M_o$  resulted in a higher perception of presence in simulations with identical content but modified visuomotor coherence. In consequence, the data supports the assumed relation between presence and the realism of the simulation response to motoric actions [70].

### 4.3 Gait Parameters

We were interested in the comparison and separation of the pose estimation models (**H2**), their relation to the baseline measure, as well as their manifestation throughout the different motion tasks. Fig. 8 visualizes interpretable gait parameter measurements.

Assessed by a Shapiro-Wilk test, 21 out of 99 variable data were not normally distributed. As simulations showed that ANOVA is relatively robust to violations of the normality assumption [85, 86] and under the circumstance that the majority of the data was normally distributed, we conducted 3x3 repeated measures ANOVAs with the model ( $M_b$ ,  $M_o$ ,  $M_s$ ) and the task (*W*, *B*, *R*) serving as factors. When sphericity was violated, we applied Greenhouse-Geisser corrections.

Table 3 summarizes the test results. Except for the interaction effect of Model X Task for the *Max. Toe Clearance* parameter, all tests revealed significant effects ( $p's < .001$ ). The main effects for the factor model show that all assessed parameters have been significantly affected by the different models. As Fig. 8 depicts, there were especially strong differences between both pose estimation models  $M_s$  and  $M_o$ , which supports that gait is affected by the pose estimation models (**H2**). Exploratory pairwise comparisons (Bonferroni adjusted) revealed that

Table 2. Cronbach's  $\alpha$  reliability results of the SSQ assessments.

	N	O	D	Total
Baseline $M_b$	-.086 *#	.725	.666 #	.833 #
Simple Model $M_s$	.780 #	.648	.817	.863 #
Optimized Model $M_o$	.494 #	.677	.443	.738 #

Note. \* indicates a violation of reliability assumption.

# indicates that items were excluded because of zero variance.



Fig. 8. Mean values of all gait parameters for the 3 models (*baseline*  $M_b$ , *simple model*  $M_s$ , *optimized model*  $M_o$ ) and the 3 movement types (Walking, Balancing, and Running). Error bars indicate 95% confidence intervals. Movement types are ordered from slowest to fastest.

Table 3. Univariate Main and Interaction Effects.

Parameter	Model X Task					Model					Task				
	<i>F</i>	<i>df</i>	Error	<i>p</i>	$\eta_p^2$	<i>F</i>	<i>df</i>	Error	<i>p</i>	$\eta_p^2$	<i>F</i>	<i>df</i>	Error	<i>p</i>	$\eta_p^2$
Heel Strike Angle [deg]	385.96	2.61	62.66	***	.941	19.10	2.00	48.00	***	.443	1829.55	1.57	37.65	***	.987
Max Toe Out Angle [deg]	12.37	1.96	47.03	***	.340	315.19	2.00	48.00	***	.929	210.77	1.40	33.63	***	.898
Gait Velocity [m/s]	101.16	2.25	54.01	***	.808	703.51	2.00	48.00	***	.967	4860.70	1.59	38.07	***	.995
Max Lateral Excursion [cm]	265.00	4.00	96.00	***	.917	190.54	2.00	48.00	***	.888	235.49	2.00	48.00	***	.908
Toe Clearance [cm]			n.s.			411.98	2.00	48.00	***	.945	41.17	2.00	48.00	***	.643
Stance Time [%]	783.59	2.62	62.87	***	.970	33.60	2.00	48.00	***	.583	795.83	1.50	35.97	***	.971
Stance Time [s]	752.20	2.27	54.41	***	.969	28.45	1.58	37.82	***	.542	9914.76	1.95	46.78	***	.998
Stride Length [cm]	30.92	4.00	96.00	***	.563	633.88	2.00	48.00	***	.964	766.691	2.00	48.00	***	.970
Stride Time [s]	7.67	2.90	96.73	***	.252	474.48	2.00	48.00	***	.952	5929.75	2.00	48.00	***	.996
Swing Time [%]	428.28	2.50	59.91	***	.947	32.66	2.00	48.00	***	.576	474.03	2.00	48.00	***	.952
Swing Time [s]	266.88	2.29	45.91	***	.917	178.34	1.53	36.83	***	.881	266.88	2.29	45.91	***	.997

Note. \*\*\*  $p < .001$ ; n.s. = non significant. Where the sphericity assumption was violated, Greenhouse Geisser corrected values are reported.

except for the gait velocity, all parameters were significantly different between  $M_b$  and  $M_s$ , whereas 5 parameters (*Toe Out Angle*, *Toe Clearance*, *Stance Time [%]*, *Stance Time [s]*, and *Swing Time [s]*) were significantly different comparing  $M_b$  and  $M_o$ ;  $ps > .05$  (Bonferroni adjusted). Overall, the differences between  $M_b$  and  $M_o$  are much smaller compared to those between  $M_b$  and  $M_s$ . The optimized Model  $M_o$  still differs compared to the  $M_b$ , yet improved the accuracy compared to  $M_s$ .

Looking at how differences distribute among the three movements, the results also show that the task (*W, B, R*) significantly affected each parameter. Furthermore, the interactions show that even tendencies and the direction of differences can change throughout the different tasks. For instance, the *Stance Time* for  $M_s$  was higher for the *B* and *W* task, whereas it was lower in the *R* task compared to the  $M_b$  and  $M_o$ , see

Fig. 8.

A straightforward assessment of sickness through gait parameters would require a constant tendency of difference. While all parameters showed significant differences amongst the different models and tasks and may, therefore, provide useful information for a sickness classification, only five parameters (*Toe Out Angle*, *Gait Velocity*, *Max. Toe Clearance*, *Stride Length*, and *Stride Time*) showed a stable difference tendency between the simple model  $M_s$  and both,  $M_b$  as well as the  $M_o$  throughout all tasks (*W, B, R*). We interpret that these parameters are candidates for sickness indicators by comparing a baseline measure and a VR exposition measure of gait motion. That partly answers **RQ**, as a baseline versus an exposure parameter assessment of the five aforementioned parameters may provide indications for simulator sickness.

Yet, one would need to assess the same motion task not to bias the result. While a quantification of this assessment would need further systematic studies, we exploratory investigate whether or not there is a linear correlation of the gait assessment and simulator sickness, which would then allow for a single assessment measure, i.e., if we can show that gait parameters and sickness correlate linearly, we could assume that a single assessment leads to a reliable result.

#### 4.4 Correlations between Gait and the SSQ

The five parameters *Toe Out Angle*, *Gait Velocity*, *Max. Toe Clearance*, *Stride Length*, and *Stride Time* showed a stable impact of the simple model  $M_s$  across the tasks Walking, Balancing, and Running compared to the baseline  $M_b$  and  $M_o$ , see Fig. 8. Therefore, these parameters qualify for a potential single assessment measure. We assessed inter-individual (i.e., within one condition between the participants) correlations for these parameters between *W, B, and R* and the SSQ factors for the simple pose estimation model  $M_s$ , which results in the highest sickness (see Table 4). Multiple small to medium correlations arise from the analyses. However, there is no clear overall image of

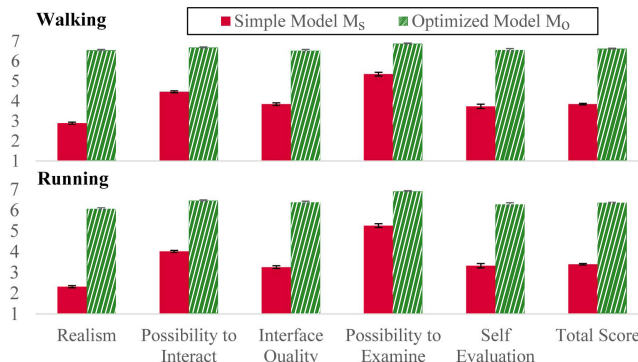


Fig. 9. Mean values of the PQ [89] measures for the Walking (*W*) and Running (*R*) assessment. Error bars denote the standard error.

the correlations. Thus, to this point, it seems that we cannot provide indications for a single assessment measure when taking the SSQ scale from [29] as a reference.

## 5 SICKNESS CLASSIFICATION

Whereas only five gait parameters (*Toe Out Angle*, *Gait Velocity*, *Max. Toe Clearance*, *Stride Length*, *Stride Time*, see abbreviations in Table 5) show a constant and linear relation, all others show very strong interaction effects (see Fig. 8) and hence, their relationship cannot be modeled linearly. Hence, we demonstrate that non-linear gait parameters can nevertheless add useful information to enable classification of sickness, improve robustness to outliers, and result in higher overall confidence. The underlying assumption is that a larger feature space describes the problem at hand in more detail and hence forces a data-driven classifier to accurately and precisely adjust its parameters to provide a better solution. To find a relationship between a user's movement, his/her simulator sickness, and the perceived presence, we state that if we can train a classifier that maps known gait parameters to a corresponding total sickness score we can also predict such sickness scores from unseen gait parameters, potentially at runtime. Below we first discuss how we obtain an optimal robust classifier that uses all gait parameters. Sec. 5.2 then introduces how we intuitively determine binary (not/sick) and 4-level sickness scores (not/slightly/sick/strongly).

### 5.1 Non-Linearities

To prove the results from Sec. 4.4 we show that we have to deal with non-linearity effects in our gait parameters. We first applied an SVM<sup>1</sup> (we implement the commonly used library from [7]) with a linear kernel function [54] to the (linear and non-linear) gait parameters, see the corresponding column in Table 5. Higher values indicate a better separability, e.g., *TOA* is highly linear (87.9%) to its corresponding sickness level whereas the *SWT<sub>p</sub>* is highly non-linear (26.2%). Second, we applied a solver with a quadratic SVM<sup>2</sup> and a cubic SVM<sup>3</sup> kernel function, see respective columns in Table 5. Whenever the SVM<sup>2,3</sup> accuracy scores are higher than the SVM<sup>1</sup> values our non-linear solvers find additional (important and non-linear dependency) information in the gait parameters that are likely to improve the reliability and robustness of the classification.

### 5.2 Feature Selection and Classifier Details

Each user provides 11 gait parameters for each of the three movement types, i.e., 11·3=33 feature values. We use their means (see Table 5) to create feature vectors, each with a size of 33×1, per model (*M<sub>s</sub>*, *M<sub>o</sub>*, and *M<sub>b</sub>*) or sickness score. In total, we have 3·33 input features to train an SVM classifier and to map these features to the corresponding model

Table 4. Pearson correlations between the SSQ dimensions and the most stable gait parameters for the simple model *M<sub>s</sub>*.

Task	Parameter	<i>N</i>	<i>O</i>	<i>D</i>	Total
Walking	Toe Out Angle	.069	.103	-.088	.014
	Gait Velocity	.077	-.015	-.045	-.002
	Max Toe Clearance	.296	.486*	.388	.422*
	Stride Length	.088	.369	.118	.197
	Stride Time	.110	.231	-.035	.085
Balance	Toe Out Angle	.104	.052	-.032	.034
	Gait Velocity	-.148	-.183	-.211	-.200
	Max Toe Clearance	.093	.243	.268	.227
	Stride Length	.300	.170	.230	.251
	Stride Time	.146	-.014	.186	.125
Running	Toe Out Angle	-.037	-.226	-.190	-.169
	Gait Velocity	-.386	-.341	-.170	-.304
	Max Toe Clearance	-.113	-.038	.089	-.007
	Stride Length	.116	-.041	.151	.144
	Stride Time	-.358	-.474*	-.395	-.440*

Note. \* indicates a significant correlation at the .05 level (2-tailed).

Table 5. Classification accuracy: separability of *M<sub>s</sub>* from *M<sub>b</sub>* and *M<sub>o</sub>*.

Features	Total Model Accuracy [%]		
	SVM <sup>1</sup>	SVM <sup>2</sup>	SVM <sup>3</sup>
Heel Strike Angle [deg] ( <i>HSA</i> )	29.8	89.1	98.2
Max Toe Out Angle [deg] ( <i>TOA</i> )	87.9	91.4	96.1
Gait Velocity [m/s] ( <i>GV</i> )	78.1	88.5	99.3
Max Lateral Excursion [cm] ( <i>MLE</i> )	63.2	86.5	97.8
Max Toe Clearance [cm] ( <i>MTC</i> )	81.4	89.7	98.4
Stance Time [%] ( <i>ST<sub>1p</sub></i> )	33.1	84.5	98.8
Stance Time [s] ( <i>ST<sub>1s</sub></i> )	29.8	87.9	99.1
Stride Length [cm] ( <i>SL</i> )	85.9	82.6	95.7
Stride Time [s] ( <i>ST<sub>2s</sub></i> )	81.7	90.5	99.3
Swing Time [%] ( <i>SWT<sub>p</sub></i> )	26.2	81.5	97.8
Swing Time [s] ( <i>SWT<sub>s</sub></i> )	57.1	89.6	99.5
* <i>TOA</i> , <i>GV</i> , <i>MTC</i> , <i>SL</i> , <i>ST<sub>2s</sub></i>	82.5	91.9	98.9
# <i>HSA</i> , <i>MLE</i> , <i>ST<sub>1p</sub></i> , <i>ST<sub>1s</sub></i> , <i>SWT<sub>p</sub></i> , <i>SWT<sub>s</sub></i>	36.8	87.4	96.3
all Gait Parameters ( <i>aGP</i> )	79.4	93.9	99.9

Note. \* indicates the gait parameters with linear dependencies.

Note. # indicates the gait parameters with non-linear dependencies.

or sickness score. Since we need a multi-class classification we use a one-vs-all SVM. We define a training set  $(\vec{x}_1, \vec{y}_1), \dots, (\vec{x}_n, \vec{y}_n)$  where a gait parameter  $\vec{x}_i$  is mapped to a model  $\vec{y}_i \in \{M_b, M_s, M_o\}$ .

Our linear solver optimizes a classification problem by adjusting  $w$  (the normal vector to the hyperplane) and  $b$  (the bias, i.e., distance to the origin), so that  $y_i(\langle w, x_i \rangle + b) \geq 1$  describes a distinct hyperplane with a minimal quadratic norm  $\frac{1}{2}||w||^2$  and the largest distance between the models  $\vec{y}_i$  per gait parameter  $\vec{x}_i$ . The non-linear solvers (SVM<sup>2,3</sup>) also optimize the same problem by adjusting  $w$  and  $b$ , but this time they also optimize a kernel function  $K(x, y) = (\sum_{i=0}^n x_i y_i)^p$  of degree  $p$ , here  $p \in \{2, 3\}$ .

We train the classifiers on gait parameters from (randomly selected) 15 subjects and test the classifiers on 10 unseen subjects (left-out) to avoid overfitting and to demonstrate robustness.

### 5.3 Results

The SVM<sup>2,3</sup> classifiers yield the highest correct classification rate when all gait parameters are used SVM<sup>2</sup>=93.9% and SVM<sup>3</sup>=99.9%, i.e., by chance of 99 of 100 cases we can correctly classify the correct underlying model or corresponding total sickness score, even if given an unseen set of gait parameters. To process the 33 features, the SVM<sup>3</sup> classifier takes 198  $\mu s$  (per set). Both non-linear classifiers SVM<sup>2,3</sup> outperform the linear SVM<sup>1</sup> classifier on every features set, see Table 5, i.e., the higher accuracy suggests that every parameter contains important information, even the non-linear ones, so all contribute to an optimal solution. The *Swing Time* [%] shows the strongest effect (i.e., difference between linear: 26.2%; quadratic: 81.5%; cubic: 97.8%).

### 5.4 Classifying Sickness Levels

Table 6 shows the results of an SVM<sup>3</sup> classifier that maps a feature set of  $x_i$ =gait parameters to the  $y_i$ =simulator sickness score. We tested our classifier on two different target sets: a binary one (0.0=not sick, and >0.0=sick) and a multi-level one with 4 sickness levels (not sick: 0.0; slightly sick: 0.0-0.25; sick: 0.25-0.50; very sick: 0.5-1.0;). We derived the sickness scale as follows: we normalize the total simulator sickness scores to [0.0=not sick, 1.0=sick] between the maximal and the minimal sickness scores of all subjects that occurred within the pose estimation

Table 6. Classification accuracy of various sets of gait parameters on Binary or 4-Level ranges.

Features	Total Model Accuracy [%]	
	SVM <sup>3</sup> to binary	SVM <sup>3</sup> to 4-levels
all Gait Parameters ( <i>aGP</i> )	96.4	85.2
<i>aGP</i> , age	97.1	86.1
<i>aGP</i> , age, height	98.4	86.7
<i>aGP</i> , age, height, VR exp.	99.6	87.3

models ( $M_b$ ,  $M_o$ ,  $M_s$ ) and movement types (Walking, Balancing, and Running). The binary sickness scale intuitively indicates sickness if the total score is  $>0.0$ . Respectively, the 4-levels sickness scale also indicates sickness if the total score is  $> 0.0$ . However, we differentiated the total sickness scores ( $>0.0$ ) based on self-reported measures, i.e., subjects that reported to suffer from negligible sickness symptoms form one cluster (0.0-0.5) and subjects that reported to suffer from significant sickness symptoms form another cluster (0.5-1.0). To demonstrate the accuracy of our classifier we split the 0.0-0.5 cluster into 0.0-0.25 and 0.25-0.50 clusters.

The results in Table 6 show that using more parameters always yields higher accuracy. The accuracy is even better if we also take the age, height, and level of previous VR experience in addition to all gait parameters ( $aGP$ ) into account. For the binary sickness level, the resulting classifier almost always (99.6%) correctly determines if a user suffers from simulator sickness. In 85.2% of the cases, the 4-level classifier even correctly estimates the degree of simulator sickness. This demonstrates a direct relation between gait parameters and a corresponding sickness level.

## 6 DISCUSSION

In summary, we presented two pose estimation models ( $M_s$  and  $M_o$ ), a simple model that results in high simulator sickness and an improved model that results in lower simulator sickness by utilizing a Kochanek-Bartels-Spline supported motion estimation.

Confirming our hypothesis **H1**, we showed that simulator sickness of users as assessed by the SSQ was lower with  $M_o$  than with  $M_s$ . Despite the low reliabilities of the instrument, we believe that these results capture what is rational and in line with previous research [26,27,35,37], i.e., that the pose estimation model  $M_s$  that results in more latency and a restricted mapping of human senses had the tendency to cause more simulator sickness. Given the fact, that the SSQ does not drop quickly between tests (but compounds over time and hence the source of the sickness in a within-subject design is hard to denote) and thus, a first test may bias a next one, and given the fact that post-test scores are always significantly higher than pre-test ones [41], we still found significant variations in the total scores. In addition, we found that the sickness increases one-way (i.e.,  $M_o$  to  $M_s$  and  $M_b$  to  $M_s$ ) and stops (i.e.,  $M_s$  to  $M_o$ ), i.e., we exclude two-way as sickness did not increase in both  $M_o$  to  $M_s$  and  $M_s$  to  $M_o$ . However, an elaborate option would be to study the same participant over the course of consecutive days.

Supporting hypothesis **H2**, significant differences in the gait parameters confirm that different pose estimation models strongly affect the assessed gait parameters. We interpret this as supporting evidence for our assumption that simulator sickness is related to the resulting gait differences. By inspecting the parameters we identified five parameters, namely *Toe Out Angle*, *Gait Velocity*, *Max. Toe Clearance*, *Stride Length*, and *Stride Time*, that showed stable differences amongst the tasks (Walking, Balancing, Running within  $M_b$ ,  $M_s$ , and  $M_o$ ) and could, therefore, be used in future studies to assess sickness in pre-post measures. The results of the correlations did not reveal a simple linear relation, i.e., a single measure of the gait parameters may not clearly indicate different individual levels of sickness. However, this may be due to the sample size and the fact that we did not include the repeated measure assessment in these analyses. To this point, our results confirm the weaknesses of perceptual measures to assess simulator sickness. In consequence, we argue that a gait assessment to identify simulator sickness can be useful, as lower reliabilities in the subjective SSQ measures arise frequently, even in controlled scenarios [90].

Confirming hypothesis **H3**, perceived presence was higher for the optimized model  $M_o$  compared to the simple model  $M_s$ . Thus, a more accurate mapping of motor behavior and visual feedback, i.e., a better matching of the human senses and thus a higher immersion, that seems to hinder simulator sickness, lead to a higher presence perception. Thus, we provide supporting evidence that inaccurate pose estimation models hinder the perception of presence. However, we cannot argue that optimized pose estimation models as such lead to a higher presence, as the perception of presence is affected by many factors, including simulation content, self-perception, and narrative [20,39,58]. Again,

the scale reliabilities restrict further conclusions and therefore more systematic studies (without complex cognitive tasks) using alternative methods of assessment are needed.

Despite the fact, that cybersickness, simulator sickness, and motion sickness [76] may differ, our gait parameter results are especially conclusive reflecting real-world examples and relating symptoms. That is, a sensory mismatch leads to a disturbance in gait behavior [80]. For example, perceiving motion sickness on a ship, sensory disturbances after a roller coaster ride, or sensory delays under the influence of alcohol or drugs does strongly affect gait behavior. With regard to a practical assessment, we thus argue that an assessment of a gait baseline in comparison to a VR exposure assessment can provide insights when using the five identified parameters and similar motion tasks which lead to an objective assessment of sickness, which answers **RQ**.

Furthermore, we showed that these gait parameters allow to identify the underlying pose estimation models they originate from. Extending this approach, we showed that using all assessed (even non-linear) gait parameters, i.e., more features, leads to a more precise classification. Beyond that, we showed that multiple gait parameters clearly indicate different pose estimation models and hence, allow to map gait parameters on one of four corresponding individual levels of sickness. Thus, if there is the possibility to assess simulator sickness before and after a walk of a specific user, our method has the potential to act as an objective sickness measurement tool for this specific user. Although the accuracy of our classification results is rather optimistic as it *only* classifies (separates) two models whose data points are *unnaturally* far apart anyway, we think that our data-driven approach may also yield accurate results even in uncontrolled scenarios with well-defined walking gait parameters and corresponding pre-post SSQ assessment. Because of the design of the presented experiment, the causality of the relation between gait impacts and sickness cannot be explained. We argue that the resulting sickness effects stand in strong relation to, and at least to certain degree affect the gait cycle. While one could argue that the gait cycle is directly affected by the disturbance of sensorimotor contingencies, we cannot provide insights into the relations on the basis of our data. However, we think that motion parameters, in our case, gait parameters, can be used to objectively assess sickness effects, as a relation is pointed out by our results.

### 6.1 Limitations

First, our tasks and the design was fairly complex and some measures were not assessed precisely at the same time, which could have biased the results. Although we assessed gait parameters only from *W*, *B*, *R*, the slalom task could have also affected the simulator sickness measures and may have different results for the gait analysis. However, we specifically designed this task as a representative task for behavior in large-scale simulations, e.g., sports applications, rehabilitation applications, or the visit of a virtual museum [63]. Furthermore, both pose estimation models had identical task exposures and thus, can still be reliably compared.

Second, our sample size was relatively small, and hence care should be taken when generalizing the result. Therefore, future replications should increase the sample size and validate the model. Obviously, it is difficult to generalize the current concept to applications not "walking in a straight line". Large-scale VR tracking systems may enable people to walk freely and allow for more straight-forward movement because of more space. However, if we find relations of sickness and motion parameters in a straight-line gait assessment, we may also find similar parameters in less controlled scenarios and of alternate tracking data (HMD IMU/position only). Yet, we argue that at the time of the current research, an exploratory study in an uncontrolled way (free movement, generic application) would have had too many potentially confusing variables to consider.

Third, our measures are drawn from a large-scale system (about  $45 \times 35$  m) that fuses IMU data with a large scale RTLS input. Logically, we cannot generalize the model to room scale or desktop VR [59,75]. However, larger-scale tracking systems will become available (e.g., radio-based [83] or optical systems such as Qualisys).

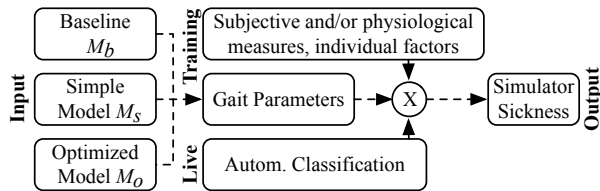


Fig. 10. The proposed pipeline for a future method.

## 7 CONCLUSION

We report novel findings on (1) the impact of simulation characteristics on gait parameters, simulator sickness, and presence, (2) the interrelationship of simulator sickness, presence, and gait, and (3) two approaches, namely a pre-post assessment of gait parameters as well as a classification approach to assess sickness through gait analysis. We developed a simple, unnatural pose estimation model that introduces a varying and higher latency to the simulation and an improved optimized pose estimation model that introduces less latency and includes a more accurate mapping of human senses. These models allowed us to conduct a controlled study with known simulation characteristics. Evaluating the impacts on gait parameters, simulator sickness, and latency, we found several notable effects. As hypothesized, the optimized model successfully decreased sickness and an increased presence. Compared to a baseline, gait parameters were most significantly impacted by the simple model. With regard to the proposed model as depicted in Fig. 1, our results, therefore, show that an unnatural mapping of senses seems to hinder immersion, leads to higher sickness, and to lower presence. We contributed a first approach to measurements for simulator sickness, which can either be applied to compare a baseline and exposure assessment or using our classification method and thus, extend existing research with important guidelines. Our insights are of great importance to both VR researchers and developers and will benefit the VR community, as our simulator sickness real-time measurement will be useful in future simulator sickness research and their application may allow dynamic content adaptation in response to a sickness increase.

### 7.1 Future Work

As we cannot estimate a precise level of sickness based on our results, we suggest that future work should assess more fine-grained gradations of simulation characteristics. Assessing multiple points of measure, and generalizing the model (trained on many users) could lead to a precise differentiation between levels of simulator sickness. Different subjective assessments such as trait measures, the History SSQ [28], individual factors [35], and physiological data may support systematic evaluations. Furthermore, future work may utilize the presented results to detect simulator sickness based on either gait parameters, or any other motion data, such as raw motion data from the users' head transformation (HMD), see Fig. 10. As we can robustly classify the models in firm real-time, our results suggest that gait parameters (or motion parameters in general) can be reliably used to establish a real-time assessment, which would be of great value for all VR applications. Thus, based on a more fine-grained training set and including individual factors, the model could then be trained on any motion parameters to allow for a real-time assessment that generalizes for all users (i.e., users that are unknown to the model). Moreover, it has been argued that gender differences in simulator sickness symptoms should be considered (see Graeber et al. for a review [14]). Yet, findings are diverse [14, 56]. While gender did not significantly affect the results in our study, it should be considered as a variable in future research. Also, the effects of rest frames or points of references, e.g., the visualization of the avatar's feet, should be considered in respect to simulator sickness. As the gait parameters are not easily accessible to all VR developers, future work should consider using raw data of end effectors and a prediction based on machine learning to assess and classify sickness.

## ACKNOWLEDGMENTS

This work was supported by the Bavarian Ministry for Economic Affairs, Infrastructure, Transport and Technology and the Embedded Systems Initiative (ESI). We thank Julius Hannink and Felix Kluge for their help in extracting the gait parameters, the Fraunhofer IIS students for their support in organizing and monitoring the measurement campaign, and all participants in our studies.

## REFERENCES

- [1] R. Baker. The history of gait analysis before the advent of modern computers. *Gait Posture*, 26(3):331–342, 2007.
- [2] J. Barth, C. Oberndorfer, C. Pasluosta, S. Schüle, H. Gassner, S. Reinfelder, P. Kugler, D. Schuldhuis, J. Winkler, and J. Klucken. Stride segmentation during free walk movements using multi-dimensional subsequence dynamic time warping on inertial sensor data. *Sensors*, 15(3):6419–6440, 2015.
- [3] R. J. V. Bertin, C. Collet, S. Espié, and W. Graf. Objective measurement of simulator sickness and the role of visual-vestibular conflict situations. In *Driving Simulation Conf. North America*, pp. 280–293. Berlin, Germany, 2005.
- [4] F. Biocca and B. Delaney. Immersive virtual reality technology. *Comm. in the Age of Virtual Reality*, 15(1):32–36, 1995.
- [5] A. E. Boone, M. H. Foreman, and J. R. Engsborg. Development of a novel virtual reality gait intervention. *Gait Posture*, 52(1):202–204, 2016.
- [6] A. Borrego, J. Latorre, R. Llorens, M. Alcaniz, and E. Noe. Feasibility of a walking virtual reality system for rehabilitation: Objective and subjective parameters. *Jo. Neuro. Eng. Rehabil.*, 13(1):68–69, 2016.
- [7] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *Trans. on Intelligent Systems and Technology*, 2(3):1–27, 2011.
- [8] M. S. Dennison, A. Z. Wisti, and M. D'Zmura. Use of physiological signals to predict cybersickness. *Displays*, 44(2):42–52, 2016.
- [9] T. Feigl, C. Mutschler, and M. Philippsen. Head-to-Body-Pose Classification in No-Pose VR Tracking Systems. In *Proc. 25th Conf. Virtual Reality and 3D User Interfaces*, pp. 1–2. Reutlingen, Germany, 2018.
- [10] T. Feigl, C. Mutschler, and M. Philippsen. Human Compensation Strategies for Orientation Drifts. In *Proc. 25th Conf. Virtual Reality and 3D User Interfaces*, pp. 1–8. Reutlingen, Germany, 2018.
- [11] A. Field, J. Miles, and Z. Field. *Discovering statistics using R*. Sage publications, 2012.
- [12] M. J. Georgiades, M. Gilat, K. A. Ehgötz Martens, C. C. Walton, P. G. Bissett, J. M. Shine, and S. J. Lewis. Investigating motor initiation and inhibition deficits in patients with parkinson's disease and freezing of gait using a virtual reality paradigm. *Neuroscience*, 337(1):153–162, 2016.
- [13] E. Goffman. The presentation of self in everyday life. *Garden City, NY*, 14(1):27–50, 1959.
- [14] D. A. Graeber and K. M. Stanney. Gender differences in visually induced motion sickness. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, pp. 2109–2113. Los Angeles, CA, 2002.
- [15] T. Grün, N. Franke, D. Wolf, N. Witt, and A. Eidloth. A real-time tracking system for football match and training analysis. *Microelectronic Systems: Circuits, Systems and Applications*, 114(3):199–212, 2011.
- [16] J. Habgood, D. Moore, D. Wilson, and S. Alapont. Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique. In *Proc. 25th Conf. Virtual Reality and 3D User Interfaces*, pp. 111–121. Reutlingen, Germany, 2018.
- [17] K. M. Hamilton, L. Kantor, and L. E. Magee. Limitations of postural equilibrium tests for examining simulator sickness. *Aviation, Space, and Environmental Medicine*, 60(3):246–251, 1989.
- [18] J. Hannink, T. Kautz, C. F. Pasluosta, J. Barth, S. Schüle, K.-G. Gaßmann, J. Klucken, and B. M. Eskofier. Mobile stride length estimation with deep convolutional neural networks. *Jo. of Biomedical and Health Informatics*, 22(2):354–362, 2018.
- [19] D. He, F. Liu, D. Pape, G. Dawe, and D. Sandin. Video-based measurement of system latency. In *International Immersive Projection Technology Workshop*, p. 111. New York, USA, 2000.
- [20] C. Heeter. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1(2):262–271, 1992.
- [21] J. H. Hollman, R. H. Brey, T. J. Bang, and K. R. Kaufman. Does walking in a virtual environment induce unstable gait?: An examination of vertical ground reaction forces. *Gait & Posture*, 26(2):289–294, 2007.

- [22] J. H. Hollman, R. H. Brey, R. A. Robb, T. J. Bang, and K. R. Kaufman. Spatiotemporal gait deviations in a virtual reality environment. *Gait & Posture*, 23(4):441–444, 2006.
- [23] O. Janeh, G. Bruder, F. Steinicke, A. Gulberti, and M. Poetter-Nerger. Analyses of gait parameters of younger & older adults during (non-) isometric virtual walking. *Trans. on Visualization and Computer Graphics*, 26(3):331–342, 2017.
- [24] O. Janeh, E. Langbehn, F. Steinicke, G. Bruder, A. Gulberti, and M. Poetter-Nerger. Biomechanical analysis of (non-) isometric virtual walking of older adults. In *Virtual Reality (VR)*, 2017, pp. 217–218. New York, USA, 2017.
- [25] O. Janeh, E. Langbehn, F. Steinicke, G. Bruder, A. Gulberti, and M. Poetter-Nerger. Walking in virtual reality: Effects of manipulated visual self-motion on walking biomechanics. *Trans. on Applied Perception (TAP)*, 14(2):12, 2017.
- [26] A. Kemeny, P. George, F. Mérienne, and F. Colombet. New vr navigation techniques to reduce cybersickness. *Electronic Imaging*, 2017(3):48–53, 2017.
- [27] R. Kennedy, L. Hettinger, and M. Lilienthal. Simulator sickness. *Motion and Space Sickness*, 10(2):317–341, 1990.
- [28] R. S. Kennedy, J. E. Fowlkes, K. S. Berbaum, and M. G. Lilienthal. Use of a motion sickness history questionnaire for prediction of simulator sickness. *Aviation, Space, and Environmental Medicine*, 63(7):588–593, 1992.
- [29] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Intl. Jo. Aviation Psychology*, 3(3):203–220, 1993.
- [30] R. S. Kennedy and K. M. Stanney. Postural instability induced by virtual reality exposure: Development of a certification protocol. *Intl. Jo. of Human-Computer Interaction*, 8(1):25–47, 1996.
- [31] B. Keshavarz and H. Hecht. Validating an efficient method to quantify motion sickness. *Human Factors*, 53(4):415–426, 2011.
- [32] B. Keshavarz, B. E. Riecke, L. J. Hettinger, and J. L. Campos. Vection and visually induced motion sickness: how are they related? *Frontiers in Psychology*, 6(1):472–481, 2015.
- [33] H. G. Kim, W. J. Baddar, H. Lim, H. Jeong, and Y. M. Ro. Measurement of exceptional motion in vr video contents for vr sickness assessment using deep convolutional autoencoder. In *Proc. 23rd Symp. Virtual Reality Software and Technology*, pp. 36–43. Gothenburg, Sweden, 2017.
- [34] D. H. Kochanek and R. H. Bartels. Interpolating splines with local tension, continuity, and bias control. In *Proc. Siggraph Comp. Graphics*, pp. 33–41. Minneapolis, USA, 1984.
- [35] E. M. Kolasinski. Simulator sickness in virtual environments. In *Army Research Inst. for the Behavioral and Social Sc.*, pp. 1–192. Alexandria, VA, 1995.
- [36] I. L. B. O. Lansink, L. van Kouwenhove, P. U. Dijkstra, K. Postema, and J. M. Hijmans. Effects of interventions on normalizing step width during self-paced dual-belt treadmill walking with virtual reality, a randomised controlled trial. *Gait & Posture*, 58(4):121–125, 2017.
- [37] J. J. LaViola Jr. A discussion of cybersickness in virtual environments. *SIGCHI Bulletin*, 32(1):47–56, 2000.
- [38] J. Lee, P. Han, L. Tsai, R. Peng, Y. Chen, K. Chen, and Y. Hung. Estimating the simulator sickness in immersive virtual reality with optical flow analysis. In *Proc. SIGGRAPH Asia Posters*, pp. 16–18. Bangkok, Thailand, 2017.
- [39] K. M. Lee. Presence, explicated. *Comm. Theory*, 14(1):27–50, 2004.
- [40] M. W. Leitman. Manual for eye examination and diagnosis. *Micro-electronic Systems: Circuits, Systems and Applications*, 114(3):199–212, 2016.
- [41] R. Li, N. Peterson, H. J. Walter, R. Rath, C. Curry, and T. A. Stoffregen. Real-time visual feedback about postural activity increases postural instability and visually induced motion sickness. *Gait & Posture*, 50(2):332–345, 2018.
- [42] P. Lincoln, A. Blate, M. Singh, T. Whitted, A. State, A. Lastra, and H. Fuchs. From motion to photons in 80 microseconds: Towards minimal latency for virtual and augmented reality. *Trans. on visualization and computer graphics*, 22(4):1367–1376, 2016.
- [43] W. Lo and R. H. So. Cybersickness in the presence of scene rotational movements along different axes. *Applied Ergonomics*, 32(1):1–14, 2001.
- [44] A. Mannini, D. Trojaniello, A. Cereatti, and A. M. Sabatini. A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington’s disease patients. *Sensors*, 16(1):134–139, 2016.
- [45] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, Jr. Physiological measures of presence in stressful virtual environments. *Trans. Graph.*, 21(3):645–652, 2002.
- [46] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks Jr. Physiological measures of presence in stressful virtual environments. *Trans. on Graphics (ToG)*, 21(3):645–652, 2002.
- [47] M. Meehan, S. Razzaque, B. Insko, M. Whitton, and F. P. Brooks. Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. *Applied Psychophysiology and Biofeedback*, 30(3):239–258, 2005.
- [48] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks Jr. Effect of latency on presence in stressful virtual environments. In *Proc. Virtual Reality*, p. 141. New York, USA, 2003.
- [49] A. Mirelman, I. Maidan, T. Herman, J. E. Deutsch, N. Giladi, and J. M. Hausdorff. Virtual reality for gait training: Can it induce motor learning to enhance complex walking and reduce fall risk in patients with parkinson’s disease? *Jo. Gerontol A. Biol. Sci. Med. Sci.*, 66(2):234–240, 2011.
- [50] B. J. Mohler, J. L. Campos, M. Weyel, and H. H. Bühlhoff. Gait parameters while walking in a head-mounted display virtual environment and the real world. In *Proc. of EuroGraphics*, pp. 85–88. Berlin, Germany, 2007.
- [51] J. D. Moss, J. Austin, J. Salley, J. Coats, K. Williams, and E. R. Muth. The effects of display delay on simulator sickness. *Displays*, 32(4):159–168, 2011.
- [52] W. T. Nelson, M. M. Roe, R. S. Bolia, and R. M. Morley. Assessing simulator sickness in a see-through hmd: Effects of time delay, time on task, and task complexity. In *Virtual Reality*, pp. 1–145. Wright-Patterson, AFB, 2000.
- [53] J. G. Nutt. Higher-level gait disorders: An open frontier. *Movement Disorders*, 28(11):1560–1565, 2013.
- [54] I. Papavasileiou, W. Zhang, X. Wang, J. Bi, L. Zhang, and S. Han. Classification of neurological gait disorders using multi-task feature learning. In *Intl. Conf. on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 195–204. Berlin, Germany, 2017.
- [55] M. E. S. Pierre, S. Banerjee, A. W. Hoover, and E. R. Muth. The effects of 0.2hz varying latency with 20–100ms varying amplitude on simulator sickness in a helmet mounted display. *Displays*, 36(4):1–8, 2015.
- [56] L. Rebenitsch and C. Owen. Review on cybersickness in applications and visual displays. *Virtual Reality*, 20(2):101–125, 2016.
- [57] C. Regan. An investigation into nausea and other side-effects of head-coupled immersive virtual reality. *Virtual Reality*, 1(1):17–31, 1995.
- [58] H. Rheingold. Virtual reality. *Summit Books*, 14(1):27–50, 1991.
- [59] J. J. Rieser, D. A. Guth, and E. W. Hill. Sensitivity to perspective structure while walking without vision. *Perception*, 15(2):173–188, 1986.
- [60] M. Rietzler, F. Geiselhart, and E. Rukzio. The matrix has you: realizing slow motion in full-body virtual reality. In *Proc. 23rd Symp. Virtual Reality Software and Technology*, pp. 2–11. Gothenburg, Sweden, 2017.
- [61] A. Rizzo, S. T. Koenig, et al. Is clinical virtual reality ready for primetime? *Neuropsychology*, 31(8):877, 2017.
- [62] G. Robillard, S. Bouchard, P. Renaud, and L. G. Cournoyer. Validation canadienne-française de deux mesures importantes en réalité virtuelle: l’immersive tendencies questionnaire et le presence questionnaire. *25e congrès annuel de la Société Québécoise pour la Recherche en Psychologie (SQRP), Trois-Rivières*, 25(7):753–759, 2002.
- [63] D. Roth, C. Kleinbeck, T. Feigl, C. Mutschler, and M. E. Latoschik. Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *Proc. of the 25th Virtual Reality (VR) Conf.*, pp. 331–342. Reutlingen, Germany, 2018.
- [64] J. C. M. Schlachetzki, J. Barth, F. Marxreiter, J. Gossler, Z. Kohl, S. Reinfelder, H. Gassner, K. Aminian, B. M. Eskofier, and J. Winkler. Wearable sensors objectively measure gait parameters in parkinson’s disease. *PloS one*, 12(10):6419–6440, 2017.
- [65] R. Skarbez, F. P. Brooks, and M. C. Whitton. A survey of presence and related concepts. *Compu. Surv.*, 50(96):1–39, 2017.
- [66] M. Slater. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence*, 8(5):560–565, 1999.
- [67] M. Slater. A note on presence terminology. *Presence Connect*, 3(3):1–5, 2003.
- [68] M. Slater. How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 13(4):484–493, 2004.
- [69] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Trans. of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

- [70] M. Slater, B. Spanlang, and D. Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *Trans. Graph.*, 29(4):54–63, 2010.
- [71] M. Slater, M. Usoh, and A. Steed. Depth of presence in immersive virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.
- [72] M. Slater, M. Usoh, and A. Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *Trans. Comp. Human Interaction*, 2(3):201–219, 1995.
- [73] A. J. J. Smith and E. D. Lemaire. Temporal-spatial gait parameter models of very slow walking. *Gait & Posture*, 61(10):125–129, 2018.
- [74] R. H. So, A. Ho, and W. Lo. A metric to quantify virtual scene movement for the study of cybersickness: Definition, implementation, and verification. *Presence: Teleoperators & Virtual Environments*, 10(2):193–215, 2001.
- [75] J. L. Souman, P. R. Giordano, M. Schwaiger, I. Frissen, T. Thümmel, H. Ulbrich, A. D. Luca, H. H. Bühlhoff, and M. O. Ernst. Cyberwalk: Enabling unconstrained omnidirectional walking through virtual environments. *Trans. Appl. Perception*, 8(4):1–22, 2011.
- [76] K. M. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is not simulator sickness. *Proc. of the Human Factors and Ergonomics Society annual meeting*, 41(2):1138–1142, 1997.
- [77] J. Stauffert, F. Niebling, and M. E. Latoschik. Towards comparable evaluation methods and measures for timing behavior of virtual reality systems. In *Proc. 22nd Conf. Virtual Reality Software and Technology*, pp. 47–50. New York, USA, 2016.
- [78] F. Steinicke, Y. Visell, J. Campos, and A. Lécuyer. Human walking in virtual environments. *Springer*, 56(7):976–985, 2013.
- [79] N. Stergiou and L. M. Decker. Human movement variability, nonlinear dynamics, and pathology: is there a connection? *Hum. Mov. Sci.*, 30(5):869–88, 2011.
- [80] M. Takahashi, A. Saito, Y. Okada, Y. Takei, I. Tomizawa, K. Uyama, and J. Kanzaki. Locomotion and motion sickness during horizontally and vertically reversed vision. *Aviation, Space, and Environmental Medicine*, 26(3):331–342, 1991.
- [81] K. E. Thomley, R. S. Kennedy, and A. C. Bittner Jr. Development of postural equilibrium tests for examining environmental effects. *Perceptual and Motor Skills*, 63(2):555–564, 1986.
- [82] J. D. Thompson and J. R. Franz. Do kinematic metrics of walking balance adapt to perturbed optical flow? *Hum. Mov. Sci.*, 54(1):34–40, 2017.
- [83] J. Tiemann, F. Eckermann, and C. Wietfeld. Atlas—an open-source tdoa-based ultra-wideband localization system. In *Proc. Intl. Conf. Indoor Positioning and Indoor Navigation*, pp. 1–6. Sapporo, Japan, 2016.
- [84] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr. Walking > walking-in-place > flying, in virtual environments. In *Proc. 26th An. Conf. Comp. Graphics and Interactive Techniques*, pp. 359–364. Los Angeles, USA, 1999.
- [85] M. W. Vasey and J. F. Thayer. The continuing problem of false positives in repeated measures anova in psychophysiology: A multivariate solution. *Psychophysiology*, 24(4):479–486, 1987.
- [86] R. R. Wilcox. *Introduction to robust estimation and hypothesis testing*. Academic press, 2011.
- [87] B. G. Witmer, C. J. Jerome, and M. J. Singer. The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 14(3):298–312, 2005.
- [88] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.
- [89] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. revised version by the uqo cyberpsychology lab. *Presence: Teleoperators and Virtual Environments. Revised version by the UQO CyberPsychology Lab*, 7(3):225–240, 2004.
- [90] P. Ziegler, D. Roth, A. Knotte, M. Kreuzer, and S. Mammen. Simulator sick but still immersed: A comparison of head-object collision handling and their impact on fun, immersion, and simulator sickness. In *Proc. 25th Virtual Reality (VR) Conf.*, pp. 1–12. Osaka, Japan, 2018.
- [91] D. J. Zielinski, H. M. Rao, M. A. Sommer, and R. Kopper. Exploring the effects of image persistence in low frame rate virtual environments. *Virtual Reality (VR)*, 7(68):19–26, 2015.