frontiers in Computer Science

TYPE Original Research
PUBLISHED 10 July 2026
DOI 10.3389/fcomp.2026.1853976

Check for updates

OPEN ACCESS

EDITED BY
Franca Delmastro,
National Research Council (CNR), Italy

REVIEWED BY
Kazuyuki Sato,
Friedrich Schiller University Jena, Germany
Wen Liang Yeoh,
Saga University, Japan

*CORRESPONDENCE
Jonas Pöhler
✉ jonas.poehler@uni-siegen.de

RECEIVED 12 April 2026
REVISED 16 June 2026
ACCEPTED 16 June 2026
PUBLISHED 10 July 2026

CITATION
Pöhler J, Vitt A, Wolling F, Michahelles F and Van Laerhoven K (2026) Wearable sensors reveal the reality gap: a full-body motion dataset for validating VR cycling simulators.
Front. Comput. Sci. 8:1853976.
doi: 10.3389/fcomp.2026.1853976

Wearable sensors reveal the reality gap: a full-body motion dataset for validating VR cycling simulators

Jonas Pöhler[1]*, Antonia Vitt[1], Florian Wolling[2], Florian Michahelles[2] and Kristof Van Laerhoven[1]

[1]Ubiquitous Computing, University of Siegen, Siegen, Germany, [2]Vienna University of Technology, Vienna, Austria

Virtual reality cycling simulators are increasingly used in research on urban mobility, rehabilitation, and road safety, yet their biomechanical fidelity remains under-validated due to the lack of standardized, publicly available datasets. We present a novel benchmark resource comprising synchronized full-body inertial measurement unit (IMU) data at 120 Hz and egocentric video from 10 participants who cycled an identical 1.4 km urban route in both Vienna and a motion-enabled VR simulator. Six body-worn sensors captured head, torso, arm, and leg movements, enabling detailed comparisons of pedaling rhythm, limb coordination, balance control, and visual attention patterns across environments. Our analyses reveal that VR successfully replicates fundamental locomotor patterns including pedaling cadence and bilateral leg coordination. However, significant differences emerge in upper-body dynamics: participants exhibited greater torso rotational variability in VR compared to real-world cycling, suggesting altered balance strategies. Head movement patterns showed broader yaw angles during virtual riding, yet these differences did not correlate with simulator sickness or immersion ratings. Emotional assessments indicated lower enjoyment and higher frustration in VR, highlighting gaps in affective realism despite mechanical similarity. This openly available dataset provides a critical resource for validating simulator designs, developing personalized VR training systems, and advancing research in embodied locomotion. The dataset, analysis tools, and documentation are freely accessible at: https://street2simulator.de/.

KEYWORDS

biomechanics, cycling, ecological validity, inertial measurement unit, motion analysis, simulator sickness, virtual reality, wearable sensors

1 Introduction

Cycling occupies a distinctive position in human locomotion research. Unlike walking or running, the lower limbs are mechanically constrained by the crank-pedal cycle, while the upper body and head remain largely free to respond to balance perturbations, steering demands, and environmental inputs. This structural distinction makes cycling an informative test case for studying motor adaptation: any perturbation introduced by a VR display, whether through visual latency, altered optic flow, or mismatch between vestibular and visual signals, must manifest in the segments that are free to respond, namely the torso, arms, and head.

VR cycling simulators have seen rapid growth in research applications, from road-safety studies (O'Hern et al., 2017; Hammami et al., 2025) to rehabilitation and physical training (McIlroy et al., 2021; Colombo et al., 2025). Compared to on-road studies, simulators offer controlled conditions, repeatable scenarios, and the ability to safely expose participants to hazardous situations. Yet their validity depends on the assumption that behavior and physiology transfer between environments (Sato et al., 2025). Systematic differences in kinematics would limit the generalizability of simulator-based findings and reduce the fidelity of training interventions that rely on movement quality rather than speed or cadence alone.

Existing validation work has examined behavioral outcomes such as speed, lateral position, and event responses (O'Hern et al., 2017; Hammami et al., 2025; Haasnoot et al., 2024), and physiological outcomes such as heart rate (Poli et al., 2024). Biomechanical studies of VR-real differences are largely concentrated in gait: Horsak et al. (2021) conducted a comprehensive analysis of full-body walking biomechanics in VR and found increased kinematic variability; Horsak et al. (2023) reported increased trunk variability during overground walking with a head-mounted display (HMD); Simonlehner et al. (2024) documented similar trends in a structured gait dataset. Cycling presents a different problem, however. The pedal constrains foot trajectory, reducing degrees of freedom in the lower limbs relative to walking, while the upper body must simultaneously manage handlebar steering, center-of-mass control, and head orientation. Currently no published study has simultaneously characterized multi-segment upper-body kinematics across matched VR and outdoor cycling conditions.

Wearable IMU sensors provide an accessible solution for multi-site movement measurement in ecological settings (Camomilla et al., 2018; Cerfoglio et al., 2023). Applied to cycling specifically, IMUs have been used to estimate knee joint angles (Cordillet et al., 2019), pedaling cadence in VR (Rojo et al., 2023), and segment-level kinematics during VR cycling tours (Prochazka et al., 2025). The present work extends this methodological line by deploying a six-site sensor array adapted from the Equimetrics system (Pöhler and Van Laerhoven, 2025) in both a VR simulator and real-world cycling on the same outdoor route.

A secondary dimension of this study concerns the subjective experience of VR cycling. Simulator sickness, operationalized through the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993), is a known confounder in VR studies, as symptom severity may alter movement strategies independently of the simulator's physical characteristics. The relationship between SSQ scores and objectively measured movement parameters has not been examined in a cycling context. Separately, presence, immersion, and affective response may explain variance in performance and movement that biomechanical metrics alone do not capture. Poli et al. (2024) showed that outdoor cycling produced higher motivation and enjoyment than VR cycling, a finding corroborated here with an extended questionnaire battery.

This paper is a direct extension of a four-page conference paper (Pöhler et al., 2025), which reported preliminary univariate analyses of head movement and provided an initial description of the dataset. The present study addresses three explicit research aims. First, we quantify differences in body segment kinematics between VR simulator and real-world cycling across six simultaneous IMU sites, with torso postural stability as the primary outcome. Second, we characterize the statistical distribution of IMU motion signals in each condition and assess whether distributional structure differs beyond what mean-level comparisons can reveal. Standard mean-and-variance comparisons assume normality and are insensitive to tail behavior, so distributional analysis can expose qualitative differences in motor control that summary statistics alone cannot capture. Third, we examine the relationship between objective movement metrics, simulator sickness, and subjective experience, including presence, achievement, and enjoyment. The present work makes the following additional contributions:

1. A distributional analysis of all IMU signals using a 97-distribution fitting procedure (Fitter library, Kolmogorov–Smirnov test), identifying condition-specific signal statistics beyond mean and variance.
2. Segment-by-segment Wilcoxon signed-rank tests across six body locations and six kinematic metrics, with effect sizes (Cohen's d) reported throughout.
3. Spectral analysis of head motion (high-frequency energy fraction >3 Hz) and its correlation with SSQ subscale scores.
4. Full questionnaire subscale analysis covering simulator sickness, presence/immersion, achievement, and comparative ratings.
5. Contextualization of all findings within the broader VR locomotion validity literature.

The dataset underlying this analysis is publicly available at https://street2simulator.de. Figure 1 provides an overview of the study design.

2 Related work

2.1 VR cycling simulator development and fidelity

Simulator validity research for cycling has grown alongside improvements in consumer VR hardware. O'Hern et al. (2017) established relative validity for a road-safety bicycle simulator using behavioral outcome measures; Haasnoot et al. (2024) validated a high-fidelity simulator incorporating lateral and roll motion against real-world reference data; Hammami et al. (2025) compared simulator performance against a matched real-world route using speed profiles and cumulative lateral position. Matviienko et al. (2023) examined four fidelity levels and found that steering control and pedaling control independently contribute to perceived realism, with their combination producing the most authentic experience. Across these studies, the common thread is that physical fidelity, particularly motion platform actuation and handlebar resistance, has disproportionate effects on behavioral validity compared to visual quality alone.

The specific simulator used in the present study incorporates a $\pm 5^{\circ}$ lateral tilt platform following the approach of Wintersberger et al. (2022), who showed that even weak tilt cues improved perceived realism over a fully static baseline. However, Bruce et al. (2025) demonstrated that incongruent VR

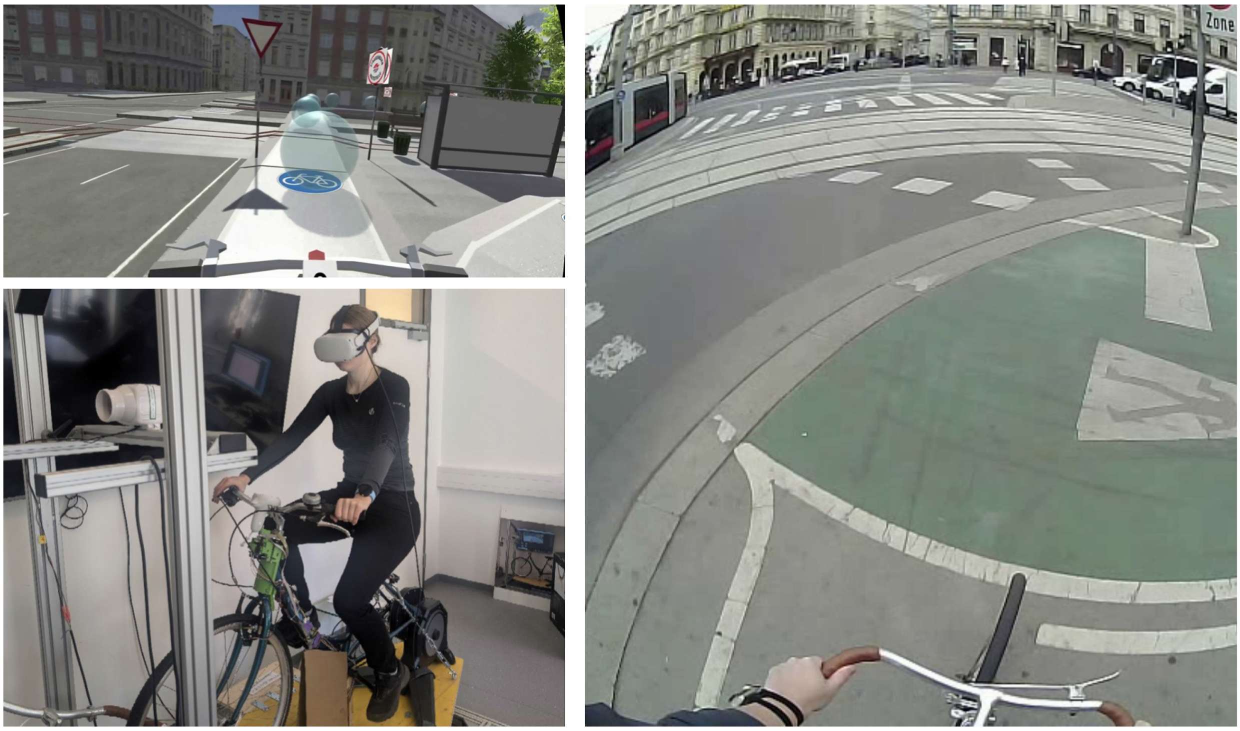

FIGURE 1
Study overview: cycling motion data were collected for the same route in VR **(left)** and in the real world **(right)**. Participants wore six body-worn IMUs and an egocentric camera.

cycling exercise revealed a role of perceived effort in cardiovascular control, raising the question of whether partial congruence, as in a simulator that provides some but not all physical cues, may produce unique physiological and biomechanical signatures not observed in either full congruence (real cycling) or full incongruence (static cycling).

A second family of cycling simulators, sometimes referred to as roller-based or smart-trainer simulators, sidesteps several of these issues by allowing the participant to ride their own outdoor bicycle on a set of motorized rollers or on a direct-drive smart trainer, typically while viewing a virtual environment on a flat display or, more recently, through an HMD. Commercial platforms such as Zwift and BKOOL combined with smart trainers (e.g., Wahoo KICKR, Tacx Neo) fall into this category, and research-oriented systems have used a similar architecture to study endurance training (McIlroy et al., 2021), cardiovascular response under congruent and incongruent visual conditions (Bruce et al., 2025), and motivational aspects of indoor cycling (Poli et al., 2024). The defining feature of these systems, from a biomechanical standpoint, is that the bicycle itself is the participant's own equipment: frame geometry, seat height, handlebar position, crank length, and bar grip shape are identical to outdoor riding by construction, and the only mechanical differences between indoor and outdoor are the support of the rear wheel by the trainer and the absence of true lateral roll. Roller-based systems are therefore a plausible alternative simulator design in which the upper-body and torso differences reported in the present study may not generalize: with the bicycle fit identical to outdoor and without an actuated tilt platform, the principal sources of postural perturbation are reduced to the HMD (if used) and the absence of self-balancing demands. We return to this point in the Section 5.

2.2 Biomechanical differences between VR and real locomotion

The most systematic biomechanical comparisons between VR and real locomotion come from gait research. Horsak et al. (2021) reported that full-body walking biomechanics were comprehensively altered in VR, with increased kinematic variability across multiple body segments. Horsak et al. (2023) confirmed that wearing an HMD increased trunk angular variability and reduced dynamic balance during overground walking. The *GaitRec-VR* dataset (Simonlehner et al., 2024) provides a large-scale reference for these effects across different gait speeds and populations. The convergent picture from gait literature is that HMD use and VR environment introduction increase proximal segment variability while distal segment (foot, shank) mechanics remain more constrained by the task.

Cycling presents a mechanically distinct case. Rojo et al. (2023) demonstrated that IMU sensors on the hip could estimate pedaling cadence in VR cycling with acceptable accuracy; Prochazka et al. (2025) extended this to full IMU analysis of VR cycling tours, characterizing segment-level kinematics across different terrain conditions. Poli et al. (2024) reported that outdoor cycling produced significantly higher heart rates, motivation, and enjoyment scores than VR cycling, pointing to systematic differences in physiological engagement even when the exercise modality is identical. The present study contributes a simultaneous six-site IMU measurement approach that extends prior single-sensor or lower-extremity-focused work.

2.3 Simulator sickness and head movement

Cybersickness in VR is a well-documented phenomenon that may compromise both participant wellbeing and data quality. Saredakis et al. (2020) conducted a meta-analysis of 50 studies ($N = 3{,}016$ participants) and found that locomotion-based VR content produced higher SSQ scores than passive viewing. Bimberg et al. (2020) critiqued the SSQ's validity for modern VR systems, noting that the questionnaire was developed for large-screen simulators and may not map cleanly onto HMD experiences. Nürnberger et al. (2021) proposed a Bayesian prediction error framework for VIMS, suggesting that accumulated sensory prediction errors, rather than a single conflict event, drive sickness onset. Palmisano et al. (2020) provided evidence that discrepancies between virtual and physical head pose are a proximal cause of cybersickness, with implications for how head motion should be measured and interpreted in VR studies. Park and Koo (2025) showed empirically that specific VR content characteristics modulate both cybersickness severity and the pattern of head movement, reinforcing the importance of content choice in sickness-sensitive protocols.

2.4 Presence, immersion, and motor transfer

The relationship between presence (felt sense of being in the virtual environment) and motor behavior has theoretical and practical importance for VR-based training. Colombo et al. (2025) showed that immersion and interaction during VR cycling influenced perceived effort and subjective experience, suggesting that the VR format can modulate physiological and psychological responses under the right conditions. Levac et al. (2019) reviewed motor skill transfer from VR, concluding that transfer is most reliable when sensorimotor contingency fidelity is high, that is, when the sensor-to-display mapping closely approximates real-world coupling. For cycling, this means that pedaling-to-speed coupling and steering-to-visual-heading coupling are critical for kinematic similarity. McIlroy et al. (2021) provided a SWOT analysis of virtual endurance cycling highlighting that while VR offers safety and weather independence, motivational disadvantages and HMD discomfort remain barriers to full ecological equivalence.

3 Materials and methods

3.1 Participants

Eleven participants were recruited from a university campus (nine male, two female; mean age $M = 24.4$, SD = 3.7 years; cycling experience score $M = 2.09$, SD = 0.70 on a five-point frequency scale, where 1 = never ride, 2 = a few times per year, 3 = a few times per month, 4 = a few times per week, and 5 = daily). Anthropometric measurements (height, body mass) were not recorded during the sessions and are acknowledged as a limitation (see Section 5.8). Prior VR experience was not formally assessed, but the participant pool was drawn from a university-aged cohort with general familiarity with consumer gaming technology; no participant reported regular use of head-mounted VR systems. One participant was excluded from quantitative analyses due to a technical recording failure (corrupted IMU timestamp file preventing session segmentation), yielding $N = 10$ complete trials. No other participants were excluded; all remaining recordings were complete across all six sensor nodes. All participants held valid cycling experience and were screened for susceptibility to motion sickness prior to enrolment. Written informed consent was obtained before each session. The study was approved by the ethics committee of the University of Siegen (reference LS_ER_03_2023).

3.2 Apparatus

3.2.1 VR cycling simulator

The simulator consisted of a stationary bicycle mounted on a motorized motion platform capable of $\pm 5°$ lateral tilt in the frontal plane, following the design principles described by Wintersberger et al. (2022). The handlebars were connected to a steering sensor, enabling realistic steering input. Brake pressure was captured via thin-film pressure sensors integrated into the braking system, facilitating real-time interaction in the virtual environment. An Oculus Quest 2 head-mounted display (HMD; Meta Platforms, Inc., Menlo Park, CA, USA; 1832 × 1920 per eye, 90 Hz refresh, $\approx 97°$ horizontal field of view, device mass approximately 503 g including the elastic strap) with 6 degrees-of-freedom tracking rendered a first-person perspective of a cycling route developed in Unity3D. The added mass of the HMD on the rider's head represents a non-negligible inertial load that may itself influence head and upper-trunk kinematics, independent of the visual content; this is considered further in the Sections 5, 5.8. The virtual route was modelled to match the physical characteristics (road width, surface type, and turn sequence) of the real-world route used in the outdoor condition, including landmarks and static obstacles, but did not incorporate dynamic traffic or pedestrians. Speed feedback was provided through the virtual scene's optic flow, driven by pedaling cadence via a smart direct-drive trainer that measured pedaling speed and adjusted resistance. No hand-force feedback was provided through the handlebars. Participants used the same model of road bicycle (*Rennrad*-style drop-bar road

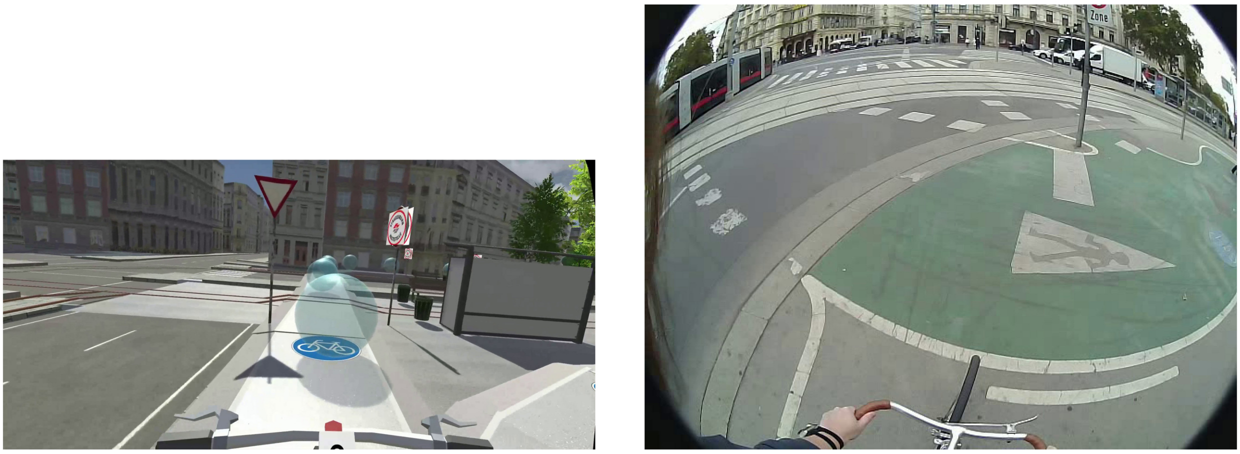

FIGURE 2
Comparison of the VR simulator environment (left) and the corresponding real-world cycling route in Vienna (right).

bike with 28-inch wheels, aluminum frame, approximate frame weight 9–10 kg, medium frame size ≈54–56 cm fitted to the average participant) in the VR and real-world sessions (Figure 2). Because the two bicycles were nominally identical in model, frame size, crank length, and handlebar geometry, these mechanical fit parameters can be regarded as matched across the two conditions by construction. Seat height was adjusted by the participant to a comfortable riding position at the start of each session and was nominally reproduced when transitioning between the VR rig and the outdoor bicycle; however, the adjusted seat height was not quantitatively recorded for each participant in each condition, so small between-condition mismatches of a few millimeters cannot be ruled out. This is acknowledged as a limitation (Section 5.8).

3.2.2 Real-world cycling

Participants cycled the same route on a standard bicycle on a paved urban road in Vienna under normal traffic conditions. The route was 1.4 km in length, encompassing typical urban features such as multilane roads, bike paths, historical streets, intersections, and tram tracks, with gentle curves and minimal elevation change (Figure 3). Participants were instructed to cycle at a self-selected comfortable pace and to complete the route once without stopping.

3.2.3 IMU system

The IMU system was adapted from the Equimetrics platform (Pöhler and Van Laerhoven, 2025). Six wireless sensor nodes were worn simultaneously (Figure 4). Sensor positions were selected to capture motion at each major body segment involved in cycling biomechanics (Camomilla et al., 2018; Cerfoglio et al., 2023): (1) head (right temporal region), (2) torso, (3) left forearm, (4) right forearm, (5) left shank (lower leg), and (6) right shank (lower leg). The torso position provides a stable attachment site proximal to the center of mass and is sensitive to both mediolateral and anteroposterior postural sway. The temporal head position captures full 3-axis angular rate of the head, which serves as an indirect indicator of visual search behavior based on head movements (Palmisano et al., 2020; Park and Koo, 2025) and is relevant to vestibular analysis. Shank (lower-leg) positions capture the rotational component of the pedaling cycle and allow cadence extraction from gyroscope signals, since the shank rotates through a larger angular range than the thigh during each pedal stroke (Rojo et al., 2023; Cordillet et al., 2019). Forearm positions capture handlebar-transmitted vibration and active steering movement.

Each node was a custom-built sensor module based on the InvenSense MPU-6050 6-axis IMU (TDK InvenSense, San Jose, CA, USA; accelerometer and gyroscope) sampling at 120 Hz, configured with a ±8 g accelerometer range and ±2000 deg/s gyroscope range. Sensors were attached using adjustable elastic straps. Participants were asked to check strap tightness immediately before each session and to report any sensor movement during the trial. Synchronization across nodes was achieved via a shared radio timestamp broadcast at session start, with data centralized on a Raspberry Pi carried by participants; clock drift was checked *post-hoc* and did not exceed 10 ms over any session. Egocentric video was recorded at 30 fps using a head-mounted eye tracker (Pupil Labs Core) for the real-world session and screen capture from the simulation software for the VR session. Video and IMU data alignment was verified post-collection by identifying specific events (e.g., initial pedaling movements) in both data streams.

3.3 Procedure

Each participant completed both conditions in a fixed order (VR first, followed by real-world cycling). Each condition was performed exactly once per participant, and both conditions for a given participant were carried out on the same day, with a minimum 30-min rest period between them during which the sensor placement was checked and the participant transitioned between the lab and the outdoor route. Before the VR session, participants were familiarized with the HMD and simulator for

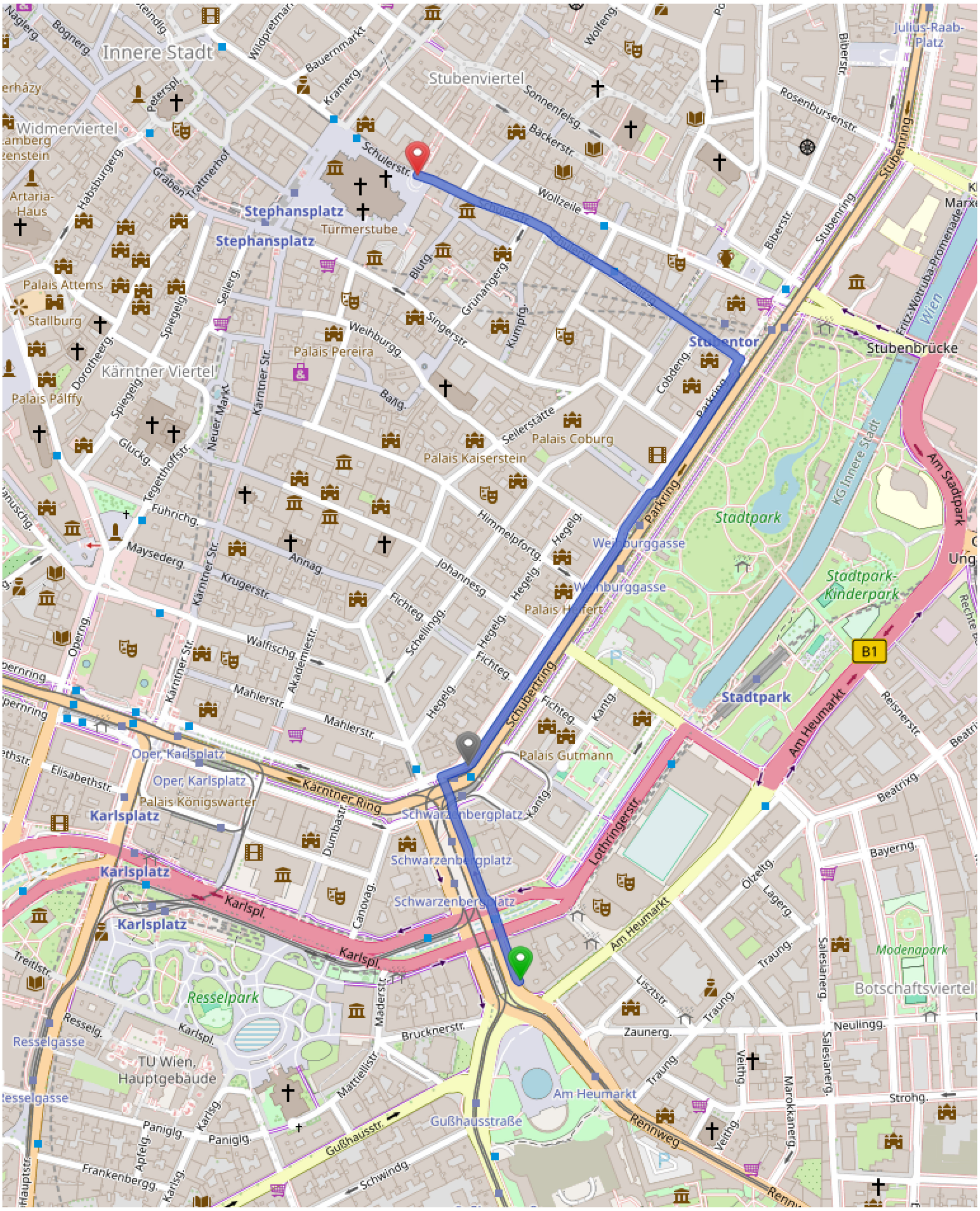

FIGURE 3
Map of the 1.4 km urban cycling route in Vienna used for both the real-world and VR conditions. Source: https://www.openstreetmap.org. The map was reproduced and adapted (route overlay added by the authors). Documentation is licensed under the Creative Commons Attribution-ShareAlike 2.0 license (CC BY-SA 2.0).

approximately 5 min; no formal practice trial was conducted. Safety measures included a spotter present during VR sessions and the use of a stationary bicycle frame on the actuated platform that prevented falls. Participants were not given instructions about how to use their body; they were told to cycle normally and to focus on the riding task. All data were collected during 1 week in October under non-rush-hour traffic conditions. The VR sessions took place in a temperature-controlled laboratory (approximately 21–23°C,

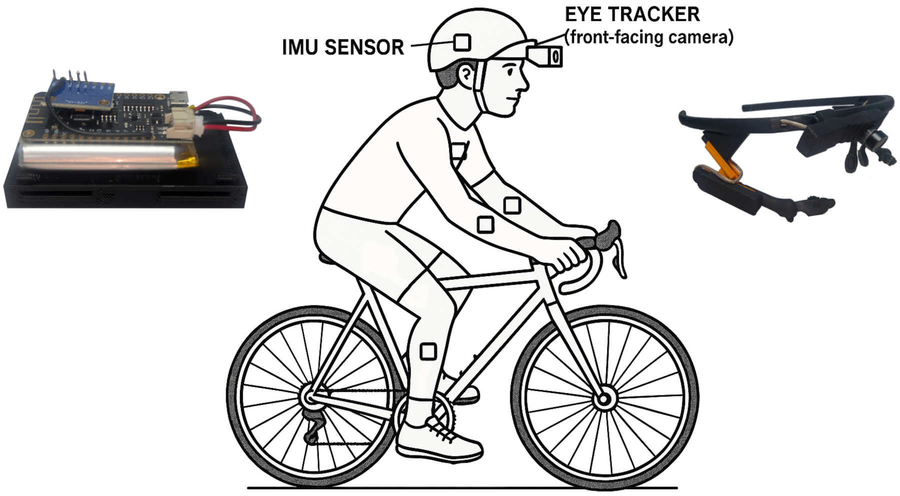

FIGURE 4
Sensor placement during data collection. Six wireless IMUs were strategically placed on the rider's body to capture whole-body motion data during both VR and real-world cycling sessions: (1) head (right temporal region), (2) torso (sternum/upper back), (3) left forearm, (4) right forearm, (5) left shank (lower leg), and (6) right shank (lower leg). A head-mounted eye-tracker (Pupil Labs Core) was used for egocentric video recording during the real-world session only; gaze data were not analyzed in the present study. All six IMU sites, including both shank positions, are indicated on the body illustration.

≈40–50% relative humidity). The outdoor sessions were conducted under mild autumn weather throughout the data collection week: ambient temperatures of 15–18°C, sunny to lightly overcast skies, light wind, and consistently dry asphalt surfaces. Session duration in VR was $M = 153$ s (SD = 27) and in real cycling $M = 354$ s (SD = 54), reflecting the self-paced protocol and the greater freedom of speed regulation outdoors. In total, the dataset comprises 20 cycling trials and 109 individual sensor streams (six sensors × 22 sessions minus instances lost to technical issues such as disconnected cables or sensor failures). The dataset includes synchronized raw tri-axial IMU data (120 Hz), egocentric video (30 fps), participant demographics, and questionnaire responses.

3.4 Questionnaires

3.4.1 Simulator sickness questionnaire

Simulator sickness was assessed using the SSQ (Kennedy et al., 1993) administered immediately after the VR session. The SSQ yields subscale scores for nausea, oculomotor disturbance, and disorientation, as well as a weighted total score; higher scores indicate greater sickness severity. Given the low total scores observed (see Section 4.8), subscale analyses were exploratory.

3.4.2 Presence and immersion

Felt presence and subjective immersion in the VR environment were assessed using a short four-item scale developed in-house for this study (items denoted *cvr1–cvr4*; rated on a 7-point Likert scale from 1 = strongly disagree to 7 = strongly agree). The items were constructed by paraphrasing recurring themes from the established Igroup Presence Questionnaire (IPQ; Schubert et al., 2001) and the Slater–Usoh–Steed presence questionnaire (Usoh et al., 2000), condensed into a short instrument suitable for end-of-session administration. The four items addressed: (cvr1) felt sense of being in the virtual environment, (cvr2) felt sense of cycling rather than operating a device, (cvr3) felt sense of the virtual environment being a real place, and (cvr4) felt sense of involvement in the experience. The composite presence/immersion score reported below is the sum of the four item ratings; the scale was not formally validated, and we therefore treat the resulting scores as exploratory descriptors rather than as a calibrated presence measure.

3.4.3 Achievement and affective response

A parallel five-item achievement and affective response questionnaire, also developed in-house for this study, was administered immediately after each of the two cycling conditions (VR items denoted *aes1–aes5*; real-world items *aer1–aer5*; rated on a 5-point Likert scale from 0 to 4). The items were constructed by adapting prompts in the spirit of the Intrinsic Motivation Inventory (Ryan, 1982) and the NASA-TLX (Hart and Staveland, 1988) into a short post-condition instrument that could be answered identically after both the VR and the outdoor session. The five items addressed: (1) enjoyment, (2) perceived effort, (3) frustration, (4) perceived competence, and (5) overall impression. A composite achievement score was computed as the sum of items 1–5 for each condition. As with the presence/immersion subscale, the achievement instrument was not formally validated; reported scores are therefore exploratory and should be interpreted in conjunction with the IMU-derived movement measures rather than as stand-alone validated psychometric outcomes.

3.5 Signal processing

Raw IMU data were loaded from binary log files and converted to physical units (deg/s for gyroscope, g for accelerometer). Session start and stop timestamps were determined by manual event markers logged by the experimenter at the moment the participant began and ended pedaling on the route, and were cross-verified *post-hoc* by detecting the first and last sustained pedaling cycles (defined as three consecutive shank-gyroscope peaks separated by intervals consistent with the participant's mean cadence) in the left-shank gyroscope signal. Each tri-axial signal was processed in two complementary ways: (i) each axis (x, y, z) was analyzed separately to obtain axis-specific descriptors such as per-axis standard deviation and yaw-range; and (ii) the three axes were combined into the orientation-invariant Euclidean norm, $\|a\| = \sqrt{a_x^2 + a_y^2 + a_z^2}$ (and analogously for the gyroscope), which was used for resultant acceleration and angular-rate variance metrics.

3.5.1 Pedaling frequency

Pedaling frequency was estimated from the left-shank gyroscope signal. Peaks corresponding to the downstroke were identified using a prominence-based algorithm (SciPy *find_peaks*). To exclude stopping phases (e.g., at intersections or while yielding to traffic), windows in which the shank-gyroscope magnitude fell below 15 deg/s for more than 2 s were removed from the peak series prior to cadence estimation, and the affected inter-peak intervals were not included in the variance computation. Acceleration phases (the first 10 s after each start or restart) and the subsequent steady-state segment were treated separately; the cadence and inter-peak interval variance values reported below refer to the steady-state segment only. Frequency was computed as the reciprocal of the mean inter-peak interval. Inter-peak interval variance was also retained as a measure of cadence regularity. Because the VR and outdoor sessions differed in duration, the number of detected pedal cycles also differed between conditions. Inter-peak interval variance, being a sample-variance estimator, is not unbiased with respect to sample size; however, because the variance estimator we used is the population-style mean of squared deviations rather than a Bessel-corrected estimator, its expectation does not depend systematically on the cycle count, and the comparison between conditions should not be biased by length differences. The modest number of cycles per participant remains a limitation and is acknowledged in Section 5.8.

3.5.2 Limb coordination

Bilateral limb coordination was quantified as the peak cross-correlation coefficient between the left- and right-shank gyroscope signals after normalizing each signal to zero mean and unit variance. Higher values indicate more synchronized alternating leg motion.

3.5.3 Upper-body and torso kinematics

For each sensor, the following descriptive statistics were computed per session: (1) variance of three-axis resultant acceleration, (2) variance of three-axis resultant angular rate, (3) standard deviation of the gyroscope signal on each individual axis. The torso gyroscope standard deviation was selected a priori as a primary outcome based on its theoretical sensitivity to postural sway. Upper-body metrics were averaged across the forearm sensors to provide a combined upper-limb measure.

3.5.4 Head motion

Head motion was characterized using the head mounted sensor. In addition to gyroscope standard deviation and range, the frequency composition of the head gyroscope signal was analyzed using the Fast Fourier Transform (FFT). The fraction of total spectral energy above 3 Hz was used as an index of high-frequency head movement, which may reflect tremor or instability rather than voluntary head-orientation behavior (Park and Koo, 2025).

3.5.5 Distributional analysis

To characterize the statistical properties of motion signals beyond mean and variance, the best-fitting probability distribution was identified for each participant, sensor, and condition combination using the Python *FITTER* library. The library fits 97 candidate distributions (SciPy *stats* module) via maximum likelihood estimation and selects the best fit using the Kolmogorov–Smirnov (KS) statistic. Distributions were grouped *post*-hoc into heavy-tailed (Cauchy family, including skew-Cauchy), moderately heavy-tailed (Laplace), and approximately normal (logistic, normal) categories for summary comparison across conditions.

3.6 Statistical analysis

All statistical analyses were conducted in Python 3.11 (Python Software Foundation, Wilmington, DE, USA) SciPy 1.11 (Open-source; cite as Virtanen et al. (2020)), Pingouin 0.5 (Open-source; cite as Vallat (2018)). For each paired comparison, the distribution of the within-participant difference scores was tested for normality using the Shapiro–Wilk test (applied to the paired differences, not to the two samples separately). When the paired differences were consistent with normality ($p > 0.05$), a paired t-test was used; otherwise the non-parametric Wilcoxon signed-rank test was applied. The Wilcoxon signed-rank test assumes that the distribution of paired differences is symmetric about the median; symmetry was inspected visually via histograms of the paired differences and considered acceptable in all reported cases. Effect sizes are reported as Cohen's d computed from the paired difference scores (for t-tests) or estimated from the z-statistic (for Wilcoxon tests) as $r = z/\sqrt{N}$, converted to d via the standard formula $d = 2r/\sqrt{1 - r^2}$. Cohen's conventions classify $|d| \geq 0.2$ as small, $|d| \geq 0.5$ as medium, and $|d| \geq 0.8$ as large effects.

For each paired comparison we additionally report the 95% confidence interval of the mean difference (VR − Real), computed from the within-participant differences using the Student t critical value with $n - 1$ degrees of freedom. For primary non-significant comparisons (pedaling frequency, head gyroscope standard deviation, head gyroscope range), we performed an additional two one-sided tests (TOST) procedure to evaluate statistical equivalence. The equivalence bound for pedaling frequency was set at ± 0.15 Hz (approximately ± 9 RPM), a band judged small but functionally meaningful for cycling motor control (Swart and Holliday, 2019). For the head motion variables, no a priori clinically meaningful bound was available, so we used ± 0.5 pooled standard deviations (Cohen's small-effect threshold) as a generic equivalence band. The two one-sided tests were each evaluated against $\alpha = 0.05$; equivalence is declared when both one-sided tests reject, equivalently when the 90% confidence interval of the mean difference lies entirely within the equivalence band.

Correlation between questionnaire scores and head motion was assessed using Pearson r with two-tailed p-values. Statistical significance was set at $\alpha = 0.05$; trends are noted at $\alpha = 0.10$. Given the exploratory nature of the study and the small sample size, results are interpreted with caution and effect sizes are considered alongside p-values. *Post-hoc* power analysis (Pingouin *power_ttest*) indicated that the primary comparison (torso gyro SD) was powered at approximately 0.77 with the observed $N = 10$ and $d = 0.963$, suggesting the significant result is reliable. Secondary comparisons with $d \approx 0.5$ require $N \approx 35$ for 80% power at $\alpha = 0.05$; non-significant results for those metrics should be treated as inconclusive rather than as evidence of equivalence. No correction for multiple comparisons was applied across the exploratory segment-level tests; adjusted α (Holm–Bonferroni) would be required in confirmatory replications.

4 Results

4.1 Pedaling frequency and cadence regularity

As visualized in Figure 5, inspection of the Euclidean norm signals from shank-mounted IMUs revealed clear periodic patterns linked to pedaling cycles, as well as characteristic movements related to balance and gaze behavior (e.g., head turns at intersections).

Pedaling frequency, estimated from the left-shank gyroscope, was lower in VR ($M = 0.59$ Hz, SD = 0.28) than in real-world cycling ($M = 0.82$ Hz, SD = 0.22). A paired t-test on $n = 9$ complete pairs (one participant missing valid bilateral data) did not reach significance [$t_{(8)} = -1.606$, $p = 0.147$, mean difference -0.28 Hz, 95% CI $(-0.68, 0.12)$], though the effect size was moderate ($d = -0.535$). To evaluate whether this non-significant result could nevertheless be interpreted as evidence of equivalence, we performed a TOST procedure with equivalence bounds of ± 0.15 Hz. The TOST procedure did not reject the null of non-equivalence ($p_{\text{TOST}} = 0.759$); the 90% confidence interval of the mean difference, $[-0.60, 0.04]$ Hz, extends beyond the lower equivalence bound, so we cannot conclude statistical equivalence at this bound either. The non-significant pedaling-frequency result is therefore inconclusive rather than supportive of equivalence. The pattern of lower cadence in VR is consistent with self-paced cycling: participants covered the outdoor route at a higher average speed, requiring a higher pedaling rate to maintain momentum.

Inter-peak interval variance was substantially higher in VR ($M = 52.07$, SD = 142.97) than in real cycling ($M = 0.44$, SD = 0.68), indicating more irregular cadence in the simulator. The large standard deviation in the VR condition is driven by a small number of trials with extended coasting or stopping events. A Wilcoxon signed-rank test was non-significant ($W = 11$, $p = 0.203$, $n = 9$), again likely reflecting insufficient power with the current sample size; the raw difference in means is nevertheless practically large.

4.2 Bilateral limb coordination

Cross-correlation of left- and right-shank gyroscope signals was available for $n = 4$ participants with complete bilateral shank data. Mean peak cross-correlation was $M = 0.332$ (SD = 0.205) in VR and $M = 0.403$ (SD = 0.275) in real cycling. The Wilcoxon test was not significant ($W = 3$, $p = 0.625$). With only four pairs, this result is underpowered and should be interpreted as preliminary; the directional trend is consistent with slightly more synchronized alternating leg motion in real-world riding.

4.3 Upper-body acceleration and angular rate

Upper-body acceleration variance did not differ significantly between conditions ($n = 10$; VR: $M = 0.069$, SD = 0.123; Real: $M = 0.072$, SD = 0.062). The Shapiro–Wilk test on the paired differences indicated non-normality ($W = 0.767$, $p = 0.006$), so the Wilcoxon signed-rank test was used [$W = 19$, $p = 0.432$, mean difference -0.003, 95% CI $(-0.108, 0.102)$, $d_{paired} = -0.02$]. Upper-body gyroscope variance showed a numerical trend in the same direction as previously reported but did not reach significance [$n = 10$; VR: $M = 1625$, SD = 1533; Real: $M = 725$, SD = 681; Shapiro–Wilk on differences $W = 0.856$, $p = 0.068$ (normal); paired $t_{(9)} = 1.664$, $p = 0.130$, mean difference 900 $(\text{deg/s})^2$, 95% CI $(-323, 2124)$, $d = 0.526$]. The direction of this trend, greater angular rate variance in VR, suggests more upper-body rotational movement in the simulator, though inter-participant variability was high and the present sample is underpowered for a medium effect.

4.4 Torso angular rate standard deviation

The torso gyroscope standard deviation was the primary kinematic outcome and yielded the clearest between-condition difference. VR produced substantially higher torso angular rate variability ($M = 26.76$ deg/s, SD = 15.25) compared to real-world cycling ($M = 12.36$ deg/s, SD = 3.77). The Shapiro-Wilk test on the paired differences was consistent with normality ($W = 0.971$, $p = 0.902$), and the paired t-test was significant [$t_{(8)} = 2.889$, $p = 0.020$, $n = 9$ participants with valid torso data in both conditions; mean difference 14.4 deg/s, 95% CI $(2.9, 25.9)$] with a large effect size ($d = 0.963$). The real-world condition also showed much lower

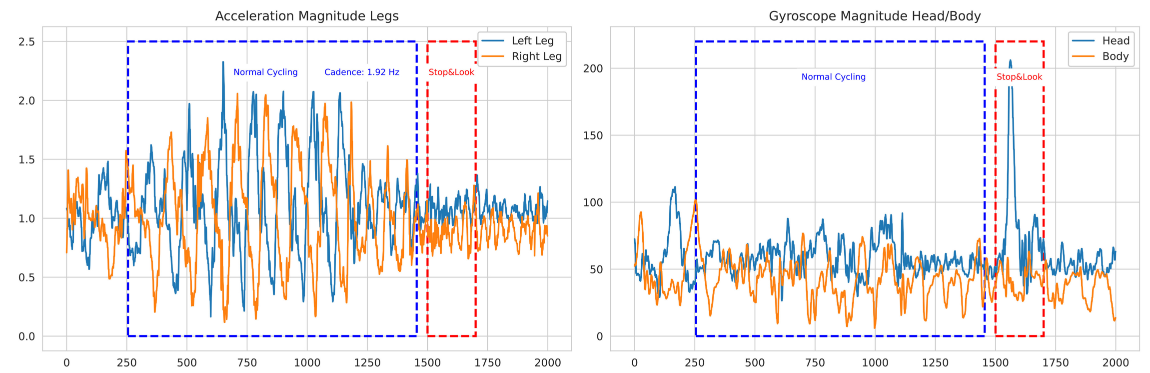

FIGURE 5
Example of pedaling rhythm visible in the IMU data for a single representative trial. The **left panel** shows acceleration magnitude of the legs (vertical axis: magnitude in *g*; horizontal axis: time in samples at 120 Hz); the **right panel** shows gyroscope magnitude of the head and body (vertical axis: angular rate magnitude in deg/s; horizontal axis: time in samples at 120 Hz). The highlighted segment marks an intersection where the participant turned their head to check for traffic.

inter-participant variability (SD of 3.77 vs. 15.25 deg/s), suggesting that outdoor cycling produces a more consistent postural pattern across individuals.

Segment-level Wilcoxon tests across all six body locations confirmed that differences were particularly pronounced at the torso and left-leg positions. Body (torso) median gyroscope rate differed significantly between conditions ($W = 2.0$, $p = 0.012$; VR: $M = -0.308$, Real: $M = -0.088$, $d = -1.501$). Left-leg median gyroscope rate also differed ($W = 6.0$, $p = 0.027$; VR: $M = -0.327$, Real: $M = -0.130$, $d = -0.949$). These results are summarized in Table 1.

4.5 Head motion

Head gyroscope standard deviation did not differ significantly between conditions [$n = 8$ participants with complete head sensor data; VR: $M = 17.81$ deg/s, $SD = 19.56$; Real: $M = 28.07$ deg/s, $SD = 14.27$; paired $t_{(7)} = -1.414$, $p = 0.200$, mean difference -10.3 deg/s, 95% CI $(-27.4, 6.9)$, $d = -0.500$]. The corresponding TOST equivalence test against bounds of ± 8.56 deg/s (± 0.5 pooled SD) did not reach equivalence [$p_{\text{TOST}} = 0.589$; 90% CI $(-24.0, 3.5)$]. Similarly, head gyroscope range was not significantly different [$n = 8$; VR: $M = 161.9$ deg/s, $SD = 177.7$; Real: $M = 235.9$ deg/s, $SD = 141.7$; paired $t(7) = -0.867$, $p = 0.415$, mean difference -74.0 deg/s, 95% CI $(-275.8, 127.8)$, $d = -0.307$]; the TOST equivalence test against bounds of ± 80.4 deg/s (± 0.5 pooled SD) also did not reach equivalence [$p_{\text{TOST}} = 0.471$; 90% CI $(-235.7, 87.7)$]. The non-significant head-motion results should therefore be treated as inconclusive rather than as evidence of equivalence between conditions.

Spectral analysis of head motion revealed that the fraction of total signal energy above 3 Hz was nearly identical between conditions ($n = 8$; VR: $M = 0.426$; Real: $M = 0.433$; $p = 0.950$). This finding indicates that VR does not selectively increase high-frequency head tremor or micro-oscillations relative to real-world cycling.

None of the head motion metrics correlated significantly with SSQ total score or subscale scores ($n = 8$; Pearson r range: 0.028–0.380; all $p > 0.3$). Table 2 presents the full correlation matrix.

4.6 Visual attention and immersion

Head movement serves as a key indicator of visual attention and immersive engagement during first-person locomotion tasks. Analysis of head orientation dynamics from the head sensor ($n = 8$ participants with complete head sensor data) revealed that the range of yaw angles was significantly higher in the VR condition compared to the real-world setting ($p < 0.05$), indicating that participants engaged in broader lateral head movements while cycling in the simulated environment. The overall frequency of head-turn events per minute was not significantly different between conditions ($p = 0.43$); this non-significant result should not be interpreted as evidence of equivalence given the small sample. These findings suggest that while the overall frequency of scanning behavior remained similar across contexts, participants in VR demonstrated a tendency for more exaggerated or exploratory head rotations, possibly reflecting differences in environmental cues or visual search strategies induced by the virtual simulation. Contrary to assumptions about reduced sensory complexity limiting visual exploration in VR, participants actively engaged in exploratory gaze behaviors within the simulated environment.

Interestingly, participants who reported higher immersion scores did not show significantly different head movement patterns compared to those reporting lower immersion, suggesting that subjective feelings of presence or immersion in VR may not be directly linked to observable head movement variability alone.

Head motion smoothness, defined as the inverse of high-frequency angular velocity changes, was also examined for its relationship to simulator sickness symptoms. No significant difference ($p = 0.53$) in head motion smoothness was found between VR and real-world conditions. Additionally, no significant correlation was found between head motion smoothness in the VR condition and participants' simulator sickness scores ($r = -0.245$, $p = 0.526$).

TABLE 1 Summary of IMU kinematic comparisons between VR and real-world cycling.

Metric	n	VR mean (SD)	Real mean (SD)	Test	Stat	p	Mean diff [95% CI]	d
Pedaling frequency (Hz)	9	0.59 (0.28)	0.82 (0.22)	paired t	−1.606	0.147	−0.28 [−0.68, 0.12]	−0.535
Inter-peak interval variance	9	52.07 (142.97)	0.44 (0.68)	Wilcoxon	$W = 11$	0.203	–	–
Bilateral coordination (r)	4	0.332 (0.205)	0.403 (0.275)	Wilcoxon	$W = 3$	0.625	–	–
Upper-body accel. variance	10	0.069 (0.123)	0.072 (0.062)	Wilcoxon	$W = 19$	0.432	−0.003 [−0.11, 0.10]	−0.02
Upper-body gyro variance	10	1625 (1533)	725 (681)	paired t	1.664	0.130	900 [−323, 2124]	0.526
Torso gyro SD (deg/s)	**9**	**26.76 (15.25)**	**12.36 (3.77)**	**paired *t***	**2.889**	**0.020***	**14.4 [2.9, 25.9]**	**0.963**
Body median gyro rate	10	−0.308	−0.088	Wilcoxon	$W = 2.0$	0.012*	–	−1.501
Left-leg median gyro rate	9	−0.327	−0.130	Wilcoxon	$W = 6.0$	0.027*	–	−0.949
Head yaw range (deg)	**8**	**338.9 (62.6)**	**220.8 (130.9)**	**paired *t***	**2.726**	**0.026***	**118.1 [18.2, 218.0]**	**0.909**

Significant results ($p < 0.05$) are marked with an asterisk. n is reported per row and varies due to sensor coverage. Effect sizes: d = Cohen's d. Mean difference = VR − Real; 95% CIs reported for paired comparisons only. Bold values indicate statistically significant differences ($p < 0.05$, Wilcoxon signed-rank test).

TABLE 2 Pearson correlations between head motion metrics and SSQ scores ($n = 8$ participants with complete head sensor data).

Head motion metric	r with SSQ total	p
Gyro standard deviation	0.028	0.942
Gyro range	0.380	0.312
High-frequency energy fraction (>3 Hz)	0.145	0.707

All $p > 0.3$.

4.7 Distributional analysis of IMU signals

Best-fit probability distributions differed between VR and real-world conditions for more than 80% of participant-sensor combinations. In the VR condition, heavy-tailed distributions from the Cauchy family (including the skew-Cauchy distribution) were the most frequently identified best-fit distributions. In the real-world condition, distributions with lighter tails (logistic, Laplace, and normal) dominated. This pattern held across all six sensor positions and was consistent for both accelerometer and gyroscope channels.

The heavy-tailed character of VR motion signals indicates that occasional large-amplitude movements, likely associated with balance corrections on the tilt platform or HMD-induced postural adjustments, occur more frequently than would be expected under a normal or logistic model. In real-world cycling, the logistic and normal fits suggest a more regular distribution of movement amplitudes, which is consistent with the lower torso angular rate variability reported in Section 4.4. The Laplace distribution, which appears at intermediate frequency in real-world data, has heavier tails than the normal but lighter than Cauchy, reflecting occasional larger movements associated with road surface variation or steering maneuvers.

These distributional differences have practical implications for statistical analyses: applying parametric tests that assume normality to VR cycling IMU data may produce inflated Type-I or Type-II errors, justifying the mixed parametric and non-parametric approach used in this study. Vitali and Perkins (2020) highlight the complexity of IMU-based motion capture and the importance of careful signal processing choices; the present findings extend the case for robust or non-parametric procedures as defaults in wearable sensor research, specifically in VR cycling contexts.

A further implication concerns the choice of summary statistics for simulator fidelity comparisons. Standard deviations underestimate the spread of Cauchy-distributed data (the Cauchy distribution has undefined variance in the theoretical sense); consequently, the reported SD values for VR signals should be understood as empirical descriptors of the observed sample rather than estimates of a population standard deviation in the classical sense. Median absolute deviation (MAD) is a more robust central tendency descriptor for heavy-tailed data and should be considered in future fidelity assessments (Camomilla et al., 2018).

TABLE 3 Questionnaire results: composite achievement and affective response items for VR and real-world cycling conditions.

Item / Scale	VR Mean (SD)	Real Mean (SD)	t	p	d
SSQ Total	6.18 (4.12)	–	–	–	–
Presence/Immersion	14.64 (5.55)	–	–	–	–
Composite Achievement	5.00 (2.45)	7.27 (2.20)	−4.077	0.002*	−1.229
Overall Impression	2.00 (0.89)	3.36 (0.67)	−4.404	0.001*	−1.328
Enjoyment	2.55	3.09	−2.206	0.052†	−0.665
Frustration	1.36	0.82	2.206	0.052†	0.665

Significant results ($p < 0.05$) are marked with an asterisk; trends ($p < 0.10$) with †.

4.8 Questionnaire results

4.8.1 Simulator sickness

SSQ total scores were low across all participants ($M = 6.18$, SD = 4.12, range 2–17). According to the severity classification of Kennedy et al. (1993), scores below approximately 15 are generally considered slight sickness; the mean value here indicates that the simulator produced minimal adverse symptoms in this sample. Subscale scores were not reported individually due to small cell sizes, but the low overall values indicate that simulator sickness is unlikely to have substantially altered movement behavior.

4.8.2 Presence and immersion

The composite presence/immersion score (cvr1–cvr4, where applicable) yielded $M = 14.64$ (SD = 5.55) on the 7-point scale. The VR comparison items showed a moderate level of felt presence: item cvr1 ($M = 5.45$), cvr2 ($M = 5.73$), cvr3 ($M = 2.64$), and cvr4 ($M = 4.91$). The lower score on cvr3 suggests that some participants did not fully feel transported to the virtual environment, despite the HMD's immersive display.

4.8.3 Achievement and affective response

Table 3 presents the full achievement and affective response results. The composite achievement score was significantly lower in VR ($M = 5.00$, SD = 2.45) than in real cycling [$M = 7.27$, SD = 2.20; paired $t_{(10)} = -4.077$, $p = 0.002$, $d = -1.229$]. Overall impression showed the largest between-condition difference: VR rated $M = 2.00$ (SD = 0.89) vs. Real $M = 3.36$ [SD = 0.67; paired $t_{(10)} = -4.404$, $p = 0.001$, $d = -1.328$].

Enjoyment (item aes1/aer1) was numerically lower in VR ($M = 2.55$) vs. real ($M = 3.09$), approaching but not reaching significance ($t = -2.206$, $p = 0.052$, $d = -0.665$). Frustration (item aes3/aer3) showed the complementary pattern: higher in VR ($M = 1.36$) than real ($M = 0.82$; $t = 2.206$, $p = 0.052$, $d = 0.665$). Both effects were medium-to-large in magnitude despite only reaching trend level, likely reflecting insufficient statistical power at $n = 11$.

The magnitude of these affective differences, with Cohen's d values exceeding 1.2, is comparable to those reported by Poli et al. (2024) for motivation and enjoyment, providing convergent evidence that VR cycling produces systematically less positive affect than real-world cycling, independent of task completion or physiological load.

5 Discussion

This study examined body segment kinematics during VR simulator and real-world cycling using a six-site IMU system. The primary finding is significantly greater torso angular rate variability in VR ($p = 0.020$, $d = 0.963$), indicating altered postural control in the simulator. IMU signal distributions were predominantly heavy-tailed in VR and lighter-tailed in real cycling, pointing to a qualitative shift in motor behavior beyond mean-level differences. Affective responses were substantially more positive in real-world cycling, while simulator sickness was mild and unrelated to head motion.

5.1 Torso instability as the primary biomechanical marker

The most consistent finding is that torso angular rate variability is substantially higher in VR than in real-world cycling ($p = 0.020$, $d = 0.963$). This pattern parallels results from VR walking studies: Horsak et al. (2021) found that kinematic variability increased across body segments in VR walking, and Horsak et al. (2023) documented increased trunk variability and reduced dynamic balance during HMD-assisted overground walking. The mechanisms may overlap: visual-vestibular conflict induced by any mismatch between the HMD display and physical head movement perturbs balance, and the trunk responds with corrective postural adjustments. In walking, the trunk's role in forward momentum and balance is primary. In cycling, where the lower body is mechanically constrained, the trunk is the segment most free to respond to balance perturbations, making it a natural locus for destabilizing effects.

The motion platform's ±5° tilt likely contributes to this finding. Although tilt cues are known to improve perceived realism (Wintersberger et al., 2022), they introduce a physical perturbation that requires active counteraction. When tilt occurs asynchronously with, or independently of, the visual scene, the mismatch may amplify rather than reduce corrective body movement. The real-world condition, where vestibular, visual, and proprioceptive signals are congruent, produces smoother and more consistent postural patterns, reflected in the lower inter-participant SD (3.77 vs. 15.25 deg/s).

The additional Wilcoxon tests at the body and left-leg positions (both $p < 0.05$, $d > 0.9$) reinforce this picture. The left-leg sensor result is notable because the shank motion is largely constrained by the crank cycle; the significant difference in median gyroscope rate may reflect differences in how participants distribute leg force across the pedal cycle, or subtly different crank geometry between the real bicycle and the simulator. Both effects warrant direct investigation in future work.

5.2 Pedaling and coordination differences

The non-significant reduction in pedaling frequency in VR ($M = 0.59$ vs. 0.82 Hz, $d = -0.535$) reflects a moderate effect that the current sample is underpowered to detect reliably. A sample size of approximately 30 would be required to achieve 80% power at this effect size and $\alpha = 0.05$. The practical significance of the difference is real: a 0.23 Hz gap in pedaling frequency corresponds to approximately 14 RPM, which is a functionally meaningful difference in workout intensity and motor pattern. Swart and Holliday (2019) note that sub-optimal cadence is associated with altered muscle activation patterns and joint loading, suggesting that cadence differences between conditions could have downstream effects on both training efficiency and injury risk.

The higher inter-peak interval variance in VR, despite not reaching significance, is consistent with more irregular pedaling in the simulator. This may reflect attentional demands of the HMD, reduced proprioceptive feedback from the simulator's fixed crank, or lower intrinsic motivation to maintain a steady cadence (consistent with the lower enjoyment scores reported in Section 4.8).

The bilateral coordination result ($W = 3$, $p = 0.625$) should be treated with caution given that only $n = 4$ participants had complete bilateral shank data. This is a limitation of the sensor deployment protocol and not a substantive null finding; future studies should ensure bilateral coverage.

5.3 Head motion and simulator sickness

The absence of significant head motion differences and the lack of correlation between head motion and SSQ scores is an important negative result. Palmisano et al. (2020) proposed that cybersickness arises from discrepancies between virtual and physical head pose; under this model, participants with higher simulator sickness scores should show greater head motion in VR as they attempt to resolve the discrepancy. The present data do not support this prediction within this sample. However, the low SSQ scores (mean total 6.18) constrain this test: with minimal sickness, there is limited range of individual differences to correlate against.

The mild sickness scores themselves merit brief commentary. Saredakis et al. (2020) found in their meta-analysis that locomotion-based VR produced the highest SSQ scores among content types (pooled mean SSQ $\approx$ 18–25 in locomotion studies); the present mean of 6.18 sits well below this range. Several factors may explain the low scores: the use of a physical motion platform providing congruent tilt cues, the participants' youth and familiarity with gaming technology, the short session duration, and the low forward speed of the cycling route. Bimberg et al. (2020) noted that SSQ norms derived from large-screen simulators may not map directly to HMD experiences, and suggested that modern VR systems tend to produce fewer sickness symptoms due to reduced latency and higher frame rates. These factors collectively suggest that the present simulator configuration represents a relatively mild VR environment, which limits the power of sickness-kinematics correlation analyses but also confirms the system's safety for repeated-measures designs.

The high-frequency spectral analysis (energy fraction >3 Hz: $p = 0.950$) suggests that VR does not introduce additional micro-oscillatory head movement. Park and Koo (2025) showed that VR content characteristics modulate cybersickness and head movement, but the content used here, a slow and familiar cycling route, may represent a low-risk scenario for cybersickness induction. Axis-level decomposition of head motion, and analysis of gaze direction via eye-tracking, are directions for follow-up investigation.

5.4 Visual attention and gaze behavior

The broader yaw range observed in VR suggests that participants engaged in more exploratory lateral head movements in the simulated environment compared to the real world. This is counterintuitive, as one might expect the reduced sensory complexity of VR to limit visual exploration. Instead, participants appeared to compensate for the absence of rich peripheral and auditory cues by actively scanning the virtual environment more broadly. The finding that head-turn frequency did not differ between conditions reinforces this interpretation: it is the amplitude, not the frequency, of exploratory behavior that changes.

The absence of a relationship between immersion scores and head movement patterns indicates that felt presence is not simply a function of motor engagement with the environment. Participants who reported feeling highly immersed did not move their heads differently from those who reported low immersion, suggesting that cognitive and perceptual factors, such as the quality of visual rendering and the coherence of the virtual narrative, may be more important determinants of subjective immersion than observable gaze behavior. This has implications for VR simulator design: improving visual fidelity may enhance presence more effectively than optimizing for natural head movement.

Eye-tracking data, collected via the Pupil Labs Core (Pupil Labs GmbH, Berlin, Germany) system during the study, were not analyzed in the present work but represent a valuable resource for future investigation. Gaze fixation patterns, saccade dynamics, and the relationship between eye and head movements during VR cycling could provide finer-grained insight into visual attention strategies and their role in simulator sickness and immersion.

5.5 Distributional structure of VR motion signals

The systematic shift from Cauchy-family distributions in VR to logistic/normal distributions in real-world cycling is a novel finding that extends beyond mean-and-variance comparisons. Heavy-tailed distributions, particularly the Cauchy, assign substantially higher probability to extreme values relative to Gaussian. In the context of cycling kinematics, a Cauchy best-fit implies that occasional large-amplitude movements contribute disproportionately to the signal's statistical structure. These events may correspond to abrupt balance corrections, HMD-induced postural reactions, or the tilt platform's lateral perturbations.

The theoretical implications for simulator validation are significant. Simulator validation studies typically compare mean values or correlation coefficients between conditions (O'Hern et al., 2017; Hammami et al., 2025). If the underlying distributions differ fundamentally, as the present data suggest, then equal means and variances can coexist with qualitatively different movement dynamics. A simulator that matches average cadence and average trunk displacement may still produce a qualitatively different motor pattern characterized by infrequent but large corrective movements. Future validity studies should report distributional characteristics, not only summary statistics.

This finding also aligns with Nürnberger et al. (2021), who proposed that VIMS arises from Bayesian prediction error accumulation: occasional prediction errors of large magnitude, encoded in the heavy tail of the sensory experience distribution, drive the sickness response even when the average experience is benign. The Cauchy-dominant signal structure in VR may reflect the same underlying mechanism at the motor level.

5.6 Affective response and its relationship to movement

The large achievement and overall impression differences between conditions ($d \approx -1.3$) indicate that participants found VR cycling substantially less satisfying. These magnitudes are larger than typically reported for short laboratory tasks, and converge with Poli et al. (2024) despite methodological differences. The frustration asymmetry (higher in VR) and enjoyment asymmetry (lower in VR) suggest that the VR experience is perceived as less rewarding and more effortful, even when sickness is low.

These affective differences are relevant to movement because negative affect and low motivation are associated with reduced motor engagement. If participants are less engaged in the VR condition, they may exert less effort to maintain postural stability or consistent pedaling, potentially contributing to the higher torso variability and lower cadence observed. The relationship between affect and kinematics in this dataset could not be statistically confirmed due to sample size, but it represents a theoretically grounded hypothesis for future study.

Levac et al. (2019) identified that motor skill transfer from VR depends on sensorimotor contingency fidelity and task complexity. The present results suggest that even simple cyclic tasks such as pedaling are not immune to VR-induced alterations, and that the affective component of the experience may mediate part of the kinematic divergence.

5.7 Implications for simulator design and research use

The findings carry several practical implications. First, the torso is a reliable, accessible measurement site for characterizing VR-induced postural effects in cycling. A single IMU is sufficient to detect significant between-condition differences and could be incorporated into routine simulator validation protocols. Second, the Cauchy distribution pattern in VR signals warrants caution when applying standard statistical models to simulator data. Non-parametric or robust statistical methods are preferable unless distributional assumptions are empirically verified. Third, the affective gap between VR and real cycling should be considered when using VR platforms as training environments: participants may not replicate real-world movement intensity if engagement is lower, limiting the ecological validity of training transfer.

A fourth implication concerns the generalizability of the present biomechanical findings to other simulator architectures. The system used here is a fixed bicycle on an actuated $\pm 5^{\circ}$ tilt platform combined with an HMD; the design intentionally introduces lateral motion cues but, in doing so, also introduces a physical perturbation that the rider must counteract. Roller-based and smart-trainer simulators, in which the participant rides their own outdoor bicycle and views the virtual environment either on a flat display or through an HMD, differ along two axes that we expect to be biomechanically relevant. First, bicycle fit (seat height, handlebar position, crank length, frame geometry) is identical to outdoor riding by construction, eliminating any contribution of fit mismatch to upper-body kinematics. Second, the absence of a motorized tilt platform removes the externally imposed roll perturbation that, in our system, plausibly contributes to the elevated torso angular-rate variability observed in VR. We therefore expect that roller-based systems with a flat display would produce torso kinematics closer to the real-world baseline than the system studied here, that roller-based systems with an HMD would isolate the contribution of the visual-vestibular component (and of the headset mass) without the confound of platform-induced roll, and that the heavy-tailed (Cauchy-family) distributional signature observed here may be specific to platforms with active mechanical perturbation rather than a general property of VR cycling. A controlled comparison between fixed-platform-with-HMD, roller-with-HMD, and roller-with-flat-display configurations using the same instrumentation would directly test this expectation and is a natural next step for the field.

5.8 Limitations

Several limitations must be acknowledged. The sample size is small, and most non-significant results should be interpreted as inconclusive rather than as evidence for absence of effects. The absence of counterbalancing (all participants completed VR first) means that potential order effects, such as fatigue, familiarity with the route, or novelty, cannot be excluded as confounds influencing subjective responses and movement patterns. The unequal session durations (VR shorter by approximately 100 s on average) introduce a confound: if fatigue or warm-up effects influence movement patterns, duration differences could account for some of the observed kinematic differences independent of the VR manipulation. Normalizing metrics to a fixed analysis window (e.g., first 300 s of each session) would address this but was not implemented here. Sensor placement relied on elastic straps, which may introduce soft-tissue artefact. The HMD itself has a non-negligible mass (approximately 503 g) that is borne by the rider's head and upper neck only in the VR condition; we cannot fully decompose the observed differences in head and torso kinematics into a component attributable to the visual-vestibular characteristics of VR and a component attributable to the added inertial load of the headset. Likewise, although both bicycles were of the same model and frame size and the seat and handlebars were adjusted to each participant's preference at the start of each session, these fit parameters (in particular seat height) were not quantitatively recorded in each condition, so small between-condition mismatches in seat or handlebar position cannot be excluded. Crank length and handlebar geometry, being fixed properties of the matched bicycle model, can be considered matched across conditions. The participant group was predominantly young male university students with moderate cycling experience, which limits generalization to other populations. Anthropometric measurements (participant height and body mass) were not recorded, which limits our ability to normalize kinematic signals by body geometry and prevents body-segment-parameter-based comparisons across the sample; future studies should collect these measures routinely. Likewise, the number of pedal cycles per participant is modest, particularly in the VR condition where session durations were shorter; although our variance estimator does not depend systematically on cycle count, larger numbers of cycles per participant would improve the stability of the cadence-regularity estimate. Finally, the outdoor route was not controlled for traffic or other cyclists, introducing environmental variability absent in the simulator.

6 Conclusion

This study provides a comprehensive characterization of body segment kinematics during VR simulator and real-world cycling using a six-site wearable IMU system. The primary finding is that torso angular rate variability is significantly higher in VR (mean 26.76 vs. 12.36 deg/s; $p = 0.020$, $d = 0.963$), indicating greater postural instability in the simulator. Segment-level Wilcoxon tests confirm that this pattern extends to median gyroscope rates at both the torso and left-leg positions. IMU signal distributions are predominantly heavy-tailed (Cauchy family) in VR and lighter-tailed (logistic, normal) in real-world cycling, suggesting a qualitative change in the motor control regime that is not captured by mean-level comparisons alone. Pedaling frequency is lower in VR ($d = -0.535$) but does not reach significance with the current sample. Affective response and perceived achievement are substantially more positive in real-world cycling, with large effect sizes ($d \approx -1.3$) that converge with prior literature. Simulator sickness is mild and head motion does not predict SSQ score within this sample.

These results reinforce that simulator fidelity assessments should go beyond mean-level behavioral comparisons to include postural dynamics, distributional structure, and subjective experience. The torso is the most sensitive IMU-accessible site for detecting VR–real differences in cycling, and single-sensor protocols may suffice for routine fidelity monitoring.

Future work should use larger samples to confirm trends identified here, implement matched session durations, and examine the within-session time course of adaptation to the VR environment.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://street2simulator.de/.

Ethics statement

The studies involving humans were approved by Ethics Committee of the University of Siegen, reference LS_ER_03_2023. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

JP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AV: Data curation, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. FW: Conceptualization, Project administration, Resources, Validation, Writing – review & editing. FM: Conceptualization, Methodology, Resources, Supervision, Writing – review & editing. KV: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), SPP 2199 – Project number 521584557 (PervaSafe Computing).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor FD declared a past co-authorship with the author KV.

The author KV declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. Generative AI was used to enhance the language of the paper, to check for grammatical errors and to adhere to a standard English. Generative AI was also used to produce some of the software used to analyze the data.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bimberg, P., Weissker, T., and Kulik, A. (2020). "On the usage of the simulator sickness questionnaire for virtual reality research," in *2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)* (Atlanta, GA), 464–467. doi: 10.1109/VRW50115.2020.00098

Bruce, R. M., Rafferty, G. F., Finnegan, S. L., Sergeant, M., Pattinson, K. T. S., and Runswick, O. R. (2025). Incongruent virtual reality cycling exercise demonstrates a role of perceived effort in cardiovascular control. *J. Physiol.* 603, 5149–5161. doi: 10.1113/JP287421

Camomilla, V., Bergamini, E., Fantozzi, S., and Vannozzi, G. (2018). Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: a systematic review. *Sensors* 18:873. doi: 10.3390/s18030873

Cerfoglio, S., Capodaglio, P., Rossi, P., Conforti, I., D'Angeli, V., Milani, E., et al. (2023). Evaluation of upper body and lower limbs kinematics through an IMU-based medical system: a comparative study with the optoelectronic system. *Sensors* 23:6156. doi: 10.3390/s23136156

Colombo, V., Mondellini, M., Aliverti, A., and Sacco, M. (2025). Immersion and interaction during cycling in virtual reality: the influence on perceived effort and subjective experience. *Front. Virtual Real.* 6:1490588. doi: 10.3389/frvir.2025.1490588

Cordillet, S., Bideau, N., Bideau, B., and Nicolas, G. (2019). Estimation of 3D knee joint angles during cycling using inertial sensors: accuracy of a novel sensor-to-segment calibration procedure based on pedaling motion. *Sensors* 19:2474. doi: 10.3390/s19112474

Haasnoot, J., Happee, R., van der Wijk, V., and Schwab, A. L. (2024). Validation of a novel bicycle simulator with realistic lateral and roll motion. *Veh. Syst. Dyn.* 62, 1802–1826. doi: 10.1080/00423114.2023.2264418

Hammami, A., Borsos, A., and Sándor, Á. P. (2025). How realistic a bicycle simulator can be?-a validation study. *Multimodal Transp.* 4:100193. doi: 10.1016/j.multra.2025.100193

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Horsak, B., Simonlehner, M., Dumphart, B., and Siragy, T. (2023). Overground walking while using a virtual reality head mounted display increases variability in trunk kinematics and reduces dynamic balance in young adults. *Virtual Real.* 27, 3021–3032. doi: 10.1007/s10055-023-00851-7

Horsak, B., Simonlehner, M., Schoffer, L., Dumphart, B., Jalaeefar, A., and Husinsky, M. (2021). Overground walking in a fully immersive virtual reality: a comprehensive study on the effects on full-body walking biomechanics. *Front. Bioeng. Biotechnol.* 9:780314. doi: 10.3389/fbioe.2021.780314

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303_3

Levac, D. E., Huber, M. E., and Sternad, D. (2019). Learning and transfer of complex motor skills in virtual reality: a perspective review. *J. Neuroeng. Rehabil.* 16:121. doi: 10.1186/s12984-019-0587-8

Matviienko, A., Hoxha, H., and Mühlhäuser, M. (2023). "What does it mean to cycle in virtual reality? Exploring cycling fidelity and control of VR bicycle simulators," in *Proceedings of the 2023 CHI conference on human factors in computing systems* (Hamburg), 1–15. doi: 10.1145/3544548.3581050

McIlroy, B., Passfield, L., Holmberg, H.-C., and Sperlich, B. (2021). Virtual training of endurance cycling-a summary of strengths, weaknesses, opportunities and threats. *Front. Sports Act. Living* 3:631101. doi: 10.3389/fspor.2021.631101

Nürnberger, M., Klingner, C., Witte, O. W., and Brodoehl, S. (2021). Mismatch of visual-vestibular information in virtual reality: is motion sickness part of the brain's attempt to reduce the prediction error? *Front. Hum. Neurosci.* 15:757735. doi: 10.3389/fnhum.2021.757735

O'Hern, S., Oxley, J., and Stevenson, M. (2017). Validation of a bicycle simulator for road safety research. *Accid. Anal. Prevent*, 100, 53–58. doi: 10.1016/j.aap.2017.01.002

Palmisano, S., Allison, R. S., and Kim, J. (2020). Cybersickness in head-mounted displays is caused by differences in the user's virtual and physical head pose. *Front. Virtual Real.* 1:587698. doi: 10.3389/frvir.2020.587698

Park, S.-Y., and Koo, D.-K. (2025). The impact of virtual reality content characteristics on cybersickness and head movement patterns. *Sensors* 25:215. doi: 10.3390/s25010215

Pöhler, J., and Van Laerhoven, K. (2025). "Equimetrics: an open-source wearable system for real-time feedback in equestrian sports," in *Companion of the 2025 ACM international joint conference on pervasive and ubiquitous computing* (Espoo: ACM), 1278–1282. doi: 10.1145/3714394.3756251

Pöhler, J., Vitt, A., Wolling, F., Michahelles, F., and Van Laerhoven, K. (2025). "From street to simulator: assessing cyclist movement consistency between virtual and real-world environments," in *Proceedings of the 2025 ACM international symposium on wearable computers (ISWC)* (New York, NY). doi: 10.1145/3715071.3750432

Poli, L., Greco, G., Gabriele, M., Pepe, I., Centrone, C., Cataldi, S., et al. (2024). Effect of outdoor cycling, virtual and enhanced reality indoor cycling on heart rate, motivation, enjoyment and intention to perform green exercise in healthy adults. *J. Funct. Morphol. Kinesiol.* 9:183. doi: 10.3390/jfmk9040183

Prochazka, A., Charvatova, H., Honzirkova, M., and Schatz, M. (2025). Integrated biomechanical motion analysis in a virtual cycling environment using wearable sensors. *IEEE Access* 13, 175069–175077. doi: 10.1109/ACCESS.2025.3619396

Rojo, A., Raya, R., and Moreno, J. C. (2023). Virtual reality application for real-time pedalling cadence estimation based on hip ROM tracking with inertial sensors: a pilot study. *Virtual Real.* 27, 3–17. doi: 10.1007/s10055-022-00668-w

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. *J. Pers. Soc. Psychol.* 43, 450–461. doi: 10.1037/0022-3514.43.3.450

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., and Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. *Front. Hum. Neurosci.* 14:96. doi: 10.3389/fnhum.2020.00096

Sato, K., Suda, Y., and Higuchi, T. (2025). Assessing and enhancing the ecological validity of human locomotion in virtual reality. *Front. Virtual Real.* 6:1699143. doi: 10.3389/frvir.2025.1699143

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: factor analytic insights. *Presence Teleoperators Virtual Environ*. 10, 266–281. doi: 10.1162/105474601300343603

Simonlehner, M., Dumphart, B., and Horsak, B. (2024). GaitRec-VR: 3D gait analysis for walking overground with and without a head-mounted-display in virtual reality. *Sci. Data* 11:1099. doi: 10.1038/s41597-024-03939-0

Swart, J., and Holliday, W. (2019). Cycling biomechanics optimization-the (r)evolution of bicycle fitting. *Curr. Sports Med. Rep*. 18, 490–496. doi: 10.1249/JSR.0000000000000665

Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using presence questionnaires in reality. *Presence Teleoperators Virtual Environ*. 9, 497–503. doi: 10.1162/105474600566989

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2

Vallat, R. (2018). Pingouin: statistics in Python. *J. Open Source Softw*. 3:1026. doi: 10.21105/joss.01026

Vitali, R. V., and Perkins, N. C. (2020). Determining anatomical frames via inertial motion capture: a survey of methods. *J. Biomech*. 106:109832. doi: 10.1016/j.jbiomech.2020.109832

Wintersberger, P., Matviienko, A., Schweidler, A., and Michahelles, F. (2022). "Development and evaluation of a motion-based VR bicycle simulator," in *Proceedings of the ACM on human-computer interaction, Vol. 6* (New York, NY), 1–19. doi: 10.1145/3546745