Predicting Motion Incongruence Ratings in Closed- and Open-Loop Urban Driving Simulation

Kolff, Maurice; Venrooij, Joost; Arcidiacono, Elena; Pool, Daan M.; Mulder, Max

DOI
10.1109/TITS.2024.3503496

Publication date
2025

Document Version
Final published version

Published in
IEEE Transactions on Intelligent Transportation Systems

Citation (APA)
Kolff, M., Venrooij, J., Arcidiacono, E., Pool, D. M., & Mulder, M. (2025). Predicting Motion Incongruence Ratings in Closed- and Open-Loop Urban Driving Simulation. *IEEE Transactions on Intelligent Transportation Systems*, *26*(1), 517-528. https://doi.org/10.1109/TITS.2024.3503496

Predicting Motion Incongruence Ratings in Closed- and Open-Loop Urban Driving Simulation

Maurice Kolff, Joost Venrooij, Elena Arcidiacono, Daan M. Pool, *Member, IEEE*, and Max Mulder, *Senior Member, IEEE*

***Abstract*— This paper presents a three-step validation approach for subjective rating predictions of driving simulator motion incongruences based on objective mismatches between reference vehicle and simulator motion. This approach relies on using high-resolution rating predictions of open-loop driving (participants being driven) for ratings of motion in closed-loop driving (participants driving themselves). A driving simulator experiment in an urban scenario is described, of which the rating data of 36 participants was recorded and analyzed. In the experiment's first phase, participants actively drove themselves (i.e., closed-loop). By recording the drives of the participants and playing these back to themselves (open-loop) in the second phase, participants experienced the same motion in both phases. Participants rated the motion after each maneuver and at the end of each drive. In the third phase they again drove open-loop, but rated the motion continuously, only possible in open-loop driving. Results show that a rating model, acquired through a different experiment, can well predict the measured continuous ratings. Second, the maximum of the measured continuous ratings correlates to both the maneuver-based ($\rho = 0.94$) and overall ($\rho = 0.69$) ratings, allowing for predictions of both rating types based on the continuous rating model. Third, using Bayesian statistics it is then shown that both the maneuver-based and overall ratings between the closed-loop and open-loop drives are equivalent. This allows for predictions of maneuver-based and overall ratings using the high-resolution continuous rating models. These predictions can be used as an accurate trade-off method of motion cueing settings of future closed-loop driving simulator experiments.**

***Index Terms*— Motion cueing, driving simulators, urban driving, subjective ratings, rating predictions.**

Received 21 June 2023; revised 28 April 2024 and 16 September 2024; accepted 17 November 2024. Date of current version 9 January 2025. This work was supported by BMW Group. The Associate Editor for this article was D.-H. Lee. *(Corresponding author: Maurice Kolff.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by BMW Group.

Maurice Kolff is with the Department of Research and Development, BMW Group, 80807 Munich, Germany, and also with the Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands (e-mail: m.j.c.kolff@tudelft.nl).

Joost Venrooij is with the Department of Research and Development, BMW Group, 80807 Munich, Germany (e-mail: Joost.Venrooij@bmw.de).

Elena Arcidiacono is with the Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zürich, 8092 Zürich, Switzerland (e-mail: earcidiacono@ethz.ch).

Daan M. Pool and Max Mulder are with the Section of Control and Simulation, Delft University of Technology, 2628 HS Delft, The Netherlands (e-mail: d.m.pool@tudelft.nl; m.mulder@tudelft.nl).

Digital Object Identifier 10.1109/TITS.2024.3503496

I. Introduction

DRIVING simulators are essential tools in the development of future driving technologies due to their ability to create safe and repeatable test conditions. When equipped with a motion system, their limited workspace often induces mismatches between vehicle and simulator inertial motion [1]. While some mismatches are not perceived by the driver, the motion is *incongruent* if the driver *does* notice a deviation between their expectation of the real vehicle motion and the simulator motion they actually perceive [2], [3]. Incongruent motion can lead to an impaired perceptual fidelity of the simulation and induce simulator sickness [4]. Therefore, the development, evaluation, and trade-off of Motion Cueing Algorithms (MCAs) typically aim at selecting the option with potentially the least incongruences. Acquiring and validating this information currently requires performing subjective evaluations in a driving simulator. Being able to *predict* such ratings through objective measures would be a crucial advancement [5]. They would allow for rapid, systematic, and cost-efficient assessment of MCAs and guide developments, e.g., of Model-Predictive Control (MPC) algorithms [6]. However, making such predictions is notoriously difficult. For example, it is known that drivers generally consider scaled motion as more realistic than fully congruent one-to-one simulator motion [7], of which the cause is not yet understood.

In most driving simulations, drivers control the simulated vehicle themselves ("closed-loop"). Due to differences in driver behavior and driving style, each drive is different, resulting in different experiences of motion. Existing models to objectively predict subjective ratings [2], [3], [8] are based on ratings of "open-loop" driving. Here, human drivers are driven around as passengers. The fact that they do not need to provide any steering control inputs has two crucial advantages. First, open-loop driving allows for performing multiple identical repetitions of exactly the same drive, e.g., to obtain more reliable subjective rating data [3]. Second, the absence of a driving task allows for a more invasive rating task, such as letting drivers *continuously* rate the motion cueing through a rating knob [2], [9], providing unmatched insights into *when* and *where* in the simulation (in)congruent motion occurs. Due to the high temporal resolution of continuous ratings, their

relation to objectively calculated mismatches between vehicle and simulated motion can be captured in mathematical models, which in turn allow for *predicting* continuous ratings [2], [3], [8]. However, as drivers are expected to continuously assess their perceived motion and operate a rating knob with one hand, the continuous rating method cannot be used in closed-loop scenarios, i.e., when drivers need to operate the steering wheel with both hands. Rating methods that *are* suitable for closed-loop driving, such as providing a single rating after each drive or maneuver, are of such lower resolution that they are much less suitable to be used in a modelling approach.

Thus, it would be extremely useful if the high-resolution open-loop prediction models of continuous ratings can be used in the design, evaluation, and testing of motion cueing for closed-loop driving simulation. However, the central assumption of the continuous rating method, i.e., that it is representative of closed-loop simulations, has never been tested. Differences between the two driving methods might occur due to perceptual differences [10], [11] or due to changes to the internal representation of motion [3]. With both the strengths and limitations of the continuous rating method in mind, three gaps are identified that would need to be answered to investigate whether continuous ratings of open-loop driving, and their predictions models, can be used to predict ratings of closed-loop driving. First, a rating model must be used to predict measured continuous ratings. This is challenging because existing rating models [2], [3], [8] have not yet been confirmed to hold predictive power *between* experiments. Second, explicit rating relationships must be developed, that can link the continuous rating method to rating methods that are possible in closed-loop driving, such as after each maneuver or after the whole drive. Finally, no work so far has investigated the equivalence of open-loop and closed-loop driving. The equivalence of these simulation methods would be a requirement to be able to make predictions of closed-loop drives based on the open-loop rating models.

This paper presents a comprehensive driving simulator experiment consisting of three phases, all performed in the Sapphire Space simulator at BMW Group. Subjective ratings were obtained from 42 drivers in both closed-loop and open-loop driving simulations. By recording the closed-loop drives of the individual drivers (first phase) and playing these back to themselves in the open-loop phase (second phase) of the experiment, it is ensured that exactly the same motion is presented. In both driving methods, the motion is evaluated through overall and maneuver-based ratings. In the third phase, drivers again perform the open-loop rating task for the same recorded drives, but rate using the *continuous* rating method.

The paper's main contribution is a complete, three-step approach that allows for predicting overall and maneuver-based subjective ratings of closed-loop driving as a function of objective motion cueing mismatch signals. First, a model for predicting continuous motion incongruence ratings from previous work [3] is employed to test whether the recorded continuous ratings (third phase) can be predicted. Although the same urban scenario of [3] is simulated, a different simulator, MCA parameters, and participant group were used. Second, it is investigated whether predictive relations exist from the continuous rating (third phase) to the overall and maneuver-based ratings which can be obtained in closed-loop driving (second phase). Reference [3] showed that the maximum of the continuous rating highly correlates to the overall rating. These methods are extended by also considering the mean and median, as well as providing a similar analysis for the maneuver-based ratings. The rating model is then used to make predictions of both rating methods. Third, Bayes' theorem [12] is used to verify whether maneuver-based and overall motion incongruence ratings provided in closed-loop and open-loop driving (first and second phases) are equivalent.

The paper is structured as follows. The driving and rating tasks are discussed in Section II. The experiment set-up is explained in Section III. Results are presented in Section IV, and discussed in Section V. Conclusions are stated in Section VI.

II. Methods

A. Driving Task

When driving closed-loop, illustrated in Figure 1 including the red elements, the driver controls the steering wheel $\delta_s(t)$, the accelerator $\delta_a(t)$ and brake $\delta_b(t)$ pedals. In a simulation, the vehicle simulation then calculates the corresponding vehicle motion states $\tilde{S}_{veh}(t)$, i.e., the specific forces $f(t)$ and rotational rates $\omega(t)$. As $\tilde{S}_{veh}(t)$ comes from a vehicle model, it is an approximation of the real vehicle motion $S_{veh}(t)$, hence the notation $\tilde{(\cdot)}$. The motion states are sent to the *Motion Control System*, consisting of the MCA and the Motion System (MS). The MCA converts the vehicle motion states to commanded platform motion. These are sent to the MS, i.e., the physical simulator, which determines the actual platform motion $\tilde{S}_{sim}(t)$ [13]. These can differ from the commanded platform motion due to a variety of factors, such as the motion system latency. Differences between the vehicle reference and simulator motion are then the objective mismatches, i.e., $\Delta\tilde{S}(t) = \tilde{S}_{veh}(t) - \tilde{S}_{sim}(t)$.

The platform motion is sensed by the driver through their sensory system. Based on the perceived inertial motion and all other non-inertial motion cues in the simulation, such as the visuals [14], the driver chooses their intended control actions based on a desired state. The motor system of the body produces the actual control actions $[\delta_s(t), \delta_a(t)$ and $\delta_b(t)]$, which are sent to the vehicle simulation, closing the driving control loop. In an open-loop driving task (Figure 1, excluding the red elements), the driver does not actively control the vehicle and the vehicle simulation is represented by a playback.

B. Rating Task

Next to the driving task, the participants also performed a rating task. They were tasked with evaluating how well the inertial motion they perceive in the simulator matched to what they would expect to feel from the simulated vehicle. This difference is defined as their Perceived Motion Incongruence (PMI) [2], see Figure 1. As the driver does not exactly know what the vehicle motion would feel like in a particular situation, they must use an *internal representation* [15] of the

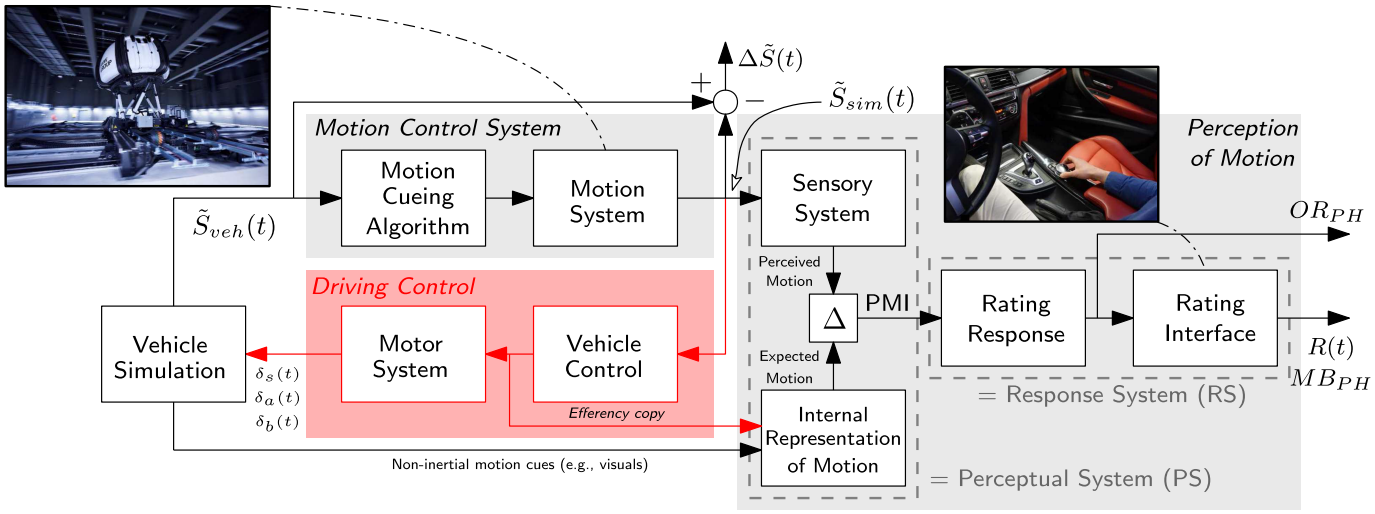


Fig. 1. Block diagram of the driving and rating tasks. The part "Driving Control" (red) is only present in closed-loop driving. The top left image shows the Sapphire Space simulator (Image: BMW Group), the top right image shows the continuous rating interface, adapted from [8].

vehicle motion based on non-motion cues (e.g., visuals) of the simulation. Individual differences can therefore arise due to familiarity or experience with the simulated vehicle [3]. Note that thus both the expected and the vehicle model motion $\tilde{S}_{veh}(t)$ can differ from the real vehicle motion $S_{veh}(t)$.

Additionally, in closed-loop control, an *efference copy* [16] of the intended control actions is used to form the expected motion. The expected motion is then not only based on what drivers expect to feel from the non-inertial motion cues, but also what they expect to feel as a result of their own intended control actions. In open-loop driving, the lack of driving control implies that no efference copy can be present. This means that the internal representation can only be generated from the non-inertial motion cues. It is therefore possible that the PMI of open-loop driving is different than in closed-loop driving, for example, because drivers might know less well what motion to expect.

As the PMI is not measurable, a Motion Incongruence Rating (MIR) was asked from the drivers. A typical choice is a numeric rating scale [2], e.g., ranging between values of 0 (fully congruent) and 10 (highly incongruent), with steps of 1 [3]. Another choice relates to the rating method, i.e., when and how these ratings are acquired. In the experiment, three types of rating methods were used:

1) Overall Ratings (OR_{PH}): After each simulation drive, a single rating value between 0 and 10, representing the overall impression of the drive, was asked from the drivers. As they do not interfere with the driving task(s), overall ratings can be used in closed-loop and open-loop driving. They can be extracted through various methods, such as verbally or through a rating interface. While beneficial due to their non-intrusive nature, the single rating values provide no direct information on which parts of the drive the overall rating is mostly based on. It has been shown that overall rating of PMI correlate with the most incongruent moment in the simulated drive [2], [3].

2) Maneuver-based Ratings (MB_{PH}): Here, a scenario is divided into different maneuvers and drivers give a single rating after those maneuvers [17]. Like overall ratings, maneuver-based ratings can be acquired from closed-loop and open-loop driving, and can be given verbally or through a rating interface. A benefit compared to overall ratings is that maneuver-based ratings provided detailed information for each maneuver separately. A downside is that they require that drivers give their rating while driving, which may cause a slight distraction from the driving task.

3) Continuous Ratings ($R(t)$): Here, drivers rate continuously throughout the drive using a rating interface [2], reflecting their current PMI at each point in time. The main benefit of this method is its high temporal resolution, which allows for modelling approaches [2], [3], [8], [18]. As it requires operating the rating interface with one hand (see Figure 1, top right), it is not possible to drive closed-loop at the same time, which requires both hands on the steering wheel. Therefore, this task can only be performed in open-loop driving experiments.

C. Validation of Rating Predictions

As explained in Section I, a main motivation is to predict how a certain motion cueing setting will be rated subjectively in closed-loop driving, based on objective signals of platform motion. For this, a three-step approach is used (Figure 2).

1) Rating Model Validation: First, a rating model is used to predict continuous ratings based on objective mismatch signals [path (1) in Figure 2]. The latter are defined as the differences in inertial motion (specific forces and rotational rates) between the vehicle motion $\widetilde{S}_{veh,m}(t)$ and the simulator motion $\widetilde{S}_{sim,m}(t)$, i. e., $\Delta\widetilde{S}_m(t)$, with $\widetilde{P}_m(t) = |\Delta\widetilde{S}_m(t)|$. Here, m represents the mismatch direction, e.g., $m \in [f_x,\ f_y,\ \ldots,\ \omega_z]$.

In [3] a linear model was proposed that predicts the continuous rating of the *average* participant. Its structure consists of a first-order low-pass filter transfer function $H_m(j\omega)$ between the absolute mismatch signal $\widetilde{P}_m(t)$ and a modeled rating

signal $\tilde{R}(t)$:

$$\widehat{\tilde{R}}(j\omega) = \sum_m K_{\tilde{P}_m} \left(\frac{\omega_c}{j\omega + \omega_c} \right) \widehat{\tilde{P}}_m(j\omega), \tag{1}$$

with the low-pass filter's cut-off frequency ω_c and the gains of the several mismatch channels $K_{\tilde{P}_m}$. The $\widehat{(\cdot)}$-terms indicate the Fourier transforms. The low-pass filter represents the participants' lagged response (Response System (RS) in Figure 1) to the mismatches $P(t)$. In [3] it was shown that the continuous ratings of a Classical Washout Algorithm (CWA) MCA condition as measured in that study could be largely explained when considering the longitudinal specific force mismatches $\tilde{P}_{f_x}$, as well as the yaw rate mismatch $\tilde{P}_{\omega_z}$ (i.e., $m \in [f_x, \omega_z]$), with the parameters: $\omega_c = 0.37$ rad/s, $K_{f_x} = 0.78$ and $K_{\omega_z} = 6.71$.

To express how well the model is able to predict the measured ratings, the Variance-Accounted-For (VAF) is used:

$$\text{VAF} = \left(1 - \frac{\text{var}[R(t) - \tilde{R}(t)]}{\text{var}[R(t)]} \right) \cdot 100\%, \tag{2}$$

with $R(t)$ and $\tilde{R}(t)$ the measured and modeled rating signal, respectively. The VAF is a measure of how much of the measured signal's variance is explained by the modeled signal [3]. A value of 100% indicates a perfect fit, whereas it is unbounded on the lower side, i.e., $[-\infty < \text{VAF} \leq 100\%]$.

2) Rating Relationships: In [3] it was shown that the maximum of the continuous ratings strongly correlate with the overall ratings [path (2a) in Figure 2], such that a linear relationship of the form $OR_{PH} = f_{OR_{PH}}[R(t)] = \alpha_{OR_{PH}} \cdot \max[R(t)] + \beta_{OR_{PH}}$ exists. A similar relationship, between maneuver-based and continuous ratings [path (2b) in Figure 2], does currently not exist. In the present work, the mean and median of the continuous ratings will also be considered as possible predictor for the overall ratings *and* the maneuver-based ratings.

3) Equivalence Testing: Finally, to investigate whether OR_{PH} and MB_{PH} ratings of open-loop driving can be used for closed-loop driving, their equivalence is investigated [paths (3a) and (3b) in Figure 2, respectively]. In frequentist statistics, data are typically tested for significant differences, i.e., tested for a 95% probability that H_0 (null hypothesis; the data are equivalent) can be rejected in favour of H_1 (alternate hypothesis; the data are different). In the present case, the interest lies not in differences, but in *equivalence*, requiring proof of H_0. This cannot be tested through the same frequentist statistics procedure, as the lack of significant differences does not necessarily imply equivalence. Instead, it only shows that an effect cannot be proven [19], which can also occur in the case of a lack of statistical power. Thus, using frequentist statistics, the H_0 cannot be accepted. This implies that the frequentist approach is not a suitable method for investigating the equivalence of the open-loop and closed-loop ratings. Specially developed alternative frequentist methods, such as the Two One-Sided Tests (TOST) [20], require normally distributed data. Furthermore, the TOST method is considered to be less reliable for testing equivalence when the sample size is relatively small [21].

As an alternative, it is possible to use Bayesian statistics [12], which does allow for explicit testing for equivalence of data. In Bayesian statistics, a degree of belief in a hypothesis is expressed as a form of conditional probability. An estimation of the distribution function is made about the data before even analyzing the data, resulting in a *prior belief*, which holds the ratio of the probability estimates of the hypotheses, i.e, $\frac{P(H_1)}{P(H_0)}$. The prior belief can stem from existing knowledge on the process under investigation, e.g., from previous experiments or from knowledge of underlying physical processes. No explicit assumptions on the distributions of the data, such as normality, are necessary [12]. After the data are observed, the degree of belief is updated [22] to a *posterior* belief. This is expressed as $\frac{P(H_1|D)}{P(H_0|D)}$, with D the observed data (in this case, the maneuver-based ratings of open-loop and closed-loop driving). The Bayes Factor can then be expressed through:

$$BF_{10} = \frac{P(D|H_1)}{P(D|H_0)} = \underbrace{\left(\frac{P(H_1)}{P(H_0)} \right)^{-1}}_{\substack{\text{Prior} \\ \text{Belief}}} \times \underbrace{\frac{P(H_1|D)}{P(H_0|D)}}_{\substack{\text{Posterior} \\ \text{Belief}}} \tag{3}$$

The Bayes Factor, denoted BF_{10}, represents the ratio in proof of H_1 over H_0. Therefore, the factor $BF_{01} = BF_{10}^{-1}$ equals the ratio of proof of H_0 over H_1. A value of $BF_{10} > 1$ indicates that H_1 is more probable [12], but only $BF_{10} > 3$ is considered evidence for H_1. In contrast, $BF_{10} < 1$ means that H_0 is more probable, whereas only $BF_{10} < 0.3$ is considered evidence for H_0 (equivalence). Thus, to prove that the open-loop and closed-loop ratings are equivalent, BF_{10} must be calculated and be shown to be below 0.3. For this analysis, the Bayes factors are calculated using the JASP software [23], which calculates values of BF_{incl}. This Bayes factor indicates the change from prior to posterior inclusion odds [24]. The same range of degrees of belief holds as for BF_{10} [12].

III. Experiment Set-up

A. Experimental Conditions

Using the driving- and rating tasks presented in Section II, the experiment was performed with the following three conditions: i) Closed-loop driving, maneuver-based rating ("CLMB"), ii) Open-loop driving, maneuver-based rating ("OLMB"), and iii) Open-loop driving, continuous rating ("OLCT"). To guarantee that drivers experienced exactly the same motion in the open and closed-loop tasks, the CLMB condition was performed first, such that in the open-loop conditions drivers could be presented with played-back recordings of their own drives. The overall rating was the only rating that was recorded in all three conditions. An overview of the conditions with the applied rating methods is shown in Table I.

B. Scenario and Data Acquisition

For increased comparability, the driven route is exactly the same as in [3], see Figure 3. The maneuvers to be rated in the maneuver-based conditions were indicated on the road using green bars and consist of several typical urban driving maneuvers: corners ('CR'), lane changes ('LC'), as well

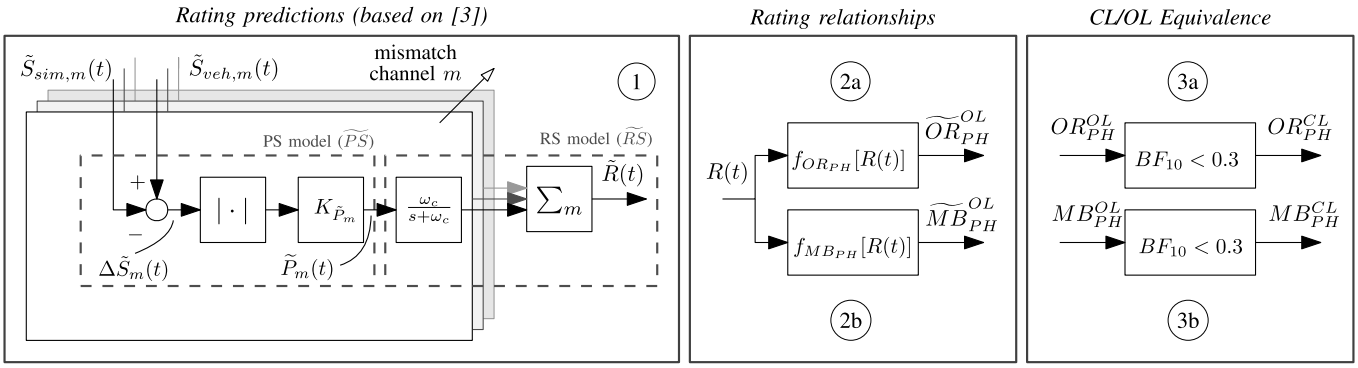


Fig. 2. Contribution steps, representing (from left to right): Open-loop continuous ratings are predicted from objective mismatches using a rating model (1). Second, rating relationships between the continuous ratings to both the open-loop (2a) overall (OR_{PH}) and (2b) maneuver-based (MB_{PH}) ratings are determined. Third, equivalence testing relates (3a) overall and (3b) maneuver-based ratings of open-loop and closed-loop driving.

TABLE I
OVERVIEW OF THE EXPERIMENTAL CONDITIONS

Condition	Driving task	Overall rating [OR_{PH}]	Maneuver-based rating [MB_{PH}]	Continuous rating [$R(t)$]
CLMB	Closed-loop	✓	✓	-
OLMB	Open-loop	✓	✓	-
OLCT	Open-loop	✓	-	✓

as decelerations ('DEC'). A traffic light was present after 'DEC1', before which drivers had to stop, wait, and accelerate again. Compared to [3], there are two changes: First, the roundabout is split-up into the roundabout turn ('RBT') and exit ('RBE') to obtain separate maneuver-based ratings for both, increasing the amount of rating information. Second, three lane change maneuvers in [3], namely after 'CR2', after 'DEC1', and after 'CR3' were not used in the current experiment, as they were not found to result in informative ratings in [3]. Furthermore, this allowed for more time between the various 'CR' maneuvers for drivers to rate. Note that in [3] the division of the maneuvers was not visible to the participants at all, as in that experiment they only rated the motion continuously. In [3], the maneuvers were only introduced and shown for clarity to the reader. Therefore, the changes of the maneuvers compared to [3] is expected to only minimally impact the results.

C. Drive Matching Approach

Due to differences in driving style, all recorded closed-loop drives are inherently unique in terms of velocity and lane position. To visualize the differences in the motion that was presented in each drive, a "drive matching approach" was developed. Here, all recorded time signals are related to a common 'reference drive' (see Figure 4). Here, the data points of each drive of interest (black points) are linearly interpolated (black lines). For each data point i (red points) in the reference run, a line is constructed perpendicular to the closest linear line piece of the drive of interest, representing the shortest distance between point i and the line piece. The point where these lines intersect (red cross) is used to calculate the ratio $r = n_{k+1}/n_k$.

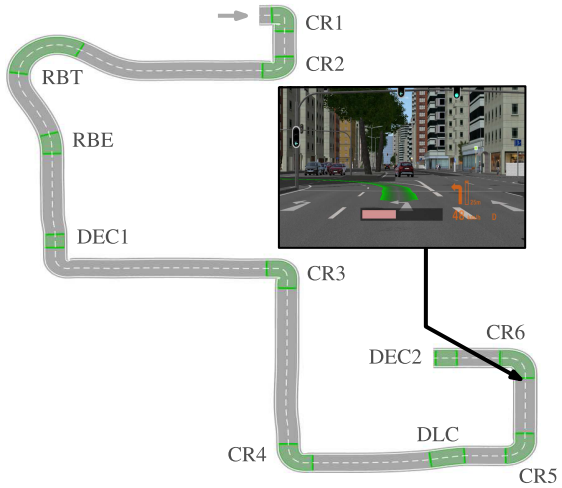


Fig. 3. Top-down view of the driven route, as in [3]. The green areas were visible in the CLMB and OLMB simulations (see screenshot) and represent the maneuvers to be rated: corners (CR), decelerations (DEC), a double lane change (DLC) and a roundabout turn (RBT) and exit (RBE).

The continuous rating signals are evaluated at these two points and the weighted average based on the ratio r is calculated.

This leads to a vector of indices of equal length for all analyzed drives at which the rating signal is evaluated. As it is arbitrary which trajectory is used as the reference drive, as long as the *same* one is used for all drives of all drivers, the trajectory of drive 1 of driver 1 is used. The method allows for relating individual drives with different velocities and lane positions, but inherent differences in driving behaviour can still be present: for example, the point in time at which drivers apply the brake can be different.

Note that the drive matching method is useful for comparing the driving behaviour of various drives. However, as the method is purely based on the position of the vehicle with respect to the reference vehicle, the method will likely not work when a certain point in the scenario is passed more than once *within* a single drive. In that case the method might incorrectly link these instances together. However, this did not occur in the present experiment. Furthermore, note that expressing the drives relative to a reference drive also implies

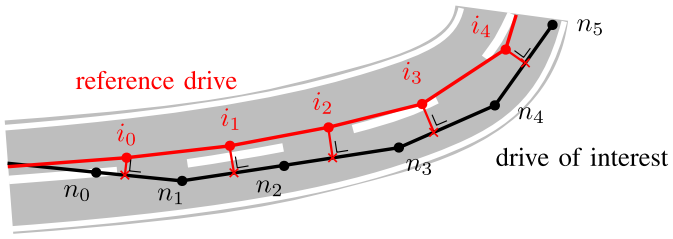


Fig. 4. Drive matching method, in which for each of the points i of the reference drive (red), the ratio of the linear line segment that connects two points n yielding the shortest distance is calculated.

that their time signals are expressed relative to the reference drive. This implies that time-domain operations need to be considered with caution.

D. Apparatus

The experiment was performed on the "Sapphire Space" simulator at BMW Group in Munich (see Figure 1, top left), a custom designed simulator constructed by Van Halteren Technologies in 2021. Its kinematic structure consists of three motion subsystems: the base is formed by a 19.14 m×15.70 m xy-drive that allows for large excursions in the x and y directions. On top of the xy-drive stands a large 1.15 m stroke hexapod that can move in all six Degrees of Freedom (DoFs). Finally, on top of the hexapod, a 360° yaw-drive is installed, allowing for additional yaw rotations of ±180°. The total motion system thus has nine DoFs. The rating model of [3] was derived from data collected on the similar, but smaller, "Ruby Space" simulator at BMW Group (xy-drive: 1.6×1.5 m, yaw-drive: ±25°, hexapod stroke: 0.34 m).

A one-to-one mock-up of a BMW 3 series (G20) was used, which was fully enclosed by the simulator dome. Visuals were rendered on the inner dome wall using 12 Norxe P1 projectors, resulting in a full 360° projection around the mock-up. During the open-loop drives, the steering wheel remained stationary. The iDrive navigation knob on the center console was used as the rating interface by the drivers to give the continuous ($R(t)$) and maneuver-based ratings (MB_{PH}), see Figure 1. The overall rating was extracted verbally for consistency with [3]. The 360° projection screen showed the visuals and the current rating value in the form of a "rating bar" [2]. The size and color of the rating bar changed (See screenshot in Figure 3) from rating 0 (short, white) to rating 10 (long, red), to make the rating method more intuitive for drivers to use. The velocity of the vehicle was visible on the tachometer on the dashboard and in the out-of-the-window visuals, together with the driving direction (arrows). The rating knob was connected to the central simulation computer using a CAN bus. This allowed for the accurate and consistent synchronization between recordings of the simulator motion and the rating signals of the participants.

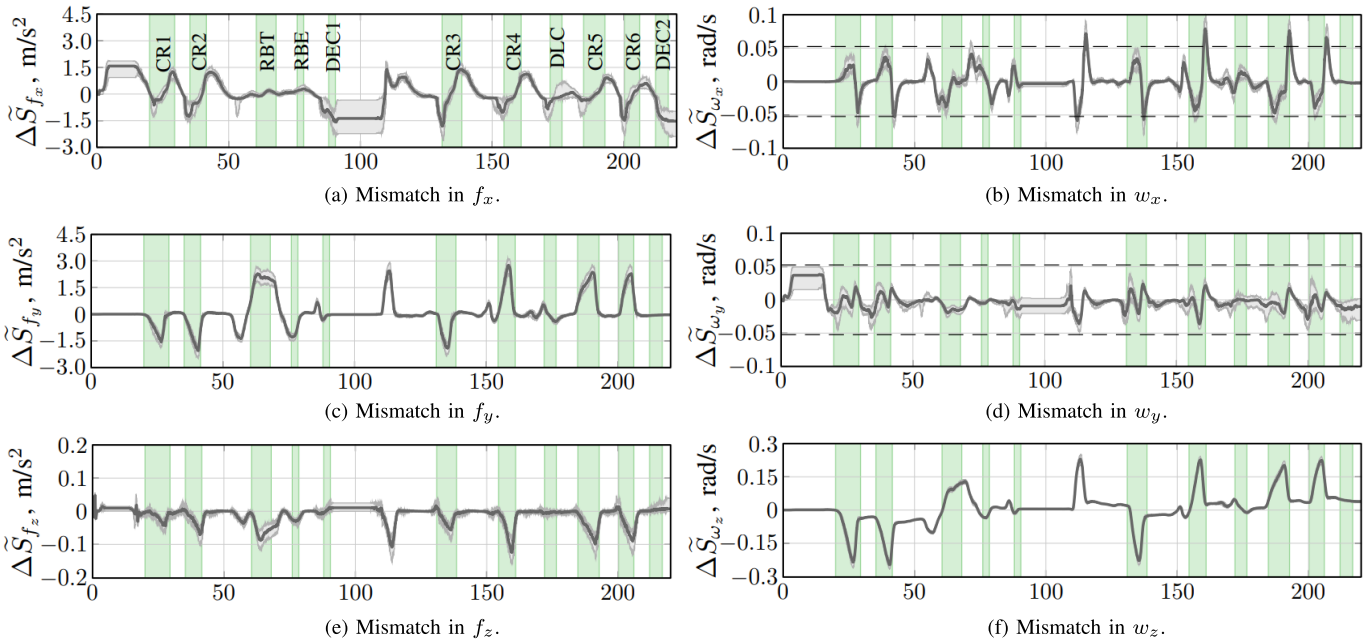
E. Motion Cueing Algorithm

A CWA was used as the MCA, as its linear filter-based structure ensures a deterministic output. As the motion cueing is calculated in real-time (see Figure 1), this is required to ensure that *identical* simulator motion is generated between the closed and open-loop driving conditions. The median mismatch signals of the MCA are shown in Figure 5, with the grey areas the interquartile ranges, and with the green areas representing the maneuvers. The CWA tuning did not fully utilize the motion system capabilities, to ensure that the limits were never reached. Tilt-coordination was used and tuned to keep the roll and pitch rate mismatches (Figures 5b and 5d) below the perceptual threshold of 3 deg/s [25] (dashed lines). In longitudinal direction, drivers drove more aggressive than expected, resulting in the median pitch rate slightly exceeding its perceptual threshold mismatch (Figure 5b).

The scaling factors used in the MCA were set to 0.5 for the specific forces and 0.6 for the rotational rates. These values lie well within the range of scaling factors considered to be the most realistic, i.e., $0.4 - 0.8$ found by [7]. First-order filters were used to distribute the low- and high-frequency motion across the motion subsystems (i.e., the xy-drive, the hexapod, and the yaw-drive). The break frequencies were set to 30 rad/s for the translational axes, such that motion below that frequency was reproduced by the xy-drive, whereas high-frequency accelerations were reproduced by the hexapod. A higher value of 50 rad/s was used for the yaw motion, such that the majority of yaw motion was reproduced by the yaw-drive, giving the hexapod more workspace to reproduce the roll and pitch motion. Finally, the lowest-frequency specific force motions in x and y directions were reproduced by the hexapod tilt-coordination, implemented using a low-pass filter break frequency of 0.5 rad/s.

Because all drivers drove themselves, the MCA output of each closed-loop drive is different. It is the longitudinal specific force mismatch ($\Delta\tilde{S}_{f_x}$, Figure 5a) that shows the largest spread, larger than $\Delta\tilde{S}_{f_y}$ (lateral specific force mismatch, Figure 5c) and $\Delta\tilde{S}_{\omega_z}$ (yaw rate mismatch, Figure 5f). This can be explained by the more varying nature of the driving behaviour in the longitudinal direction (i. e., braking and accelerating at different points in time) [26], whereas the lateral and yaw mismatches are mostly determined by the road shape [27] and result in more similar experiences across all drives. Furthermore, although there were only two distinct braking maneuvers in the maneuver-based conditions (DEC1 and DEC2), this does not mean that there was no longitudinal maneuvering present in the other maneuvers. In fact, as is visible in Figure 5a, the longitudinal specific force mismatch was also present during corner maneuvers, where participants braked into and accelerated out of the corner. Thus, the ratings of these corner maneuvers should also partially consist of a response to the longitudinal specific force mismatch.

Between 10-20 s, and 90-110 s, a constant average mismatch is present in all six signals. In the reference drive, the vehicle is standing still here, such that the drive matching approach selects the same position of the other drives. In these other drives, however, the vehicle can still be moving. This leads to constant values for as long as the reference vehicle is standing still. Therefore, all further time-domain operations (such as applying the rating model) are calculated for each drive separately, rather than using the median mismatch.



(a) Mismatch in f_x. (b) Mismatch in w_x.

(c) Mismatch in f_y. (d) Mismatch in w_y.

(e) Mismatch in f_z. (f) Mismatch in w_z.

Fig. 5. Median mismatches (grey lines) and interquartile ranges (grey shaded areas) over all drives between the reference and simulator as a function of equivalent time in seconds (participant 1, run 1). Green areas are the maneuvers; dashed lines in 5b and 5d are the perceptual thresholds (±3 deg/s, [25]).

F. Participants and Procedures

The experiment was performed by forty-two participants to ensure a large enough sample size [28] and to account for possible dropouts due to simulator sickness and/or technical problems. All participants were BMW employees and had a European car driver's license B for at least five years ($M = 14.5$ years, $SD = 9.1$ years) and an average yearly driven distance of $M = 16,278$ km ($SD = 16,408$ km). The average age was $M = 32.9$ years ($SD = 9.4$ years). Thirty drivers had previous experience in driving simulators. All drivers provided informed consent and the experiment was approved following BMW's internal ethics review procedures. Due to drop-outs (technical issues or simulator sickness), 36 complete data sets were obtained. The incomplete data sets of the drop-outs are not considered in further analysis.

All experiment sessions were ran by a single experimenter to ensure consistency in the interaction with the participants. All participants completed the experiment in a single session. Before entering the simulator, they all read a written briefing that explained the three rating methods, as well as the rating scale. All drivers performed one training drive for each of the three conditions, to get accustomed to the simulator, the sensation of motion, and the rating methods. The training drive of the OLCT condition contained inverted longitudinal (f_x) motion, to create large false cues and anchor the highest value of the rating scale (10), as in [3]. The OLCT and OLMB training drives were not based on the participant's own CLMB training drive, but used a pre-recorded drive, such that the anchoring of the rating scale was identical for all participants.

Drivers were instructed to drive as they normally would. As the closed-loop drive recordings were played back in the open-loop conditions, the CLMB condition was always tested first. For half of the drivers this was followed by OLMB and then by OLCT. For the other half, the order of the open-loop conditions was switched to average out order effects. Drivers performed three repetitions of each condition, resulting in a total of nine runs. The open-loop drives followed a different order than the closed-loop drives (1)-2-3): For OLMB (2)-3-1) and for OLCT (3)-1-2), to minimize recognition of the drives.

For the maneuver-based ratings, drivers were asked to give their impression of the maneuvers (Figure 5) using the rating knob. Drivers were instructed to rotate towards their intended rating, leave the rating at this value for at least two seconds, and then rotate back to zero. The selected maneuvers were spaced to give drivers enough time to give their rating and refocus on the driving task.

The recorded continuous, maneuver-based and overall rating signals are represented by $R^{cjp}(t)$, MB^{cjp}_{PH}, and OR^{cjp}_{PH}, respectively. Here, subscript c represents the experimental condition, j the condition repetition and p the driver. Note that if, in further notation, the subscript is missing, this indicates that the average along this dimension is taken.

IV. Results

A. Modeling of Continuous Ratings

Figure 6 shows the measured median continuous ratings (blue) over all drives. Given that the rating scale runs from 0 (congruent motion) to 10 (highly incongruent motion), ratings are generally low (< 2), i.e., the MCA setting was rated well. The rating peaks generally coincide with the end of the maneuver (vertical line), showing the lagged response to the incongruences, as expected from the estimated rating dynamics represented in the rating model in Eq. 1. The figure also shows the model of [3] (grey), which predicts the peaks of the continuous ratings quite well, although the quality of the fit is low at VAF= 11.0%. This VAF excludes the initial acceleration (between 0 and 20 s) and final deceleration 'DEC2' sections, as these were generally ignored by most

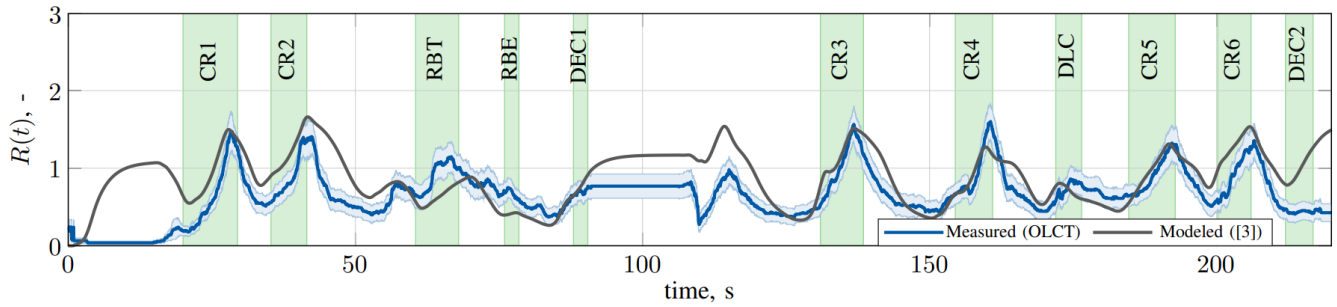


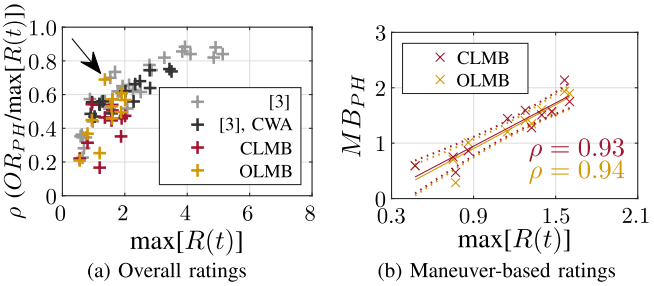
Fig. 6. Mean continuous motion incongruence rating (blue line) and standard error (blue shaded area) over all OLCT drives as a function of equivalent time in seconds (participant 1, run 1). The predicted rating based on the model of [3] is displayed by the grey line. Green areas are the maneuvers, although these were not highlighted to the participants in this OLCT condition.

participants; it might have been unclear to them that these were also to be rated. A second point of interest lies at the plateau between 90-110 s which is, as explained in Subsection III-E, caused by the drive matching approach. For further model calculations, the modeled ratings of the rating model are calculated for each drive separately, such that this plateau is not present (but cannot be compared in a single figure).

B. Rating Relationships

1) Relationship Overall/Continuous Ratings: A relationship linking the continuous rating to the overall ($OR_{PH} = f_{OR_{PH}}[R(t)]$) [path (2a) in Figure 2] is investigated. For the overall ratings, these fits are determined for each maneuver separately. Thus, each maneuver has a regression coefficient as to how much it correlates to the overall rating. As this requires a single data point for the continuous ratings in each maneuver, the continuous ratings are summarized through three methods: i) the maximum of $R(t)$ of the maneuver (CLMB: $\rho = 0.46$, OLMB: $\rho = 0.69$, ii) the mean (CLMB: $\rho = 0.46$, OLMB: $\rho = 0.65$), and iii) the median (CLMB: $\rho = 0.40$, OLMB: $\rho = 0.44$). A value closer to 1 indicates a stronger linear relationship, such that the maximum best explains the relationship between the rating methods. These values correspond to the roundabout ('RBT'), for which the correlation was always highest. Figure 7a shows how well the overall rating correlates to the maximum rating of each maneuver, expressed as the maximum continuous rating of that maneuver. Here, the grey values show the correlation values as determined by [3], the dark grey indicates such data points that correspond to a CWA condition. The red (CLMB) and orange (OLMB) points indicate the present study. To obtain an explicit predictive relationship, the regression fit with the highest Pearson correlation ($\rho = 0.69$, indicated by the arrow in Figure 7a) in the OLMB condition is taken: $OR_{PH} = 0.79 \cdot \max[R(t)] + 1.63$.

2) Relationship Maneuver-Based/Continuous Ratings: To investigate the relationship between continuous and maneuver-based ratings ($MB_{PH} = f_{MB_{PH}}[R(t)]$), [path (2b) in Figure 2], also the Pearson correlation is calculated. Here, the applied method differs from the overall rating. As a single data point exists in each maneuver for both the continuous and maneuver-based rating methods, a single regression fit can be made on all data points of the various maneuvers



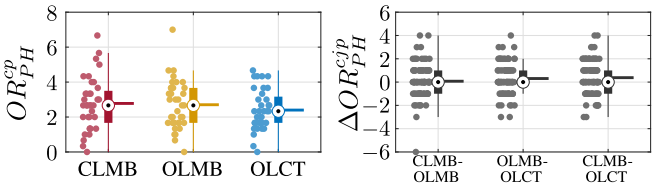
(a) Overall ratings (b) Maneuver-based ratings

Fig. 7. Correlations between the maximum of the continuous ratings and the overall ratings (a) / maneuver-based ratings (b). The data points represent each maneuver as defined in Figure 5. In (a), the arrow indicates the maneuver ('RBT') with the highest correlation. In (b), the lines are the regression fits, the dotted lines represent the 95% confidence bounds.

together. The continuous ratings are again summarized through three methods: i) the maximum of $R(t)$ in that maneuver (CLMB: $\rho = 0.93$, OLMB: $\rho = 0.94$, ii) the mean (CLMB: $\rho = 0.80$, OLMB: $\rho = 0.76$), and iii) the median (CLMB: $\rho = 0.59$, OLMB: $\rho = 0.49$). Similar to the overall ratings, it is the maximum of the continuous rating in the maneuver with the highest Pearson correlation (Figure 7b) that is the best predictor for the maneuver-based ratings: $MB_{PH} = 1.32 \cdot \max[R(t)] - 0.29$.

C. Equivalence of CL/OL Ratings

1) Overall Ratings: The overall rating distributions are shown in Figure 8a. The OLCT is also shown for reference, as the overall ratings were recorded in all three conditions. The box plots show the median (circles), the box edges indicate the 25th and 75th percentiles, and the whiskers show the range of the non-outlier data points. All individual data points are plotted as dots. The horizontal bars represent the means of the distributions. The data are normally distributed; the means for the CLMB, OLMB, and OLCT conditions are 2.78, 2.70, and 2.40, respectively, showing that the OLCT condition was rated slightly lower. The Bayes factor of the single effect between the CLMB and OLMB conditions [path (3a) in Figure 2] is $BF_{incl} = 0.263$ (Table II under 'OR_{PH}'), indicating moderate evidence of equivalence (< 0.3) [12]. Note that when including the overall ratings obtained in the OLCT condition, the Bayes factor increases to $BF_{incl} = 0.419$, providing no more evidence of equivalence. However, as the prime focus



(a) Overall ratings. (b) Overall rating differences.

Fig. 8. Overall ratings of the three tested conditions.

TABLE II

BAYES FACTORS OF THE DRIVING METHOD, MANEUVER, AND THEIR INTERACTION, FOR THE OVERALL (OR_{PH}) AND MANEUVER-BASED (MB_{PH}) RATINGS. BOLD VALUES INDICATE EQUIVALENCE (< 0.3)

	Effect	BF_{incl}
OR_{PH}	CLMB/OLMB	**0.263**
	CLMB/OLMB/OLCT	0.419
MB_{PH}	CLMB/OLMB	**0.143**
	Maneuver	$1.018 \cdot 10^{14}$
	CLMB/OLMB×Maneuver	**0.023**

here is on the comparison between CLMB and OLMB, this does not affect any further conclusions on equivalence between closed-loop and open-loop driving.

Even though the CLMB and OLMB overall ratings are equivalent, individual differences can still be present. For example, two drivers could have rated the two conditions differently, but in exactly opposite ways. Although this would lead to equivalent data, it would ignore insights into individual differences. Figure 8b shows the distributions of ΔOR_{PH}^{cjp}, i.e., the difference in overall ratings per condition pair of each individual run pair. With the presence of the OCLT condition, this results in three ΔOR_{PH}^{cjp} distributions. As this only has one distribution per condition pair, no statistical test is possible. The horizontal bars indicate the means: 0.074 for CLMB - OLMB (median = 0), showing that individuals rated the CLMB and OLMB conditions the same. Furthermore, for OLMB - OLCT the mean is 0.31 (median = 0), and for CLMB - OLCT: 0.38 (median = 0). These mean values show that OLCT was also rated slightly lower *within* individuals.

2) Maneuver-based Ratings: Equivalence of the maneuver-based ratings is investigated next [path (3b) in Figure 2]. Figure 9a shows the distributions of the maneuver-based ratings of the CLMB and OLMB conditions for each maneuver. Differences exist between the maneuvers, with CR3 the worst rated maneuver (i.e., the highest means). The corners (involving lateral and yaw motion) are generally rated the worst (e.g., CR3 and CR4), whereas maneuvers involving longitudinal motion (DEC1 and DEC2) are rated best.

To investigate the equivalence of the two conditions, the Bayes factors are calculated. The results are shown in Table II under 'MB_{PH}'. Three possible effects are analyzed for the MB_{PH} data: 'CLMB/OLMB', 'Maneuver' and 'CLMB/OLMB×Maneuver', where the latter represents the interaction effect. For the maneuver effect, the BF_{incl} is $1.018 \cdot 10^{14}$, indicating extremely decisive evidence (> 30) [12] that the maneuvers were rated differently.

In contrast, when considering CLMB/OLMB, $BF_{incl} = 0.143$, providing moderate to strong evidence [12] that the maneuver-based ratings of the two conditions are equivalent, supporting the earlier findings on equivalence of the overall ratings. For the combination of the two effects, no interaction effect exists between the CLMB/OLMB and the maneuvers ($BF_{incl} = 0.023$). This indicates that the equivalence within the driving method does not depend on the (type of) maneuver. Thus, although the maneuvers are rated differently, these differences are *equivalent* in the CLMB and OLMB conditions.

Similar to the analysis on the overall ratings, Figure 9b shows the distributions of ΔMB_{PH}^{cjp}, i.e., the difference of each individual run pair. All medians are 0, and the means are generally very close to 0 (highest $\Delta MB_{PH}^{cjp} = 0.24$, for 'RBT'). This provides further evidence that the drivers rated both conditions equivalently.

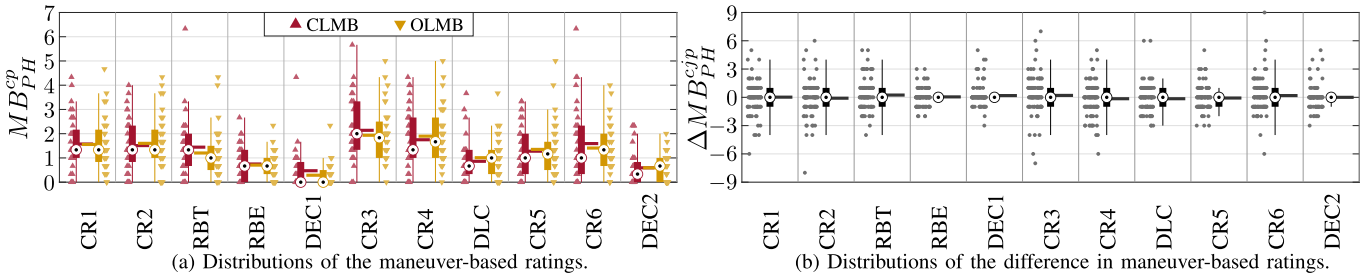
D. Rating Prediction Framework Evaluation

The three steps defined in Figure 2 have now been evaluated. First, due to their equivalence, maneuver-based ratings of open-loop drives can be used to predict ratings of closed-loop drives (see red and orange data in Figure 10, representing their means). Second, using the estimated regression fits that relate the overall and maneuver-based ratings to the (measured) continuous ratings, both the overall and maneuver-based ratings can be predicted (blue). This holds for both ratings of open-loop and closed-loop driving due to their equivalence. Third, the continuous ratings can be predicted using a rating model, based on objective mismatch signals (grey). Therefore, the steps combined allow for predicting maneuver-based and overall ratings of closed-loop drives using a continuous rating model. Between the predicted maneuver-based ratings of the rating model and the measured closed-loop ratings, the deviations are smaller than half a rating point. Considering a ten-point rating scale, where only steps of 1 were possible, these errors can be considered acceptable. Exceptions are 'RBE' and 'DEC2', where the differences are 0.65 and 0.79, respectively. For the overall post-hoc ratings, the rating predictions also work well, with a difference between the measured closed-loop (red) and the modeled (grey) ratings of 0.16.

V. DISCUSSION

A. Model Predictions

The model proposed by [3] was used to predict the measured continuous ratings as a function of the objective mismatch between vehicle reference and simulation motion. Using the same model parameters of [3] resulted in a reasonably accurate prediction of the ratings. Although the VAF was low, the resulting predictions of the maneuver-based and overall ratings were accurate. Between the present work and the model of [3], the scenario, the rating set-up, instructions, and the MCA were the same. However, the simulator, the MCA parameters, and the participant group were different, which may have affected the ratings. Overall, the presented results show that



(a) Distributions of the maneuver-based ratings. (b) Distributions of the difference in maneuver-based ratings.

Fig. 9. Maneuver-based ratings for CLMB and OLMB conditions per maneuver.

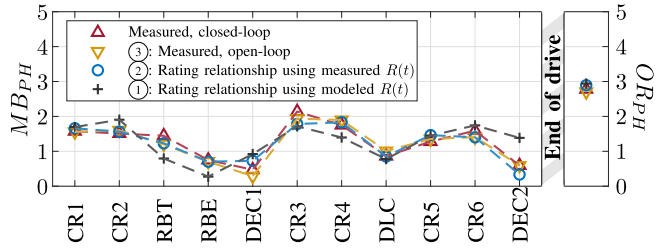


Fig. 10. Steps of determining maneuver-based (left) and overall (right) ratings.

the predictive model still provides accurate results across these variables when the averages of these participant groups are considered. For these three variables there is thus no combined effect. This shows that the rating prediction methodology is effective even across these experiment variables and can thus be applied for predictions of motion cueing quality of future, as of yet untested, driving simulator studies.

Especially the difference in simulator is notable. The nine DoF Sapphire Space simulator used in the present study was significantly larger than the nine DoF Ruby Space simulator, on which the model of [3] was developed. The smaller workspace of the simulator in [3] resulted in larger mismatches compared to the present experiment, on which the rating model was fit. This thus also shows that the model still works when considering different ranges of incongruent motion. This is an important quality, as it shows the general applicability of the model across various simulators, which is an important property considering the various simulators to choose from at BMW's Driving Simulation Center. For smaller, hexapod-only systems, the rating model, including its ability to predict ratings of closed-loop driving, can be further tested by applying it to predict the motion cueing quality of practical driving experiments. For example, the effectiveness of the rating model can be confirmed by comparing predictions and actually obtained overall post-hoc ratings, as the latter can be obtained with limited interference in the experiment itself.

Predicting ratings for significantly smaller simulators (e.g., small hexapods) might in fact prove to be the largest future difficulty, especially for systems that cannot approximate the high cueing quality of the Sapphire Space and the Ruby Space. As the applied rating scale has a fixed lower anchoring ('no incongruence = 0'), but no upper anchoring ('large incongruence = 10'), it is possible that the rating that drivers associate with 'large incongruence' can depend on the intensity of the incongruences presented in the experiment. [29] showed that in some cases, transferability between experiments can be an issue if the difference in presented motion is too large (e.g., between a low- and high-fidelity simulator). She introduced a *Model Transfer Parameter (MTP)* to apply a linear scaling between the acquired ratings, normalized for the presented motion in each experiment. A further investigation on transferability between experiments is therefore suggested, in which larger differences between the motion are present, as this would need to be corrected for in the rating model.

Another crucial direction for future work is to investigate the validity of the findings and the rating model prediction in different driving scenarios. Here, the first step would be to test a different urban route, as this might alter the balance of the presented mismatches and therefore require the introduction of an MTP. Second, extending the results to completely different scenario types (e.g., highway or rural) is another important step. As discussed in [3], different scenario's can, for example, induce more interaction with surrounding traffic, which may induce different types of motion (e.g., more lane changes in highway scenarios). Here, maneuvers may be more difficult to rate, as anticipating responses to traffic is more difficult than the road-geometry driven maneuvers of an urban scenario.

B. Relationships Between Rating Signals

To understand how the overall (OR_{PH}) and maneuver-based (MB_{PH}) ratings relate to the continuous ratings ($R(t)$), it was determined which metrics best correlate. Analyzing the correlation between the maximum of the continuous ratings per maneuver and the overall ratings, it can be concluded that the higher the maximum of the continuous ratings, the more these ratings correlate with the overall ratings. This reproduces findings by [3]. The point with the highest correlation ($\rho = 0.69$) was the roundabout maneuver 'RBT'. The analysis between the maneuver-based and continuous ratings shows a similar result: the maneuver-based ratings are highly correlated with the maximum of the continuous ratings in that maneuver ($\rho = 0.94$). These results show that maneuver-based and overall ratings can be predicted using continuous ratings.

Two limitations remain, however. The correlation analysis could only be applied on the average driver level, rather than for each drive separately. This analysis might therefore be somewhat confounded due to the inherently different drives that were present, which can affect the correlation. Second, the acquired relations could only be evaluated for a limited part of the rating scale. Although this range of presented motion in the present experiment corresponded to a realistic MCA for the considered simulator, further research could investigate how these relationships hold at better or worse motion cueing. Similarly, it is suggested to extend the findings

on correlations towards other scenarios. As the urban driving scenario generally results in reliable rating data [3], less strong relationships might be present in scenarios that inherently have a lower reliability, such as rural [17] or highway scenarios.

The maneuver-based rating method itself, using the rating knob after each maneuver, worked well and was noted by participants to be an intuitive task. This might thus be a suitable alternative to the commonly used overall ratings.

C. Equivalence of Closed- and Open-Loop Ratings

Through the estimation of Bayes factors, the ratings of closed-loop and open-loop driving of the overall ratings were shown to be equivalent. For the maneuver-based ratings, the driving methods (CL/OL) also show equivalence, whereas the maneuvers are rated differently. No interaction effect exists between the driving methods and the maneuvers.

The equivalence analysis further shows that the differences in simulator motion as perceived in the various maneuvers did have an impact on the provided ratings, as expected based on the between-maneuver objective cueing error variations. The lack of a significant interaction effect indicates that these differences between maneuvers are equivalent for closed-loop and open-loop driving. Differences in ratings are therefore caused by the differences in maneuver, and not by whether closed-loop or open-loop driving is active. The implications of these results are two-fold. First, it enables using (predictions of) the continuous rating method to identify where and to which extent incongruences occur with high resolution. Second, it enables predicting maneuver-based and overall ratings with high accuracy, which is especially useful for comparisons of motion cueing (i.e., MCA "A" is better than MCA "B").

A main application for these results is to improve methods to objectively select the best possible motion cueing settings (simulators, MCAs, parameters) prior to inviting participants for closed-loop testing. To do this well, a prediction of drivers' PMIs as a function of a simulator's (objective) movement (i.e., the paper's main contribution) is crucial. While the application of our findings is useful for all driving simulators, it is especially important for experiments at BMW, due to the wide range of different simulators and MCAs available. Thus, the presented work can be directly used to improve the decision making for driving simulation motion cueing selection.

Even though the ratings were equivalent, note that the underlying perception does not necessarily have to be equivalent as well. It has been shown [10], [11] that perceptual thresholds can in fact change under closed-loop and open-loop single-axis settings, hinting at differences in perception. However, even if these perceptual differences would be present in multi-axis car driving simulations, the equivalent ratings show that these differences are small enough to not be of practical significance.

A point of attention lies within the fact that the participants had to rate their own drives, also in the open-loop conditions. This was a crucial choice, as it allowed for the explicit comparison between open-loop and closed-loop driving. Although participants were not told that they would rate their own drives in the open-loop conditions, a potential bias could occur when participants recognize their own drives: then their rating could be affected by their memory of what the motion felt like in the closed-loop condition. To mitigate this, the order in which the three drives were presented in the open-loop conditions was different than that of the closed-loop drives. Furthermore, while the vehicle's trajectory was replicated directly, the traffic in the simulation was still random every time.

The final point of attention concerns the OLCT condition, which resulted in consistently lower overall ratings, relative to both the CLMB and OLMB conditions. It is possible that the continuous rating method itself affects the rating measurements. For example, continuous ratings require more workload than maneuver-based ratings, potentially decreasing drivers' sensitivity to motion incongruences. Furthermore, in the OLCT condition participants rated the complete driving scenario, while in the maneuver-based drives (OLMB and CLMB) participants only focused on the outlined maneuvers. Even though the overall rating is intended to represent the whole drive, it is possible that the OLMB and CLMB conditions are biased towards the maneuvers rated in those conditions, which are the most incongruent points. Therefore, the overall ratings in these conditions might be higher than the OLCT results.

VI. Conclusion

This paper described a driving simulator experiment of which the data of 36 participants was used to develop a method to predict motion incongruence ratings of closed-loop driving through three key findings. First, a model of continuous rating signals from literature was validated by showing it can successfully predict the measured continuous ratings. Second, the maximum of the continuous ratings (i.e., the worst motion) was shown to correlate strongly with the drivers' overall ($\rho = 0.69$) and maneuver-based ratings ($\rho = 0.94$). This allows for predicting such ratings based on measured and modelled continuous rating signals. Third, performing a Bayes analysis showed that both maneuver-based and overall ratings are equivalent between closed-loop and open-loop driving methods. All findings combined show that the open-loop continuous rating method is a valid method for obtaining high-resolution information on incongruences of closed-loop driving. Moreover, it shows that both overall and maneuver-based ratings of closed-loop driving can be predicted through objective mismatch signals between vehicle and simulator motion. Both allow for improved objective predictions of subjective ratings to guide the design, testing, and assessment of future motion cueing algorithms, while greatly reducing the required on-site simulator testing time.

References

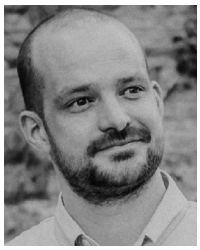
[1] M. R. C. Qazani et al., "A new prepositioning technique of a motion simulator platform using nonlinear model predictive control and recurrent neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 23268–23277, Dec. 2022.

[2] D. Cleij, J. Venrooij, P. Pretto, D. M. Pool, M. Mulder, and H. H. Bülthoff, "Continuous subjective rating of perceived motion incongruence during driving simulation," *IEEE Trans. Hum.-Mach. Syst.*, vol. 48, no. 1, pp. 17–29, Feb. 2018.

[3] M. Kolff, J. Venrooij, M. Schwienbacher, D. M. Pool, and M. Mulder, "Reliability and models of subjective motion incongruence ratings in urban driving simulations," *IEEE Trans. Hum.-Mach. Syst.*, early access, Sep. 18, 2024, doi: 10.1109/THMS.2024.3450831.

[4] T. Irmak, D. M. Pool, and R. Happee, "Objective and subjective responses to motion sickness: The group and the individual," *Exp. Brain Res.*, vol. 239, no. 2, pp. 515–531, Feb. 2021.
[5] S. Casas-Yrurzum, C. Portalés-Ricart, P. Morillo-Tena, and C. Cruz-Neira, "On the objective evaluation of motion cueing in vehicle simulations," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 3001–3013, May 2021.
[6] A. Lamprecht, D. Steffen, K. Nagel, J. Haecker, and K. Graichen, "Online model predictive motion cueing with real-time driver prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12414–12428, Aug. 2022.
[7] A. Berthoz, "Motion scaling for high-performance driving simulators," *IEEE Trans. Hum.-Mach. Syst.*, vol. 43, no. 3, pp. 265–276, May 2013.
[8] F. Ellensohn, J. Venrooij, M. Schwienbacher, and D. Rixen, "Experimental evaluation of an optimization-based motion cueing algorithm," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 62, pp. 115–125, Apr. 2019.
[9] M. Kolff, J. Venrooij, M. Schwienbacher, D. M. Pool, and M. Mulder, "Quality comparison of motion cueing algorithms for urban driving simulations," in *Proc. Driving Simul. Conf. Eur.*, Munich, Germany, 2021, pp. 141–148.
[10] A. Nesti, S. Nooij, M. Losert, H. H. Bülthoff, and P. Pretto, "Roll rate perceptual thresholds in active and passive curve driving simulation," *Simulation*, vol. 92, no. 5, pp. 417–426, May 2016.
[11] A. R. V. Pais, D. M. Pool, A. M. de Vroome, M. M. Van Paassen, and M. Mulder, "Pitch motion perception thresholds during passive and active tasks," *J. Guid., Control, Dyn.*, vol. 35, no. 3, pp. 904–918, May 2012.
[12] H. Jeffreys, *Theory of Probability*. Oxford, U.K.: Oxford University Press, 1961.
[13] M. Kolff, J. Venrooij, M. Schwienbacher, D. M. Pool, and M. Mulder, "The importance of kinematic configurations for motion control of driving simulators," in *Proc. IEEE 26th Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2023, pp. 1000–1006.
[14] M. Sivak, "The information that drivers use: Is it indeed 90% visual?" *Perception*, vol. 25, no. 9, pp. 1081–1089, Sep. 1996.
[15] H. G. Stassen, G. Johannsen, and N. Moray, "Internal representation, internal model, human performance model and mental workload," *Automatica*, vol. 26, no. 4, pp. 811–820, Jul. 1990.
[16] M. Mulder, D. M. Pool, K. van der El, and R. M. van Paassen, "Neuroscience perspectives on adaptive manual control with pursuit displays," *IFAC-PapersOnLine*, vol. 55, no. 29, pp. 160–165, 2022.
[17] F. Ellensohn, M. Spannagl, S. Agabekov, J. Venrooij, M. Schwienbacher, and D. Rixen, "A hybrid motion cueing algorithm," *Control Eng. Pract.*, vol. 97, Apr. 2020, Art. no. 104342.
[18] M. Kolff, J. Venrooij, M. Schwienbacher, D. M. Pool, and M. Mulder, "Motion cueing quality comparison of driving simulators using Oracle motion cueing," in *Proc. Driving Simul. Conf. Eur.*, Strasbourg, France, 2022, pp. 111–118.
[19] D. Lakens, "Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses," *Social Psychol. Personality Sci.*, vol. 8, pp. 1–8, May 2017.
[20] D. Lakens, A. M. Scheel, and P. M. Isager, "Equivalence testing for psychological research: A tutorial," *Adv. Methods Practices Psychol. Sci.*, vol. 1, no. 2, pp. 259–269, Jun. 2018.
[21] M. Linde, J. Tendeiro, R. Selker, E. Wagenmakers, and D. van Ravenzwaaij, "Decisions about equivalence: A comparison of tost, hdi-rope, and the Bayes factor," *Psychol. Methods*, vol. 28, no. 3, pp. 740–755, 2023.
[22] W. M. Bolstad and J. M. Curran, *Introduction To Bayesian Statistics*, 2nd ed., Hoboken, NJ, USA: Wiley, 2007.
[23] JASP Team. (2023). *JASP (Version 0.17)[Computer Software]*. [Online]. Available: https://jasp-stats.org/
[24] R. Kelter, "Bayesian alternatives to null hypothesis significance testing in biomedical research: A non-technical introduction to Bayesian inference with JASP," *BMC Med. Res. Methodol*, vol. 20, no. 1, pp. 1–12, Dec. 2020.
[25] G. Reymond and A. Kemeny, "Motion cueing in the Renault driving simulator," *Vehicle Syst. Dyn.*, vol. 34, no. 4, pp. 249–259, Oct. 2000.
[26] J. Eppink, M. Kolff, J. Venrooij, D. M. Pool, and M. Mulder, "Probabilistic prediction of longitudinal driving behaviour for driving simulator pre-positioning," in *Proc. Driving Simul. Conf. Eur.*, Antibes, France, 2022, pp. 119–126.
[27] P. Bosetti, M. Da Lio, and A. Saroldi, "On the human control of vehicles: An experimental study of acceleration," *Eur. Transp. Res. Rev.*, vol. 6, no. 2, pp. 157–170, Jun. 2014.
[28] B. Cai and X. Wang, "Estimation of the smallest acceptable sample size in bilateral approaches to coefficient estimation and accuracy prediction," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2678, no. 8, pp. 690–699, Aug. 2024.
[29] D. Cleij, "Measuring, modelling and minimizing perceived motion incongruence," Ph.D. thesis, Fac. Aerosp. Eng., Delft Univ. Technol. (TU Delft), Delft, The Netherlands, 2020.

Maurice Kolff received the M.Sc. degree from Delft University of Technology (TU Delft), The Netherlands, in 2019. He is currently pursuing the joint Ph.D. degree with the Section Control and Simulation, TU Delft, and the Department of Research and Development, BMW Group, Munich, Germany. His work focuses on the evaluation methods of motion cueing in driving simulation. His research interests include motion perception, mathematical modeling, and model-predictive control algorithms.



Joost Venrooij received the M.Sc. and Ph.D. degrees (cum laude) in aerospace engineering from Delft University of Technology, Delft, The Netherlands, in 2009 and 2014, respectively. After the Ph.D. degree, he was a Project Leader with the Max Planck Institute for Biological Cybernetics, Tubingen, Germany. He is currently a Research Specialist with BMW Group, Munich, Germany. His research interests include motion cueing, driving simulation, and vehicle modeling.

Elena Arcidiacono received the B.Sc. degree from the Technical University of Munich, Germany, in 2022. She is currently pursuing the joint M.Sc. degree in medical technology and neuroscience with ETH Zürich, Switzerland, and the Technical University of Munich, Germany. Besides her research interest in the validation and development of driving simulators, she focuses on the fields of rehabilitation engineering, signal processing, and neurotechnology.

Daan M. Pool (Member, IEEE) received the M.Sc. and Ph.D. degrees (cum laude) from TU Delft, The Netherlands, in 2007 and 2012, respectively. He is currently a tenured Assistant Professor with the Control and Simulation Section, Faculty of Aerospace Engineering, TU Delft. His research interests include cybernetics, vehicle simulation, simulator-based training, and mathematical modeling, identification, and optimization techniques.

Max Mulder (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees (cum laude) in aerospace engineering from Delft University of Technology (TU Delft), Delft, The Netherlands, in 1992 and 1999, respectively, for his work on the cybernetics of tunnel-in-the-sky displays. He is currently a Full Professor and the Head of the Control and Simulation Section, Faculty of Aerospace Engineering, TU Delft. His research interests include cybernetics and its use in modeling human perception and performance, and cognitive systems engineering and its application in the design of "ecological" interfaces.