

Continuous Subjective Rating of Perceived Motion Incongruence During Driving Simulation

Cleij, Diane; Venrooij, Joost; Pretto, Paolo; Pool, Daan M.; Mulder, M.; Bulthoff, Heinrich H.

DOI

[10.1109/THMS.2017.2717884](https://doi.org/10.1109/THMS.2017.2717884)

Publication date

2018

Document Version

Final published version

Published in

IEEE Transactions on Human-Machine Systems

Citation (APA)

Cleij, D., Venrooij, J., Pretto, P., Pool, D. M., Mulder, M., & Bulthoff, H. H. (2018). Continuous Subjective Rating of Perceived Motion Incongruence During Driving Simulation. *IEEE Transactions on Human-Machine Systems*, 48(1), 17 - 29 . <https://doi.org/10.1109/THMS.2017.2717884>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Continuous Subjective Rating of Perceived Motion Incongruence During Driving Simulation

Diane Cleij, Joost Venrooij, *Member, IEEE*, Paolo Pretto, Daan M. Pool, *Member, IEEE*,
Max Mulder, *Member, IEEE*, and Heinrich H. Bülthoff, *Member, IEEE*

Abstract—Motion cueing algorithms are used in motion simulation to map the inertial vehicle motion onto the limited simulator motion space. This mapping causes mismatches between the unrestricted visual motion and the constrained inertial motion, which results in perceived motion incongruence (PMI). It is still largely unknown what exactly causes visual and inertial motion in a simulator to be perceived as incongruent. Current methods for measuring motion incongruence during motion simulation result in time-invariant measures of the overall incongruence, which makes it difficult to determine the relevance of the individual and short-duration mismatches between visual and inertial motion cues. In this paper, a novel method is presented to subjectively measure the time-varying PMI continuously throughout a simulation. The method is analyzed for reliability and validity of its measurements, as well as for its applicability in relating physical short-duration cueing errors to PMI. The analysis shows that the method is reliable and that the results can be used to obtain a deeper insight into the formation of motion incongruence during driving simulation.

Index Terms—Cueing, human factors, perception, simulation, simulator validation, virtual reality.

I. INTRODUCTION

MOTION-BASED vehicle simulators are used for a wide variety of applications. They are an increasingly important tool for training, research, and vehicle system development in both car [1] and aerospace industries [2]. However, one of the main challenges in motion-based simulation is to cope with the typically limited workspace of the simulator. To map the vehicle inertial motions onto the simulator motion space, a motion cueing algorithm (MCA) is used [3]. As the simulator motion space typically is much smaller than the vehicle motion space, this process inherently results in motion mismatches: differences between the unconstrained visual and the constrained

inertial motion cues. These mismatches result in a decrease of simulator motion fidelity and unrealistic simulations [4].

For motion simulation fidelity, a distinction is made between physical and perceptual motion fidelity [5]. Physical fidelity is defined as the match between objectively measured motion cues in the simulator and in the vehicle. Perceptual fidelity is defined as the match between simulator and vehicle motion cues as perceived by the human. The main reason for using a vehicle simulator is not to replicate the physical vehicle motions, but rather replicate the human perception of these motions [6]. van der Steen [7] investigated the effect of physical incongruence between visual and inertial motion on the perceived realism of the combined motion in a passive flight simulation. He introduced the term coherence zone for the range of inertial motion amplitudes that were still perceived as coherent with a given visual motion amplitude. In [8], the effect of motion frequency on these coherence zones in passive flight simulation is investigated and in [9] the term phase coherence zone is introduced as the range of phase shifts for which inertial and visual motion are still perceived as realistic. As in real vehicles, where all motion stimuli are congruent, motion simulators should provide inertial motions that are within these coherence zones. If this is not possible, at least the perceived incongruence between different motion stimuli should be minimal. The current study therefore focuses on measuring any, linear or nonlinear, incongruence between visual and inertial motion that is perceived in a passive vehicle simulation. The degree to which this incongruence results in unrealistic motion is hereby called the perceived motion incongruence (PMI).

To improve motion cueing, we need to understand how this PMI is related to the physical motion mismatches presented in the simulator. Currently, there are methods to directly or indirectly measure PMI, but they only provide time-invariant overall results. These discrete results can be used to quantify and compare the overall quality of an MCA, but cannot be correlated to the time-varying short-duration motion mismatches. It therefore remains unclear which motion mismatches are responsible for the overall PMI. A time-varying measure of PMI, which can be correlated to these mismatches, is therefore needed. Relevant motion mismatches can then be identified and, eventually, minimized. Besides being instrumental to improve motion cueing, such a measure can also be used to gain a better understanding of human motion perception.

Perceptual fidelity is measured using human-in-the-loop experiments. During these experiments, participants are usually

Manuscript received June 23, 2016; revised January 27, 2017; accepted May 14, 2017. This paper was recommended by Associate Editor M. Tanelli. (Corresponding author: Diane Cleij.)

D. Cleij is with the Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany, and also with the Faculty of Aerospace Engineering, Delft University of Technology, Delft 2629 HS, The Netherlands (e-mail: diane.cleij@tuebingen.mpg.de).

J. Venrooij, P. Pretto, and H. H. Bülthoff are with the Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany (e-mail: joost.venrooij@tuebingen.mpg.de; paolo.pretto@tuebingen.mpg.de; heinrich.buelthoff@tuebingen.mpg.de).

D. M. Pool and M. Mulder are with the Faculty of Aerospace Engineering, Delft University of Technology, Delft 2629 HS, The Netherlands (e-mail: d.m.pool@tudelft.nl; m.mulder@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/THMS.2017.2717884

subjected to vehicle simulations using different MCA tunings. This fidelity can currently be measured directly via questionnaires or subjective ratings on the MCA quality. In [10], information on MCA quality during car motion simulation was obtained via questionnaires after each simulation run and overall MCA quality ratings at the end of the experiment. In [11], the simulation fidelity rating scale together with an overall motion fidelity rating was used to subjectively rate the motion fidelity of a helicopter motion simulation for different MCAs. In both the cases, the only time-varying information of MCA quality was obtained via questionnaires on specific parts of the simulation. In [12], an offline rating (OR) method based on magnitude estimation with cross-modality matching was developed and used to detect differences between MCAs during car motion simulation. The MCA rating results obtained in these studies are time invariant and thus cannot be easily correlated with the time-varying motion mismatches. Direct objective and time-varying measures of PMI could possibly be done via physiological measures. Currently though only physiological measures related to motion sickness, a possible effect of sustained or extreme PMI, have been found. Physiological measures such as heart rate and skin temperature were measured and compared with the continuously rated subjective estimate of discomfort during a car simulation in [13], whereas in [14] similar physiological measures during a driving and flying simulation were compared to ORs of motion sickness. Instead of direct measurements, objective indirect measurements of PMI can be attained by observing the induced control behavior for different MCAs. In [15], different MCAs were analyzed based on objective measures such as overall control activity and tracking performance throughout an active driving simulation, whereas in [16] similar measurements during a flight simulation were used to determine the effect of heave washout filter settings on the parameters of a pilot model. To understand which differences in control behavior indicate a “better” MCA, in [17] and [18] control behavior in a simulator for different MCAs is compared to real in-flight and in-vehicle recordings, respectively. In [19], the effect of time-varying filter gains on pilot control model parameters is investigated. However, the changes in control behavior described by both time-invariant and time-variant models currently available do not have the temporal resolution needed to identify the relevant short-duration motion mismatches.

This paper therefore presents a novel subjective measurement method, which allows for measurement of the time-varying PMI continuously during a vehicle simulation (first described in [20]). The method is based on continuous subjective rating used in other research fields, such as a three-dimensional (3-D) television [21]. The validation of the novel method is done by analyzing the results of a human-in-the-loop experiment, where continuous rating (CR) of PMI was performed in a motion-based simulator during a passive driving simulation. First, the measurements are tested for reliability and validity. Subsequently, the applicability of this method is analyzed to determine if relevant short-duration motion mismatches can indeed be identified from the measured PMI. More information on the method and the validation process is given in Section II. Section III describes the experiment setup, whereas the results with respect to reliability, validity, and applicability of the method are presented in

Section IV. A discussion of these results and the corresponding conclusions are presented in Sections V and VI, respectively.

II. CR METHOD

A. Background

CR refers to an online subjective rating, based on the method of magnitude estimation [22]. This method allows for the measurement of the perceived intensity of any physical stimulus. In the more traditional OR method, the observer provides a single rating (magnitude) to a certain property of the sensory stimulus via a dedicated rating interface. In the CR method, the observer is asked to provide this rating continuously throughout the sensory stimulus, resulting in a rating that varies over time. In the field of 2-D/3-D television, CR methods are used to assess video quality by rating the visual stimuli on visual comfort. In [23], this method is used to relate the measured visual comfort to disparity and motion, whereas in [24] the influence of 3-D video properties, such as perceived depth, on the feeling of presence is rated. In the field of music analysis, CR is also used. In [25], the method is used to measure the predictability of music over time, whereas in [26] a CR method is used to relate levels of emotion to specific aspects of music. Finally, in [27] and [28], a CR method is used to gauge strain and workload in, respectively, a motion-based and a fixed-based driving simulator, continuously.

B. Procedure

The proposed rating method is based on the rating methods described above and used to measure PMI during a passive vehicle motion simulation in a motion-based simulator. The participants thus did not use the steering wheel or pedals, but were instead asked to continuously rate the PMI. The resulting motion incongruence rating (MIR) is a measure for the perceived incongruence between visual and inertial motion cues presented in the simulator.

The CR is performed using a dedicated rating interface, as shown in Fig. 1, consisting of a rotary knob to express the rating and a rating bar displayed on the screen, serving as visual feedback on the current rating. A maximum rating of one is given by turning the rotary knob fully to the right and this will result in a fully colored rating bar. A minimum rating of zero, given by turning the knob fully to the left, will result in a fully black rating bar. The method makes use of simulation trials: vehicle simulations of maneuvers of interest, each of which includes the complete range of motion incongruence that will be presented during a specific experiment. In the experiment described in this paper, the simulation trials all consist of the same segments, combinations of maneuver and MCA, but ordered differently for each trial. To anchor both ends of the rating scale, participants are instructed to provide the minimum rating of zero when no motion incongruence is perceived. When motion incongruence is perceived, the rating should increase proportional to the incongruence intensity, with the maximum rating anchored at the highest incongruence perceived during the simulation trial.

Participants can only use such a rating scale properly, if the maximum incongruence during the experiment is known. The complete range of motion incongruence presented during an

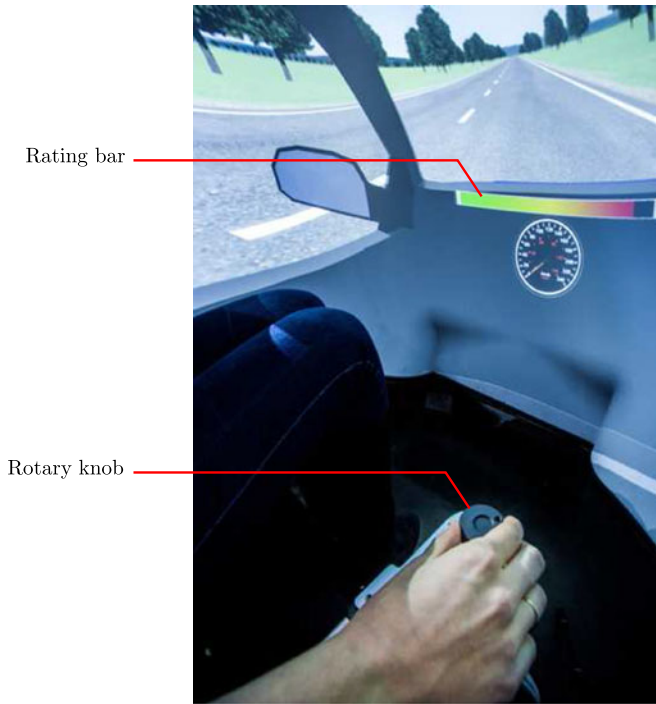


Fig. 1. Rating interface consisting of a rating device and a rating bar.

experiment thus should be observed at least once, before a proper rating can be performed. Therefore, participants first receive training that consists of two procedures: rating interface training and congruence range training, based on [29], and [23] and [24], respectively. During the rating interface training, participants familiarize themselves with the rating interface via a simple control task, where they are asked to follow a second automatically adjusted rating bar. Subsequently, in the congruence range training, the participants familiarize themselves with the full range of motion incongruence that can occur during the experiment. They also familiarize themselves with the task of rating this incongruence continuously. To this end, participants are instructed to continuously rate the motion incongruence during a simulation trial. The training is repeated several times to check if the participant can provide a consistent rating. At the end of this training, the participants thus should have learned to use the full range of the rating bar, i.e., when no motion incongruence is felt, it provides a rating of zero and when the maximum motion incongruence during the simulation trial is felt, it provides a rating of one.

In the measurement part of the experiment, participants are asked to continuously rate the motion incongruence in a simulation trial, using the rating interface. For verification of consistency of the rating, this procedure is repeated three times.

During the experiment described here, a second measurement, a retrospective OR, was done. For this OR, the simulation trial is split into several smaller segments. After observing a segment, participants are asked to provide one overall rating of the PMI during this segment using again the rating interface shown in Fig. 1. This OR method is commonly used to measure MCA quality [10], [11], [18] and is here assumed to be an accepted measure of PMI.

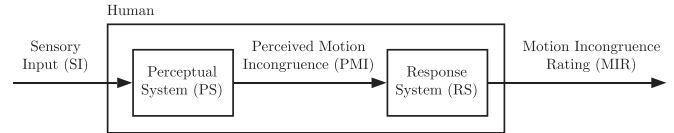


Fig. 2. Simplified block diagram of the human subject during the CR of PMI in a simulator.

C. Measurements

To better understand what is measured with the CR method, a simplified block diagram of the human subject during the CR task is shown in Fig. 2.

This diagram shows that the sensory input (SI), generated by the simulator motion and visualization systems during the simulation trial, is processed by the human perceptual system (PS) into, among other signals, the PMI. Here, the subjects use their response system (RS) to translate the PMI into a continuous MIR. The latter is the CR data obtained during the experiment. When using the block diagram for the OR task mentioned above, the RS will yield a time-invariant rating, the offline MIR.

D. Validation

The most important properties of a measurement method are reliability and validity of the measurements [30]. The main advantage of this measurement method, in particular, is the possibility to correlate the various physical motion mismatches to the measured MIR. In the following paragraphs, the validation process and the applicability of the method in finding these correlations are further explained.

1) *Reliability*: To validate the novel measurement method, the results need to show that participants gave consistent ratings. The reliability analysis will determine within-subject consistency by comparing the three consecutive ratings of the same simulation trial by the same participant. The between-subject reliability analysis will be done by comparing all mean CRs across participants. The reliability estimate Cronbach's Alpha [31] is calculated in both cases. This parameter measures internal consistency and serves as a metric for the expected correlation between the ratings. A value of 0.7 or higher is generally considered to reflect acceptable reliability [32].

2) *Validity*: In addition to reliability, the CR method should also provide a valid measure of PMI. One way of analyzing this validity is to compare the continuous MIR to a generally accepted measure of PMI. The continuous MIR will therefore be compared to the offline MIR introduced in Section II-B. To pass the validity test, the continuous MIR should show a significant correlation with the offline MIR per segment. For this correlation calculation, the continuous MIR, containing measurements for each time step, should be reduced to one value per segment. The reduction method will be chosen based on the measurement results and, for example, could include the mean or the maximum MIR per segment. The resulting correlation coefficient between the offline and continuous MIR per segment will be tested for significance [33]. The *t*-test used to calculate the significance is shown in (1), where *N* is the amount of test

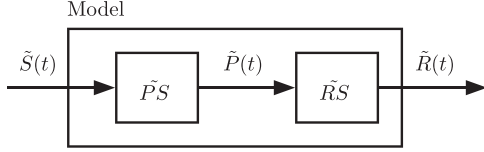


Fig. 3. Block diagram of a model mimicking a human subject during the CR of PMI in a simulator.

items and r the correlation coefficient

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}. \quad (1)$$

In this paper, it is assumed that the participants can use the visual motion cues presented in the simulator, together with their real-world driving experience, to derive the desired vehicle motion, while the simulator motion is represented by the perceived inertial motion cues generated by the motion platform. This means that motion mismatches as defined previously can be represented by the difference between the desired vehicle motion and obtained simulator motion. In the experiment, multiple MCAs and maneuvers are used to generate specific physical motion mismatches between the vehicle and the simulator in different motion channels. It is hypothesized that if the continuous MIR is indeed a valid measure of PMI, these motion mismatches can be clearly identified from the continuous MIR. This second validity check will be done via visual comparison of the mean continuous MIR over all participants and the induced physical motion mismatches.

3) *Applicability*: As mentioned in Section I, a major advantage of this measurement method is that its results can be used to obtain a deeper insight into the correlation between the SI generated by the simulator and the PMI. Subsequently, this correlation can be used to identify relevant short-duration motion mismatches and, eventually, minimize them. For this purpose, the block diagram of Fig. 2 is transformed into the model shown in Fig. 3.

The measurement method provides a continuous MIR $R(t)$, which can be compared to the modeled continuous MIR $\tilde{R}(t)$ to provide insight into the correlation between SI and PMI. As the model presented here is merely a first example of the applicability of the measurement method, the models $\tilde{P}S$ and $\tilde{R}S$ will be kept simple and in accordance with the previous literature.

In the field of motion simulation, SI $\tilde{S}(t)$ is often described as the specific force and rotational velocity in longitudinal (x), lateral (y), and vertical (z) direction [4], which, for simplicity, will also be done here. The perceptual system $\tilde{P}S$ translates these SIs into motion mismatches that together form the PMI $\tilde{P}(t)$. In the literature, these motion mismatches are often described as the absolute difference between vehicle and simulator SIs in individual degrees of freedom [34], [35], which will also be used here. The PMI is then calculated as the weighted sum of these motion mismatches, resulting in the perceptual system $\tilde{P}S$ shown in Fig. 4.

The modeled response system $\tilde{R}S$ should account for certain dynamics in the human rating process. In a previous research

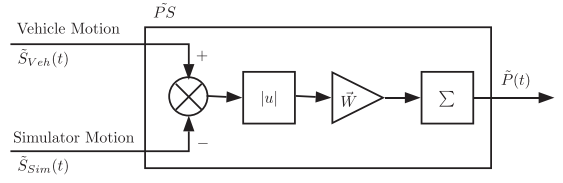


Fig. 4. Block diagram of the PS model, when rating PMI in a motion simulator.

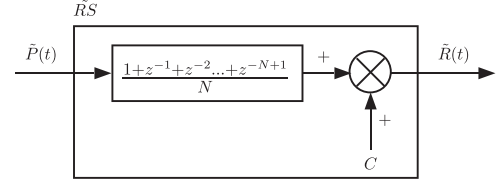


Fig. 5. Block diagram of the rating system model when continuously rating PMI in a motion simulator.

where a CR method was used for perceived positive emotion [36] and melody predictability [25], the CR was found to be a smoothed and delayed version of the expected signal. Hence, in this paper, the continuous MIR is expected to be a smoothed and delayed version of the PMI. Without available data to support an explicit model, the response system $\tilde{R}S$ is modeled as a simple moving average filter with a window length of N s. A constant C is added to account for the nonzero minimum mean rating, due to the spread between participants. The resulting rating system model $\tilde{R}S$ is shown in Fig. 5.

Assuming these representations of $\tilde{S}(t)$, $\tilde{P}S$, and $\tilde{R}S$, experimental CR data $R(t)$ can be used to, using linear least squares, fit the model parameters: the 6×1 motion mismatch weight vector \tilde{W} , the filter window length N , and the constant C . The resulting model weights \tilde{W} show the strength of the correlation between a specific motion mismatch and PMI. The perceptual system model $\tilde{P}S$ can be used to minimize the motion mismatches, by implementing it as a cost function in the optimization algorithms for MCAs. The weight parameters of the simple model described here $\tilde{P}S$ could for example be used to replace the tuned weights in cost functions for MCA optimization based on adaptive [34] or model predictive control [12] algorithms.

III. EXPERIMENT

An experiment was performed to investigate whether a CR method can be used to measure time-varying PMI. For this experiment, participants were exposed to a passive driving simulation in a motion-based simulator. During the simulation, different levels of motion incongruence were induced by varying the simulator MCA settings for different maneuvers.

A. Independent Variables

The independent variables in this experiment were maneuver (three levels) and MCA setting (three levels), which were all embedded in a simulation trial, resulting in nine different simulation segments. The following maneuvers were used in the simulation:

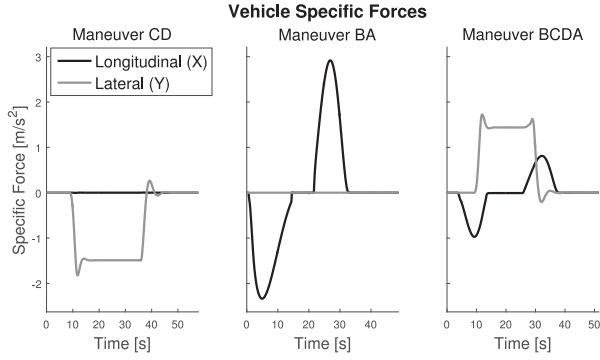


Fig. 6. Longitudinal and lateral vehicle specific force for the three maneuvers.

- 1) *CD*: Curve Driving at 70 km/h, on a curve with a 257-m radius and a 120° deflection angle;
- 2) *BA*: Braking from 70 km/h to full stop and again Accelerating to 70 km/h on a straight road;
- 3) *BCDA*: Braking from 70 to 50 km/h while entering the curve, Curve Driving at 50 km/h and Accelerating again to 70 km/h when exiting the curve, on a curve with a 131-m radius and a 120° deflection angle.

With these three maneuvers, the simulation consists of motion incongruence in different motion channels. As shown in Fig. 6, maneuvers CD and BA primarily affect the longitudinal (X) and lateral (Y) specific forces, respectively, whereas the BCDA maneuver combines both forces.

The MCAs were all classical washout filters [37], which map the vehicle specific force and rotational velocity vectors onto the simulator workspace. These algorithms make use of motion washout, returning the simulator to a neutral position with accelerations and rotations below human perception threshold, and tilt-coordination, tilting of the simulator cabin to simulate sustained acceleration. Tilt-rate limiting is applied to keep the rotation rate below human perception thresholds, for which values of ~ 3 deg/s are often used [38].

The washout filter parameters that serve as a basis for the three MCAs used here were tuned to reproduce the above-mentioned maneuver motions within simulator limits, while making maximum use of tilt-coordination and not applying scaling or tilt-rate limiting. To induce specific motion mismatches, only the scaling or the tilt-rate limiting parameters were adjusted, which resulted in the following three MCAs:

- 1) MCA_{Scal} : Scaling
 - a) Motion scaling (gain = 0.6), which leads to scaling and small rotational errors (< 4 deg/s);
- 2) MCA_{TRL} : Tilt-rate limiting
 - a) Rotation rate limiting to 1 deg/s, which leads to missing or false cues, and very small rotational errors;
- 3) MCA_{NL} : No limiting
 - a) Neither tilt-rate limiting nor scaling is applied, which leads to large rotational errors (< 8 deg/s).

In Fig. 7, it can be seen that the vehicle motions for each maneuver are shown together with the measured and commanded simulator motions resulting from the use of different MCAs.

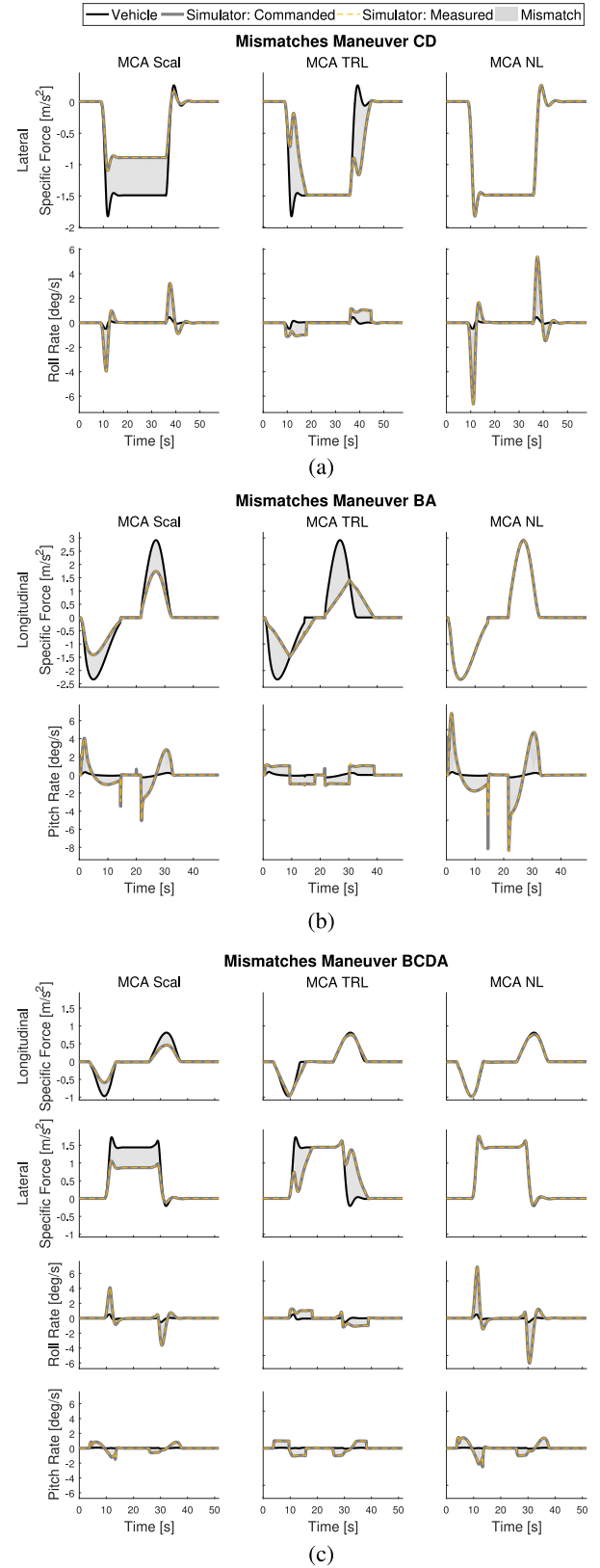


Fig. 7. Typical motion mismatches for three different MCAs during maneuvers (a) CD, (b) BA, and (c) BCDA. The figures show vehicle motion as calculated by CarSim (black line), the commanded simulator motion resulting from the different MCAs (gray line), the motion that was measured in the simulator (yellow line), and the mismatch between vehicle and commanded simulator motion (light gray area).

As the measured and commanded motions are very similar, the physical motion mismatches are hereby defined as the difference between vehicle and commanded simulator motion and are indicated with the light gray area in Fig. 7. The longitudinal specific force and pitch rate for maneuver CD, as well as the lateral specific force and roll rate for maneuver BA, are zero for both vehicle and simulator motion and are not shown in Fig. 7. A scaling error, visible in the specific force during the turn, the acceleration, and the deceleration motions for MCA_{Scal} , is caused by a constant gain between vehicle and simulator motion. A missing cue, visible in the specific force at the beginning of these motions for MCA_{TRL} , is here defined as a simulator motion that has a lower amplitude than the vehicle motion but, unlike the scaling error, the gain between vehicle and simulator motion is not constant over all frequencies. A false cue, visible in the specific force at the end of these same motions for MCA_{TRL} is similar to the missing cue, but here the variable motion gain is greater than one. False cues can also refer to simulator motion when no vehicle motion is present, such as the rotation errors visible in all rotational velocity plots in Fig. 7.

B. Dependent Variables

The dependent variables were the continuous MIR throughout the simulation trial, repeated three times, and the offline MIR for each of the nine simulation segments.

C. Apparatus

The experiment was performed in the CyberMotion Simulator at the Max Planck Institute for Biological Cybernetics. This dynamic simulator was developed to expand the limited workspace and dexterity of traditional hexapod-based simulators. It is an eight-degrees-of-freedom serial robot derived from an industrial robot manipulator (Kuka GmbH, Germany), where a six-axes industrial robot manipulator is mounted on a linear rail and equipped with a motorized cabin at the end effector. The cabin is equipped with two WUXGA (1920×1200 pixels) projectors (Eyevis, Germany) and interference filter stereo projection system (Infitec GmbH, Germany), which provide up to $160 \times 90^\circ$ field-of-view on the cabin inner side. The visuals and vehicle inertial motions were generated using the simulation software CarSim (Mechanical Simulation, USA). The rating interface, shown in Fig. 1, consisted of a rotary knob (SensDrive GmbH, Germany) to express the rating and a rating bar rendered on the dashboard of the virtual vehicle for visual feedback on the current rating.

D. Participants

In total 16 participants, one female, aged between 22 and 38 years partook in the experiment. Their levels of simulator experience ranged from no simulator experience (7), participated in simulator studies before (5), to motion cueing expert (4), and all had a valid driving license.

E. Procedure and Instructions

Participants were first trained to use the rating interface and familiarize themselves with the simulation via the rating interface

and congruence range trainings as described in Section II-B. For the congruence range training, two simulation trials were rated, after which the within-subject consistency was visually checked by the experimenter. If a low consistency was detected, a third training trial was given. After a short break, the CR measurement part started, where participants were asked to observe and continuously rate three simulation trials, each including all nine combinations of maneuvers and MCAs. After a second break, the OR measurement part was started, where the participants were asked to observe nine short simulation trials, containing only one segment each, and provide one OR after each trial using the rating interface. The same simulation segments were used throughout the experiment.

The simulation trial used for the congruence range training had a fixed segment order. The three trials used for the CR measurements had a different segment order and were never the same as the training trial. For the OR measurement part, each trial always consisted of the same initial acceleration and final deceleration and one of the nine segments, such that the simulation always had a natural start and ending. The order of these trials was randomized per participant. Per participant, the experiment lasted approximately 2 h.

F. Hypothesis

It is hypothesized that the continuous MIR will show sufficient consistency within- and between-subjects and that this rating will significantly correlate with the corresponding ORs. It is also hypothesized that the continuous MIR will show an increase in motion incongruence during the physical motion mismatches shown in Fig. 7. This in turn leads to the hypothesis that a simple model, which makes use of these motion mismatches, described in Section II-D3, can explain a significant portion of the motion incongruence measured with the CR.

IV. RESULTS

A. Reliability

As mentioned in Section II-D1, the reliability within- and between-subjects is determined by Cronbach's Alpha, using each time step sample as a separate measurement. The within-subjects reliability is calculated using the three simulation trial repetitions. The raw rating data for each of these three simulation trials from the first three participants are shown in Fig. 8. As each simulation trial has a different sequence of the same simulation segments, for comparison, the rating data for each trial have been reordered to fit the same base sequence.

The rating data in Fig. 8 show that participants rated the three trials consistently. The alpha for within-subject reliability had a median across all participants of 0.771 and an interquartile range between 0.727 and 0.897. Fig. 8 also shows that there is variability between participants. For example, during the vehicle acceleration and deceleration in maneuver BA for MCA_{NL} , which causes rotation rate mismatches, participant 1 rated the motion as being much more incongruent than participant 2. This difference, which is also visible in the other maneuvers for MCA_{NL} , could be explained by a difference in rotation rate perception threshold between these participants. Instead,

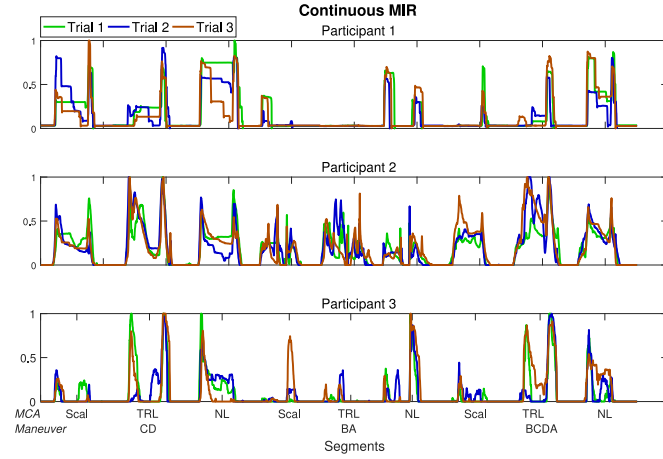


Fig. 8. Raw continuous MIR during three simulation trials.

participant 2 gave higher incongruence ratings than the other two participants during maneuver BA for MCA_{TRL} . This could possibly be explained by the ability of participant 2 to extract the vehicle motion more accurately from the visuals than the two other participants and therefore observing the incongruence better. Another explanation could be the preference of errors in one motion channel over another. Rotational errors might have had a stronger influence on the perceived incongruence of participant 1, whereas participant 2 was more focused on, or had a preference for, accurate linear acceleration. As humans have different motion sensitivities and thresholds, but also dissimilar higher level processes such as motion preferences, expectations, and experiences, the observed rating differences between participants are to be expected. However, an alpha of 0.855 for between-subject reliability indicates that, in general, participants did agree on the occurrence and magnitude of the PMI during the simulation. From these results it is concluded that the method provides sufficient reliable measurements.

B. Validity

As mentioned in Section II-D2, the validity analysis is done by comparing the offline and continuous MIR for each of the nine simulation segments, as well as comparing the continuous MIR to the physical motion mismatches shown in Fig. 7. For this analysis, Fig. 9 is used, which shows the mean offline and continuous MIR and their standard error during each segment.

For the correlation calculation between offline and continuous MIR, the latter needs to be reduced to one variable. Fig. 9 shows that, with the exceptions of the BA/ MCA_{TRL} and BCDA/ MCA_{NL} cases, the offline MIR can be accurately predicted by the maximum continuous MIR during that segment. To compare the offline and continuous MIR, the latter is therefore “summarized” as the mean across all participants of the maximum rating per participant per segment. As the set of maximum ratings per segment for each participant has a smaller variance and higher mean than the set of ORs per participant, for the correlation calculation, both sets are standardized to have zero mean and unit variance for each participant. Fig. 10 shows

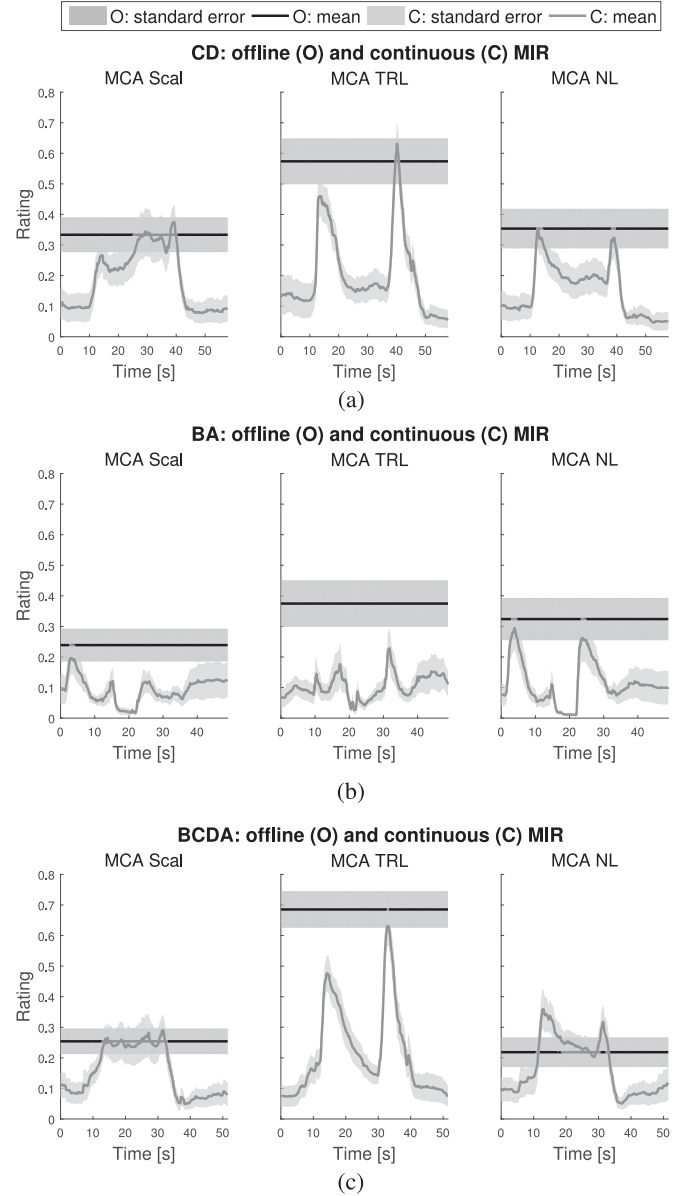


Fig. 9. Mean values for offline and continuous MIR and their standard error for each of the MCAs during maneuvers (a) CD, (b) BA, and (c) BCDA.

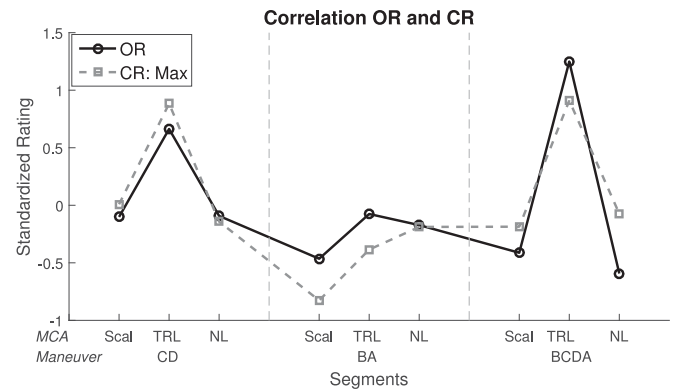


Fig. 10. Mean standardized rating per simulation segment for the two rating methods.

the means over all participants of 1) the standardized offline MIR and 2) the standardized maximum continuous MIR per segment.

The Pearson correlation coefficient between the standardized mean OR and maximum CR is $r = 0.86$ ($p < 0.01$). This indicates that there is a significant linear relationship between the offline and continuous MIR, and it can therefore be reasonably assumed that both methods measure the same PMI.

The time variations in the continuous MIR were hypothesized to correlate with the induced physical motion mismatches illustrated in Fig. 7. The continuous MIR for maneuver CD, shown in Fig. 9(a), clearly shows differences between the three MCAs. The false cue generated by MCA_{TRL} , starting at around 37 s, is clearly rated to induce the strongest motion incongruence. The second strongest PMI can be related to the missing cue starting at 11 s for the same MCA. The scaling error throughout the turn caused by MCA_{scal} is also clearly visible, but there is an unexpected increase in rating toward the end of the turn. More detailed analysis showed that this increase is visible in the ratings of seven of the sixteen participants. Finally, the peaks seen in the right graph in Fig. 9(a) can be related to the roll rate error at the beginning and end of the turn for MCA_{NL} .

The ratings for maneuver BA, shown in Fig. 9(b), are overall much lower than the ratings for maneuver CD. The main peaks in the continuous MIR for this maneuver are found when using MCA_{NL} , which causes large tilt rates at the onset of braking and accelerating. At around 20 s, the continuous MIR seems to approach zero, which can be attributed to the absence of any vehicle or simulator motion during the full stop (see Fig. 7). The continuous MIR for MCA_{TRL} does not approach zero as observed for the other two MCAs and, in fact, shows a small peak, which can be attributed to the missing high-frequency motion cue at the end of the full stop. This peak, however, was only visible in the ratings of six participants. Some participants reported verbally that they did not rate this missing cue, even though it did clearly increase the motion incongruence. They reported waiting for the cue to arrive but, when realizing it would not occur, felt it was too late to rate accordingly, which could explain the difference between offline and maximum continuous MIR for this maneuver.

Ratings for maneuver BCDA, shown in Fig. 9(c), are very similar to the ratings for maneuver CD, which reveals that similar physical motion mismatches indeed result in a very similar continuous MIR, indicating consistent rating behavior. Again the scaling error, missing/false cues and roll rate errors caused by MCA_{scal} , MCA_{TRL} , and MCA_{NL} , respectively, result in increased continuous MIR. The main difference with the continuous MIR is seen for MCA_{scal} : The increase in rating toward the end of the turn that was visible in the CD maneuver is not observed in the BCDA maneuver. Overall, the continuous MIR can be visually correlated to the physical motion mismatches during each simulation segment rather well, suggesting that the CR method can indeed be used to measure the PMI.

C. Applicability

To show how the results of the novel measurement method can be used, the simple model described in Section II-D3 is fit

TABLE I
ESTIMATED MODEL PARAMETERS

Motion Channel		\vec{W}	I
Specific Force	X	0.000	0
	Y	0.227	37
	Z	0.389	18
Rotational Velocity	Roll	0.047	17
	Pitch	0.003	2
	Yaw	0.012	26

Parameter	Value
C	0.087
N	3 [s]
δ_{avg}	1.45 [s]

to the measured continuous MIR. Such models may lead to a better insight into how the PMI results from physical motion mismatches. Furthermore, the fitted motion mismatch weights \vec{W} can be used in MCA optimization.

As explained in Section II-D3, we do not directly measure the true PMI $P(t)$, but rather the MIR $R(t)$. The latter being the output of the human response system RS, rather than the output of the human perception system PS. For this reason, not only a model of the perceptual system $\tilde{P}\tilde{S}$, but also a model of the human response system $\tilde{R}\tilde{S}$ needs to be fit. The model parameters motion mismatch weight \vec{W} , filter window length N , and rating constant C are estimated by fitting the modeled $\tilde{R}(t)$ to the mean measured continuous MIR over all participants.

When comparing the estimated mismatch weights \vec{W} , it should be taken into account that the corresponding mismatches did not have equal strength in the simulation trial. The motion mismatches in longitudinal specific force, for example, were mainly present during one third of the total simulation, i.e., during the BA maneuver. For this reason, an additional parameter, the influence factor I , is calculated to represent the percentage of $\tilde{P}(t)$ caused by mismatches in a specific motion channel

$$I_i = 100 \cdot \frac{\sum(|\tilde{S}_{veh_i}(t) - \tilde{S}_{sim_i}(t)|) * W_i}{\sum \tilde{P}(t)}. \quad (2)$$

Here, I_i is the influence factor, W_i the weight, \tilde{S}_{veh_i} the vehicle motion and \tilde{S}_{sim_i} the simulator motion for the i th motion channel. The resulting estimated parameters and the influence factor per motion channel are listed in Table I(a) and (b). Additionally the average filter delay δ_{avg} , resulting from the fitted filter window length N , is shown.

Fig. 11(a) shows the resulting modeled continuous MIR, as well as the mean and standard error of the measured continuous MIR averaged over all participants. Fig. 11(b) shows the PMI components per motion channel: The absolute difference between vehicle and simulator motion multiplied with the estimated weight vector.

To determine the goodness of fit of the modeled PMI, the coefficient of determination r^2 [33] is calculated with

$$r^2 = \frac{\sum(\bar{R} - \bar{R})^2}{\sum(R - \bar{R})^2} \quad (3)$$

where \bar{R} is the mean R over all time steps. r^2 was found to be 0.79, i.e., 79% of the variations in $R(t)$ can be accounted for by the model.

The lateral specific force and yaw rate motion channels had the highest influence factors. Fig. 11 shows that the motion mismatches in these channels only occurred during the curve

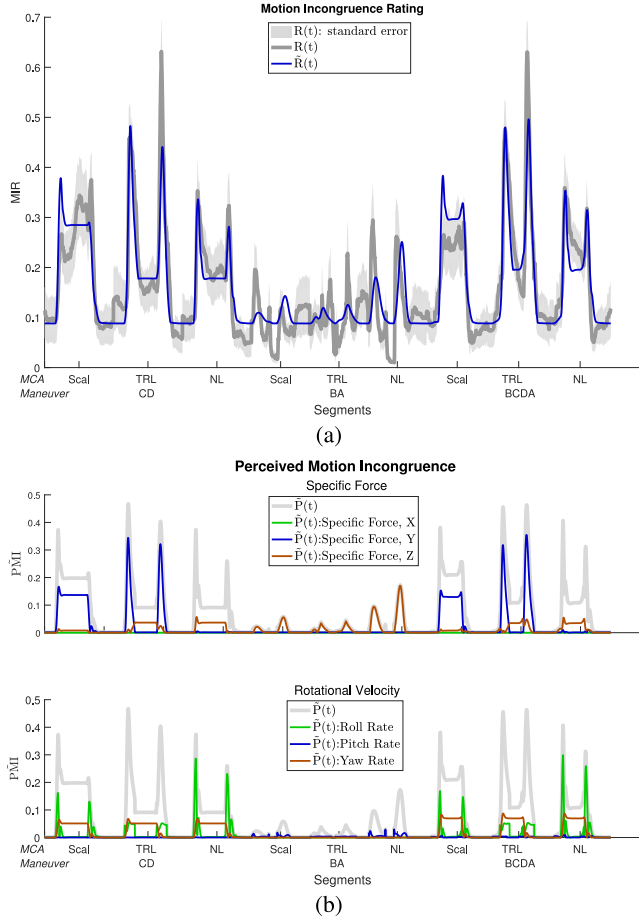


Fig. 11. (a) Measured continuous MIR $R(t)$, its standard error over all participants, and the modeled continuous MIR $\hat{R}(t)$. (b) Modeled PMI $\hat{P}(t)$ and its components.

driving maneuvers CD and BCDA. As expected, the specific force mismatches are the main contributor for the PMI when using MCA_{scal} and MCA_{TRL} , whereas the roll rate mismatches are the main contributor for the clear peaks in PMI when using MCA_{NL} . The PMI measured throughout the curve for all MCAs is best modeled with the yaw rate mismatches.

During the BA maneuver, $\hat{P}(t)$ is mainly based on the motion mismatches in the vertical specific force, caused by tilt-coordination. It is surprising that the motion mismatches in the longitudinal specific force did not influence $\hat{P}(t)$ at all. The pitch rotation rate had a very small influence on $\hat{P}(t)$.

The perceptual system model \tilde{PS} together with the estimated weights \tilde{W} from Table I(a) could now be used to minimize the PMI for maneuvers and motion mismatches similar to those used in this experiment, by implementing it as a cost function in an MCA optimization algorithm. The estimated weights would then replace tuned weights normally used in the cost function. As the range of motion incongruence used in this experiment is relatively small, compared to those that can possibly be present during a vehicle simulation, the perception model \tilde{PS} should be further improved using additional experimental data, before it can be used for a larger range of maneuvers and motion mismatches.

V. DISCUSSION

A. Summary

The experiment analysis described in this paper shows that the proposed CR method can indeed be used to obtain a valid and reliable measure of time-varying PMI during vehicle motion simulation. Within- and between-subject reliability of the continuous MIR was shown to be sufficiently high to assume reliable measures. The validity of the measurement was investigated by comparing the continuous MIR first to an offline MIR generally used to measure overall PMI and second to the physical motion mismatches in six motion channels over time. A significant correlation between the offline MIR and the maximum continuous MIR was found, indicating that the CR method indeed results in a measure of the PMI similar to that measured with the OR method. Also because the different physical motion mismatches can be identified from the continuous MIR, it is reasonable to assume that the continuous MIR is indeed a measure for PMI. Finally, the applicability of the results from this method was investigated by fitting a simple model, describing the formation of PMI and the resulting continuous MIR from the SI in a motion simulator, to the measured CR. This simple model could already explain a large part of the measured continuous MIR. Next to giving more insight into the importance of certain motion mismatches on the formation of PMI; these estimated weights for motion mismatches in six motion channels, specific force and rotational velocity, can be used in the cost functions for MCA optimization.

B. Specific Findings

Even though the current experiment was set up to validate the measurement method, it also provided some interesting results on motion cueing that require further discussion. The correlation between the OR per segment and the CR seemed to depend on the maximum, rather than the mean CR during a maneuver. This finding is in accordance with the earlier findings on the relation between overall and a CR of video quality [39], where it is found that the OR can be best predicted from the CR by the peak impairment. This finding should also be taken into account when designing a cost function for MCA optimization, where currently the overall motion incongruence is assumed to be a summation of the motion incongruence over all time steps of a certain maneuver [40] or the prediction horizon [41].

The correlation between the CRs and motion mismatches in different motion channels showed that the false cues during the curve driving maneuvers were rated with a higher motion incongruence than the scaling and missing cues during these maneuvers. This is in accordance with general knowledge on false cues [4], but has now been measured directly for the first time.

The missing cue in longitudinal specific force at the end of the full stop in the "Braking, Accelerating" maneuver, unlike the missing cues during the "Curve Driving" maneuver, was not clearly rated as highly incongruent with the CR. The OR for this segment, however, did show high overall incongruence. The missing cue makes one feel like the car never came to a full stop,

while the visuals do not show any vehicle motion anymore. The timing of this cue is difficult to deduce from the visual motion, but, from experience, participants do expect the cue to appear at the end of the full stop. Because of this, some participants had reported to be too late in rating this PMI during the CR, but said they did take it into account during the OR. This could explain the difference between offline and CR, which is an interesting topic for further research. Depending on the application of the measurement method, it can be useful to clearly instruct the participants to rate any motion incongruence, even if their rating is delayed. It is often more important to obtain the incongruence rating than it is to avoid time delays, as the latter can be detected and removed during the data analysis.

The “Braking, Accelerating” maneuver, was clearly rated to be less incongruent for all MCAs than the curve driving maneuvers, even though the objective motion mismatch magnitude, shown in Fig. 7, was similar. An explanation for this can be that participants are less capable of perceiving longitudinal vehicle acceleration, derived from the changes in velocity observed in the simulator visuals, than they are at extracting vehicle yaw rate from these visuals during curve driving due to differences in the optic flow for these two degrees-of-freedom, as explained in [42]. Less accuracy in the visually perceived vehicle motion results in a larger range of simulator motions still being perceived as congruent.

A simple model, mapping SI to a MIR can already produce a good fit, explaining 79% of measured mean CRs. The fit was best for the curve driving maneuvers, where the model shows that the rating was most likely caused by the lateral specific force mismatches.

The peak in motion incongruence due to the false cues for the curve driving maneuvers, when using an MCA that includes tilt-rate limiting, is modeled to be much lower than the peak that was measured with the CR, whereas the peaks related to the missing cues are a much better fit. Both these peaks are caused by motion mismatches in the same motion channel, i.e., lateral specific force. As mentioned before, false cues are in general perceived as more incongruent than missing cues, but with the current implementation of the model this difference cannot be emphasized. It is therefore advised for future work that the motion mismatches defined as false cues are assigned a different weight than mismatches defined as missing cues. The fit for the “Braking, Accelerating” maneuver is much less accurate than for the curve driving maneuvers. A surprising finding is that the longitudinal specific force and the pitch rate are modeled to have, respectively, no and very little impact on the PMI. Instead, the motion mismatch in vertical specific force is modeled as the main influence on PMI during this maneuver. The latter mismatch exists due to tilt-coordination used in washout filters, where linear acceleration is simulated via the gravitational vector by tilting the simulator cabin. The motion mismatch in vertical specific force is thus related to the pitch angle. This implies that instead of responding to the perceived pitch rate, participants might have responded to the resulting pitch angle. This is consistent with the results presented in [43], where it was concluded that subjects relied strongly on mismatches in

attitude to determine the goodness of the motion cueing, at least during the tested longitudinal acceleration maneuver.

The estimated weight of zero for longitudinal specific force can be related to the aforementioned less accurate perception of this motion from simulator visuals, leading to a large range of simulator motions still being perceived to be congruent with the visual motion. In [44], it is concluded that simulator jerk has a large influence on perceived motion. In future work, it is therefore advised to investigate if additional SIs, such as rotational angle and linear jerk, would result in a better fit for the “Braking, Accelerating” maneuver.

The human RS during the rating task was modeled as a moving average filter and an added constant offset. The estimated window length for the filter resulted in an average delay of 1.45 s. This delay seems reasonable as it is in the same range as delays found in previous research where a CR method was applied: Winkler and Campos [45] report delays between 1.5 and 2 s, Eerola *et al.* [25] report delays between 0.5 and 0.7 s, and Buchinger *et al.* [46] report delays between 0.9 and 1.2 s. The constant offset should have accounted for the nonzero minimum mean rating, due to the spread between participants, at points in the simulation where instead a minimum rating of zero was expected. A minimum rating is expected, for example, while the car is stopped and no simulator motion is present. The nonzero minimum mean CRs found at these points in the simulation, however, are much lower than the estimated constant offset. This could be an indication that the constant offset value was not estimated correctly, or that the constant offset fulfills a different role in the model than the role that was intended. Some participants verbally reported that they rated the absence of road rumble. It is possible that such ratings were accounted for in the model by an inflated value of the constant offset. In future work, it is therefore advised to include the absence of road rumble as one of the motion mismatches in the perception system.

C. Method Considerations

This section gives a short overview of several aspects of the rating method that should be taken in consideration when using it to measure PMI.

- 1) *Time resolution*: The major advantage of this method is that PMI is measured continuously over time, resulting in much more information than the currently existing methods that only provide overall measurements for a certain simulation trial. Even though the rating is continuous, it cannot be assumed to be instantaneous. Due to processing delays and filtering, which can also differ somewhat between participants, the time resolution of the mean MIR is limited.
- 2) *Passive driving*: The main drawback of the presented method is that it involves a passive driving simulation, where participants are not asked to control the vehicle. In [47], it is shown that, at least in some cases, perceptual thresholds differ between passive and active driving tasks, which can be an indication that PMI will also differ. The perceptual thresholds, however, seem to increase, which

could lead to less sensitivity to motion incongruence during active driving. If this is the case, measuring PMI during a passive rather than an active driving task might actually lead to a more sensitive measurement. Another drawback of passive driving simulations is that only the incongruence between visual and inertial motion stimuli can be measured. Perceived visual-inertial motion incongruence, however, is not the only aspect of motion quality in a simulator. Motion quality depends on both the congruence between all motion stimuli, including stimuli such as proprioceptive feedback from control devices during active driving, and on how well these stimuli combined represent the actual vehicle motion. For motions such as the high-frequency vehicle motion at the end of a full stop maneuver, which is not easily derived from visual information, incongruence between the proprioceptive stimuli during active driving and the inertial motion might be a more important measure. This, however, cannot be measured with passive driving simulations. In [27], an experiment is described where CRs are used to measure strain during driving in both active and passive driving simulations. Here, CRs were taken during a simulation while performing an active driving task, followed by CR while passively observing a repetition of this simulation. To include the effect of performing an active driving task on PMI, it would be interesting to use a similar experiment setup in future work.

- 3) *Direct measurement*: An advantage of the method is that it provides direct measurements of perceived visual-inertial motion incongruence rather than indirect measurements such as control behavior or performance. Critical aspects of motion cueing therefore can be more easily identified.
- 4) *Memory workload*: An advantage of the CR method as compared to OR is that the memory workload is reduced, as participants do not need to evaluate an entire trial and compare it to another one. Instead the PMI is only compared to the one instant of maximum incongruence that is unchanged throughout the experiment. This decreased memory workload also allows for rating of longer simulation trials, as compared to OR. Longer rating trials in turn might help participants reach a higher level of immersion into the simulation.
- 5) *Measurement scale*: Another advantage of the method is that it allows for measurements on an interval scale, i.e., containing information about order as well as having equal intervals, rather than an ordinal scale, i.e., only containing information about order, such as the often used the Cooper–Harper scale [48] or the paired comparison method [49]. However, unlike the Cooper–Harper scale, the scale in this study is derived from a specific experiment, which has the drawback that comparison between experiments is more difficult.
- 6) *Participant engagement*: A negative effect of passive driving simulations is that participants can lose concentration due to the lack of activity. In the CR method this is much reduced by requesting participants to actively rate throughout the simulation.

- 7) *Realistic simulation*: The experiment described in this paper was deliberately set up to create very clear and specific cueing errors, such that the method could be validated. For future work, it would be interesting to perform an experiment using a more realistic environment by using a dedicated driving simulator and realistic visuals and maneuvers.

VI. CONCLUSION

This paper describes a first experiment using a CR method to measure time-varying PMI in a motion-based simulator. Results show that participants with different backgrounds and expertise in motion cueing and motion simulation are able to continuously rate PMI during passive driving simulations in a consistent manner. The correlation between retrospective offline and CR methods suggests that both methods indeed measure the same underlying variable, i.e., the PMI. This result is strengthened by the similarities between the CR and the presented physical mismatches between vehicle and simulator motion. The CR could therefore be used to determine the relative importance of short-duration motion mismatches such as scaling errors, missing, and false cues. A simple model, mapping a selected set of SIs to the MIR, was fitted to the measured CR data. The estimated model parameters showed the relative importance of each of the selected SIs on the formation of PMI. Using this novel measurement method more complex models can be designed, which can significantly increase the knowledge on PMI and that can also be used to further improve simulator motion cueing.

REFERENCES

- [1] D. H. Weir, "Application of a driving simulator to the development of in-vehicle human-machine-interfaces," *IATSS Research*, vol. 34, no. 1, pp. 16–21, Jul. 2010.
- [2] D. J. Allerton, "The impact of flight simulation in aerospace," *Aeronaut. J.*, vol. 114, no. 1162, pp. 747–756, Dec. 2010.
- [3] N. J. I. Garrett and M. C. Best, "Driving simulator motion cueing algorithms—A survey of the state of the art," in *Proc. Int. Symp. Adv. Veh. Control*. Loughborough, U.K.: Loughborough University/Society of Automotive Engineers of Japan, Aug. 2010, pp. 183–188.
- [4] P. R. Grant and L. D. Reid, "Motion washout filter tuning: Rules and requirements," *J. Aircr.*, vol. 34, no. 2, pp. 145–151, Apr. 1997.
- [5] Advisory Group for Aerospace Research and Development, "Fidelity of simulation for pilot training," North Atlantic Treaty Organization, Neuilly sur Seine, France, Tech. Rep. AGARD-AR-159, 1980.
- [6] Y. J. Brown, F. M. Cardullo, and J. B. Sinacori, "Need-Based evaluation of simulator force and motion cueing devices," in *Proc. AIAA Flight Simul. Technol. Conf. Exhib.* Boston, MA, USA: Amer. Inst. Aeronaut. Astronaut., Aug. 1989, pp. 78–85.
- [7] F. A. M. van der Steen, "Self-motion perception," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, 1998.
- [8] A. R. Valente Pais, M. M. van Paassen, M. Mulder, and M. Wentick, "Perception coherence zones in flight simulation," *J. Aircr.*, vol. 47, no. 6, pp. 2039–2048, Nov. 2010.
- [9] P. Jonik, A. R. Valente Pais, M. M. van Paassen, and M. Mulder, "Phase coherence zones in flight simulation," in *Proc. AIAA Model. Simul. Technol. Conf.*, vol. 47, Reston, VA, USA: Amer. Inst. Aeronaut. Astronaut., Aug. 2011, pp. 2039–2048.
- [10] A. R. Valente Pais, M. Wentink, M. M. van Paassen, and M. Mulder, "Comparison of three motion cueing algorithms for curve driving in an urban environment," *Presence, Teleoperators Virtual Environ.*, vol. 18, no. 3, pp. 200–221, Jun. 2009.
- [11] S. E. Reardon, S. D. Beard, and B. L. Aponso, "Effects of motion filter parameters on simulation fidelity ratings," in *Proc. AHS 70th Annual Forum*, Montreal, Canada, May 2014, pp. 2112–2123.

- [12] P. Pretto, J. Venrooij, A. Nesti, and H. H. Bühlhoff, "Perception-Based motion cueing: A cybernetics approach to motion simulation," in *Recent Progress in Brain and Cognitive Eng.*, vol. 5, S.-W. Lee, H. H. Bühlhoff, and K.-R. Müller, Eds. Dordrecht, The Netherlands: Springer, 2015, ch. 9, pp. 131–152.
- [13] R. J. V. Bertin, C. Collet, S. Espié, and W. Graf, "Objective measurement of simulator sickness and the role of visual-vestibular conflict situations," in *Proc. Driving Simul. Conf., North Amer.*, Orlando, FL, USA, Nov. 2005, pp. 280–293.
- [14] D. P. Jang, I. Y. Kim, S. W. Nam, B. K. Wiederhold, M. D. Wiederhold, and S. I. Kim, "Analysis of physiological response to two virtual environments: Driving and flying simulation," *CyberPsychol. Behavior*, vol. 5, no. 1, pp. 11–18, Feb. 2002.
- [15] H. J. Damveld, M. Wentink, P. M. van Leeuwen, and R. Happee, "Effects of motion cueing on curve driving," in *Proc. Driving Simul. Conf.*, Paris, France, Sep. 2012, pp. 1–9.
- [16] D. M. Pool, P. M. T. Zaal, M. M. van Paassen, and M. Mulder, "Effects of heave washout settings in aircraft pitch disturbance rejection," *J. Guid., Control, Dyn.*, vol. 33, no. 1, pp. 29–41, Feb. 2010.
- [17] D. M. Pool, P. M. T. Zaal, H. J. Damveld, M. M. van Paassen, and M. Mulder, "Evaluating simulator motion fidelity using in-flight and simulator measurements of roll tracking behavior," in *Proc. AIAA Model. Simul. Technol. Conf.*, Reston, VA, USA: Amer. Inst. Aeronaut. Astronaut., Aug. 2012, pp. 1–25.
- [18] M. Bruenger-Koch, "Motion parameter tuning and evaluation for the DLR automotive simulator," in *Proc. Driving Simul. Conf., North Amer.*, Orlando, FL, USA, Nov. 2005, pp. 262–270.
- [19] P. M. T. Zaal and D. M. Pool, "Multimodal pilot behavior in multi-axis tracking tasks with time-varying motion cueing gains," in *Proc. AIAA Model. Simul. Technol. Conf. Exhib.*, National Harbor, MD, USA: Amer. Inst. Aeronaut. Astronaut., Jan. 2014, pp. 174–190.
- [20] D. Cleij, J. Venrooij, P. Pretto, D. M. Pool, M. Mulder, and H. H. Bühlhoff, "Continuous rating of perceived visual-inertial motion incoherence during driving simulation," in *Proc. Driving Simul. Conf.*, Tübingen, Germany, Sep. 2015, pp. 191–198.
- [21] ITU, "Methodology for the subjective assessment of the quality of television pictures," Int. Telecommun. Union, Geneva, Switzerland, Tech. Rep., ITU-R Recommendation BT.500-13, 2012.
- [22] S. S. Stevens, "The direct estimation of sensory magnitudes: Loudness," *Amer. J. Psychol.*, vol. 69, no. 1, pp. 1–25, Mar. 1956.
- [23] M. Lambooi, W. A. IJsselstein, and I. Heynderickx, "Visual discomfort of 3D TV: Assessment methods and modeling," *Displays*, vol. 32, no. 4, pp. 209–218, Jun. 2011.
- [24] J. Freeman, S. E. Avons, D. E. Pearson, and W. A. IJsselstein, "Effects of sensory information and prior experience on direct subjective ratings of presence," *Presence, Teleoperators Virtual Environ.*, vol. 8, no. 1, pp. 1–13, Feb. 1999.
- [25] T. Eerola, P. Toivainen, and C. L. Krumhansl, "Real-Time prediction of melodies: Continuous predictability judgments and dynamic models," in *Proc. Int. Conf. Music Perception Cogn.*, Sydney, NSW, Australia, Jul. 2002, pp. 473–476.
- [26] R. Cowie, A. Camurri, and D. Glowinski, "SIEMPRE Consortium deliverable D4.1—Results from the first series of experiments and first evaluation report," Social Interaction and Entrainment using Music PeRformance Experimentation (SIEMPRE) Project no. 250026 Consortium, Tech. Rep., 2011. [Online]. Available: <http://www.infomus.org/siempre/publications.php>
- [27] C. Schiebl, "Subjective strain estimation depending on driving manoeuvres and traffic situation," *IET Intell. Transport Syst.*, vol. 2, no. 4, pp. 258–265, Dec. 2008.
- [28] J. M. Girard, M. Wilczyk, Y. Barloy, P. Simon, and J. C. Popieul, "Towards an on-line assessment of subjective driver workload," in *Proc. Driving Simul. Conf., North Amer.*, Orlando, FL, USA, Nov. 2005, pp. 382–391.
- [29] T. Liu, G. Cash, N. Narvekar, and J. Bloom, "Continuous mobile video subjective quality assessment using gaming steering wheel," in *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron.*, Scottsdale, AZ, USA, Jan. 2012, pp. 1–6. [Online]. Available: <http://events.engineering.asu.edu/vpqr/vpqr12/>
- [30] E. G. Carmines and R. A. Zeller, *Reliability and Validity Assessment*, 9th ed., J. L. Sullivan and R. G. Niemi, Eds. Beverly Hills, CA, USA: Sage Publications, 1987.
- [31] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, Sep. 1951.
- [32] J. F. Hair, Jr., W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 7th ed. Upper Saddle River, NJ, USA: Pearson Education, 2009.
- [33] D. C. Howell, *Statistical Methods for Psychology*, 8th ed., T. Matray and P. Leeds, Eds. Belmont, CA, USA: Cengage Learning, 2012.
- [34] A. Naseri and P. R. Grant, "An improved adaptive motion drive algorithm," in *Proc. AIAA Model. Simul. Technol. Conf. Exhib.*, San Francisco, CA: American Institute of Aeronautics and Astronautics, Aug. 2005.
- [35] Y.-H. Chang, C.-S. Liao, and W.-H. Chieng, "Optimal motion cueing for 5-DOF motion simulations via a 3-DOF motion simulator," *Control Eng. Pract.*, vol. 17, no. 1, pp. 170–184, Jul. 2009.
- [36] D. S. Messinger, T. D. Cassel, S. I. Acosta, Z. Ambadar, and J. F. Cohn, "Infant smiling dynamics and perceived positive emotion," *J. Nonverbal Behavior*, vol. 32, no. 3, pp. 133–155, Sep. 2008.
- [37] L. D. Reid and M. A. Nahon, "Flight simulation motion-base drive algorithms: Part 1—Developing and testing the equations," Inst. Aerosp. Studies, Univ. Toronto, Toronto, ON, Canada, Tech. Rep. UTIAS 296, 1985.
- [38] E. L. Groen, M. Wentink, A. R. Valente Pais, M. Mulder, and M. M. van Paassen, "Motion perception thresholds in flight simulation," in *Proc. AIAA Model. Simul. Technol. Conf. Exhibit*, Keystone, CO, USA, Aug. 2006, pp. 1–11.
- [39] D. S. Hands and S. E. Avons, "Recency and duration neglect in subjective assessment of television picture quality," *Appl. Cogn. Psychol.*, vol. 15, no. 6, pp. 639–657, Nov. 2001.
- [40] J. Venrooij et al., "Perception-Based motion cueing: Validation in driving simulation," in *Proc. Driving Simul. Conf.*, Tübingen, Germany, Sep. 2015, pp. 153–161.
- [41] M. Dagdelen, G. Reymond, A. Kemeny, M. Bordier, and N. Maizi, "Model-Based predictive motion cueing strategy for vehicle driving simulators," *Control Eng. Pract.*, vol. 17, no. 9, pp. 995–1003, Apr. 2009.
- [42] B. J. Correia Grácio, J. E. Bos, M. M. van Paassen, and M. Mulder, "Perceptual scaling of visual and inertial cues," *Exp. Brain Res.*, vol. 232, no. 2, pp. 637–646, Nov. 2014.
- [43] J. E. Bos, W. Bles, and R. J. A. W. Hosman, "Modeling human spatial orientation and motion perception," in *Proc. AIAA Model. Simul. Technol. Conf. Exhibit*, Montreal, QC, Canada: Amer. Inst. Aeronaut. Astronaut., Aug. 2001, pp. 1–11.
- [44] P. R. Grant and B. Haycock, "Effect of jerk and acceleration on the perception of motion strength," *J. Aircraft*, vol. 45, no. 4, Jul. 2008, pp. 1190–1197.
- [45] S. Winkler and R. Campos, "Video quality evaluation for internet streaming applications," *Electron. Imag.*, vol. 5007, pp. 104–115, Jun. 2003.
- [46] S. Buchinger et al., "Evaluating feedback devices for time-continuous mobile multimedia quality assessment," *Signal Process., Image Commun.*, vol. 29, no. 9, pp. 921–934, Jul. 2014.
- [47] A. Nesti, S. Nooij, M. Losert, H. H. Bühlhoff, and P. Pretto, "Roll rate perceptual thresholds in active and passive curve driving simulation," *Simul.*, vol. 92, no. 5, pp. 417–426, May 2016.
- [48] G. E. Cooper and R. P. Harper, Jr., "The use of pilot ratings in the evaluation of aircraft handling qualities," Nat. Aeronaut. Space Administration, Washington, DC, USA, Tech. Rep. NASA TN D-5153, Apr. 1969.
- [49] L. L. Thurstone, "A law of comparative judgment," *Psychol. Rev.*, vol. 34, no. 4, pp. 273–286, 1927.



Diane Cleij received the M.Sc. degree in aerospace engineering from Delft University of Technology (TU Delft), Delft, The Netherlands, in 2011. She is currently working toward the Ph.D. degree in predicting and minimizing perceived motion incongruence in vehicle motion simulations the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and the Faculty of Aerospace Engineering, TU Delft.

Her research interests include cybernetics, modeling of human perception and control behavior and human-machine interface design.



Joost Venrooij (S'10–M'14) received the M.Sc. and Ph.D. degrees (both *cum laude*) in aerospace engineering from Delft University of Technology, Delft, The Netherlands, in 2009 and 2014, respectively.

After receiving the Ph.D. degree, he was a Project Leader in the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, where his research interests included motion simulation, biodynamics, and cybernetics. He is currently a Research Specialist in BMW Group, Munich, Germany.



Paolo Pretto received the M.Sc. degree in psychology from University of Padova, Italy, in 2002 and the binational Ph.D. degree in perception and psychophysics from University of Padova, Italy, in 2008.

He is currently the leader of the Motion Perception and Simulation Research Group, Max Planck Institute for Biological Cybernetics, Tübingen, Germany. His research interests include multisensory self-motion perception, VR and simulation, and user experience.



Daan M. Pool (M'09) received the M.Sc. degree in aerospace engineering and Ph.D. degrees (*cum laude*) in development of an objective method for optimization of flight simulator motion cueing fidelity based on measurements of pilot control behavior from Delft University of Technology (TU Delft), Delft, The Netherlands, in 2007 and 2012, respectively.

He is currently an Assistant Professor in the Section Control and Simulation, Aerospace Engineering, TU Delft. His research interests include cy-

bernetics, manual vehicle control, simulator-based training, and mathematical modeling, identification, and optimization techniques.



Max Mulder (M'14) 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 Section Control and Simulation, 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.



Heinrich H. Bühlhoff (M'95) received the Ph.D. degree in natural sciences from Eberhard Karls University, Tübingen, Germany, in 1980.

From 1980 to 1985, he was a Research Scientist in the Max Planck Institute (MPI) for Biological Cybernetics, Tübingen, Germany, and from 1985 to 1988, he was a Visiting Scientist in the Massachusetts Institute of Technology, Cambridge, MA, USA. Between 1988 and 1993, he was an Assistant Professor, an Associate Professor, and a Full Professor of cognitive science at Brown University, Providence, RI, USA.

In 1993, he was the Director of the Department of Human Perception, Cognition, and Action at the MPI for Biological Cybernetics, and a Scientific Member of the Max Planck Society. Since 1996, he has been an Honorary Professor at Eberhard Karls University, and an Adjunct Professor at Korea University, Seoul, South Korea, since 2009. His research interests include object recognition and categorization, perception, and action in virtual environments, and human–robot interaction.