Driving Simulator Experiment Stakeholder Perspectives on Motion Cueing Algorithm Quality

Kolff, M.J.C.; Venrooij, Joost ; Pool, D.M.; Mulder, Max

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the 22nd driving Simulation & Virtual Reality Confrence

Citation (APA)
Kolff, M. J. C., Venrooij, J., Pool, D. M., & Mulder, M. (2023). Driving Simulator Experiment Stakeholder Perspectives on Motion Cueing Algorithm Quality. In *Proceedings of the 22nd driving Simulation & Virtual Reality Confrence* (pp. 99-106)

Driving Simulator Experiment Stakeholder Perspectives on Motion Cueing Algorithm Quality

Maurice Kolff[1,2], Joost Venrooij[1], Daan M. Pool[2], Max Mulder[2]

(1) BMW Group, Research, New Technologies, Innovations, 80788 Munich, e-mail: {Maurice.Kolff, Joost.Venrooij}@bmw.de
(2) Delft University of Technology, Faculty of Aerospace Engineering, section of Control & Simulation, 2629 HS Delft, e-mail: {m.j.c.kolff, d.m.pool, m.mulder}@tudelft.nl

Abstract - *In driving simulation, the choice of a simulator, motion cueing algorithm, and associated set of tuning parameters for an experiment is typically made with an exclusive focus on the quality of the motion. In practice, many other metrics could affect this choice as well, such as tuning complexity, algorithm stability, or the financial costs of the simulation. Arguably, the complete motion cueing algorithm quality is thus more than the quality of the motion alone. This paper presents results of a survey which attempted to identify the most important metrics from the perspective of the main experiment stakeholders. Four stakeholder groups in typical driving simulator experiments are defined: The experimenters, motion cueing engineers, operators, and participants. All groups received the same survey, asking them to indicate how important various metrics are for them. Results show that, next to the quality of the motion, experimenters and participants are generally interested in reducing simulator sickness. The motion cueing engineers rank tuning effort and tuning complexity as most important metrics. Operators prefer an easy to use and overall stable motion cueing. A typical BMW experiment is discussed as example, which shows that the choice for a simulator and motion cueing algorithm can indeed differ when including these metrics in a trade-off, compared to when only motion quality is considered. The presented methods allow for a better, multi-faceted selection of the simulator, motion cueing algorithm, and associated tuning parameters, improving future driving simulation experiments.*

Keywords: *Motion cueing, quality comparison, objective assessment, stakeholder survey.*

1. Introduction

In motion-base driving simulators, a Motion Cueing Algorithm (MCA) is required to reproduce the motion of the simulated vehicle, while remaining within the limited workspace of the motion system. Typically, the choice of an MCA and its tuning parameters is made to provide the highest possible perceptual or behavioural quality (Romano, et al., 2019). In practice, operational qualities and costs can also play a role in this choice, such as how easy it is to tune an MCA on a certain simulator, or its induced energy consumption. Depending on the experiment research question, such metrics can even be of equal or higher importance than the attained level of motion quality. For example, a lower-quality MCA might be perfectly acceptable for an experiment, and additional improvements to cueing quality may not be worth the additional time and money investments. Arguably, the overall 'quality' of an MCA thus goes beyond consideration of motion cueing quality alone.

In early literature, Nahon and Reid (1990) already stressed the importance of other metrics next to the motion cueing quality: the number of tuning parameters of the MCA (as a measure of tuning effort), its number of differential equations (as a measure of its required computational effort), and the 'transparency' of the MCA (how clear it is how certain parameter settings affect the MCA output). In this light, Kolff, et al. (2020) included workspace management (how well the MCA is able to exploit the available workspace) and energy consumption next to objective metrics of perceptual motion cueing quality.

Although these works provide example assessments based on the view of the MCA designer, an evaluation of what metrics are actually of importance to *all those* involved in an experiment (i.e., the stakeholders) does not yet exist. Overall MCA quality should ultimately be defined as a function of all the wishes and requirements of these stakeholders, which may be conflicting. For example, an experimenter may want the best possible perceptual quality during the simulation, whereas the motion cueing engineer might prefer as little tuning parameters as possible, at the cost of a lower overall motion cueing quality.

This paper aims to provide a more comprehensive view on "MCA quality" through two contributions. First, four different stakeholders typically involved in motion cueing experiments are defined: the experimenter, the motion cueing engineer, the operator, and the participant. Second, to identify what metrics are seen as important, a single survey containing 35 questions is performed amongst all four stakeholder groups, in which they indicate how important various metrics of MCA quality are for them. The results of a future BMW driving simulator experiment, which has not been performed yet, are highlighted. Based on these results, an example evaluation is performed. Ultimately, the goal is not to develop a trade-

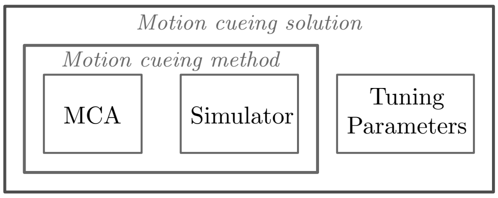


Figure 1: Definitions of the motion cueing method and solution, relating the MCA, simulator, and tuning parameters.

off methodology by weighing these various qualities and costs, as relative weighting between metrics would be highly subjective. Rather, the goal is to provide more insight in which metrics are of importance for an experiment and how these compare between various configurations.

This paper is structured as follows. Section 2 introduces various metrics and introduces the survey. Section 3 presents the obtained survey results, followed by the future experiment evaluation in Section 4. A discussion is presented in Section 5 and the paper is concluded in Section 6.

2. Methods

2.1. Selection Process

The process of selecting the simulator, MCA, and tuning parameters generally consists of various steps. We define the simulator and MCA choice together as the *motion cueing method*, see Figure 1, as they define the method of the motion cueing in which the later tuning is performed. In practice, it is typically the simulator choice that has to be made first, as simulators might have to be reserved well beforehand. For the MCA choice, it must be determined which type of algorithm is used. Different MCA options exist, each with their own (dis)advantages, such as the Classical Washout Algorithm (CWA) (Conrad, Douvillier, and Schmidt, 1973; Reid and Nahon, 1985) or the Model-Predictive Control (MPC) algorithm (Dagdelen, et al., 2009). Note that the quality of the MCA can be strongly related to the simulator. For example, whereas for a hexapod system the small angle approximation is typically used (Ellensohn, 2020), this assumption is no longer valid on a system with a yaw-drive (allowing large yaw angles). Vice versa, a complex algorithm might only be able to exploit the simulator workspace better if it is given enough workspace. It is therefore paramount to realize that a single MCA can have different types of implementations across different available simulators. Hence, the complexity of adding various simulators to the analysis is *coupled* to the MCA choice, as different MCA approaches can perform differently across simulators.

We name the combination of a motion cueing method and a set of tuning parameters as selected for an experiment the *motion cueing solution*. The goal of any MCA selection and tuning is thus to choose the best-suited motion cueing solution for a given experiment. This choice is ideally made before the simulator choice, in which the exact eventual tuning is still unknown. Thus, an approximated version of each motion cueing solution must be used, as the actual tuning is not yet performed. It depends on the motion cueing method choice if time will be spent on tuning that MCA. It would be highly cost-inefficient to tune all available algorithms for all simulators. This is a causality dilemma: the quality of an MCA can only be tested once it is tuned, but its selection can only be justified if it is known how well it will perform. It is therefore difficult to make exact and informed decisions on which motion cueing method suits the use-case best. A prediction must therefore be made to evaluate how the different options perform with respect to each other. The more accurate the prediction tool, the more reliably the process of motion cueing method selection can be performed.

2.2. Experiment stakeholders

To identify the importance of various quality and cost metrics, we focus on the typical stakeholders that are involved in driving simulator experiments at BMW:

Experimenter - Wants to perform an experiment to answer a research question. Therefore, generally experimenters want a simulation that is most suitable for answering this research question.

Motion Cueing Engineer - Designs and/or tunes the MCA according to the wishes of the experimenter, and possibly of other stakeholders, to achieve the best-suited motion cueing solution.

Simulator Operator - Operates the simulator from the control room during the experiment and governs the safety and well-being of the participant. At BMW, this role is often performed by an external company.

Participant - Participates in the experiment. They might have the smallest impact on the design of the experiment, but can still have clearly defined wishes *during* the experiment.

Note that in some cases, the same individual may have multiple stakeholder roles: e.g., an individual could be the experimenter and the simulator operator at the same time.

2.3. Quality metrics

Based on experience, several quality metrics are proposed, which are all ideally as high as possible.

Perceptual Fidelity - When the presented motion is perceived as unrealistic, perceptual fidelity is low (Cleij, et al., 2018). To increase perceptual fidelity, more accurate MCAs and/or larger motion simulators are generally required. The importance of perceptual fidelity can differ per experiment, and can also be used as a means to the goal of reducing simulator sickness (Hogerbrug, et al., 2020).

Behavioural Fidelity - The behavioural fidelity of the simulation is determined through the degree of similarity in driving behaviour compared to the behaviour in its real world equivalent, rather than whether what they perceive feels realistic. This can occur on various levels independently, such as the lateral and longitudinal driving behaviour or the interaction with traffic, as well as handling secondary driving tasks (such as operating a navigation system).

Stability - A motion cueing solution that is unstable can lead to unexpected and/or dangerous behaviour of the system. This can affect the quality of the motion, but also decrease the system safety, resulting in a higher risk of incidents. Especially for model-predictive control algorithms the associated inherent stability has shown to sometimes be an issue (Fang and Kemeny, 2016).

2.4. Cost metrics

In contrast to quality metrics, cost metrics are generally to be minimized as much as possible.

Simulator Sickness - Simulator sickness can occur when the stimuli presented in the simulator are inconsistent or deviate from what is expected from reality (Hogerbrug, et al., 2020). Light simulator sickness may affect the outcome of an experiment and is thus generally to be avoided. Severe simulator sickness can even lead to participants dropping out of the experiment, resulting in incomplete experiment data.

Cost - The cost refers to the total financial cost of performing the experiment. This includes the cost of preparing the simulation (man hours) and the cost of running the simulator (energy consumption, maintenance, etc.), which generally both scale with the platform size and MCA tuning complexity. For example, the smallest simulator with the easiest to tune MCA will likely result in the lowest cost.

Energy Consumption - Unnecessary energy consumption can be undesired to save costs and to minimize the environmental impact. Therefore, this metric can be an argument for choosing the smallest simulator with the least motion available that is still able to answer the research question.

Noise - Noise and/or vibrations of the motion system perceived inside the cabin can possibly decrease the perceptual fidelity. Larger, heavy systems tend to generate both more noise and vibrations due to the combination of a high weight and large platform movements. Secondary, noise can be perceived inside the building and affect other experiments.

Maintenance - Maintenance increases the cost and reduces the time in which the simulator can be operational for other experiments. Although this aspect is somewhat specific to the simulator design, the maintenance is likely to scale with the size of the system and the amount of platform movement.

2.5. Tuning metrics

Finally, tuning metrics are generally only relevant to the motion cueing engineer and describe inherent properties of finding a suitable set of parameters (i.e., tuning) of a motion cueing method.

Number (#) of tuneable parameters per axis - As each simulator axis requires tuning, the complexity of each axis depends on how many parameters have an effect on that axis. For example, a CWA might require several parameters in each axis (gains, split frequencies, and washout parameters). For larger motion systems, which typically have more Degrees of Freedom (DoFs) (DoF$= 9$ for BMW's Sapphire Space (Kolff, et al., 2022)) than hexapod-only simulators (DoF$= 6$), the total number of parameters for each axis is generally also higher, because these are required to divide the motion over the available DoFs.

Total Number (#) of parameters - Furthermore, the total number of parameters is considered, as in Nahon and Reid (1990). Ideally, the number of tuneable parameters would be as low as possible. Depending on the use-case, however, it is possible that not all vehicle axes require tuning. This can, for example, occur when a lateral-only use-case is performed, in which the longitudinal direction does not require tuning. Thus, this metric depends on the simulator choice, MCA, as well as the considered use-case.

Transparency of parameters - The cost due to the number of parameters also depends on how fast the desired change in tuning can be obtained through changing the parameters (Nahon and Reid, 1990). Thus, the clearer it is which effect a parameter has, the better.

Determinism of System - Some MCA types will always give the same output when the same input is provided. An example of this is CWA, as the linear filter structure follows a deterministic approach. In contrast, MPC is based on optimization, which highly depends on the initial values that are given to the algorithm. Thus, it is highly unlikely that the same outputs are provided, given the same inputs. A non-deterministic system arguably leads to a higher tuning complexity, because its outcome is more difficult to predict.

Offline testing capabilities - When an MCA can be tested outside of the normal simulation environment, this allows for additional debugging and testing capabilities, as no simulator is required for testing. Thus, having no offline testing option can also lead to a higher tuning complexity.

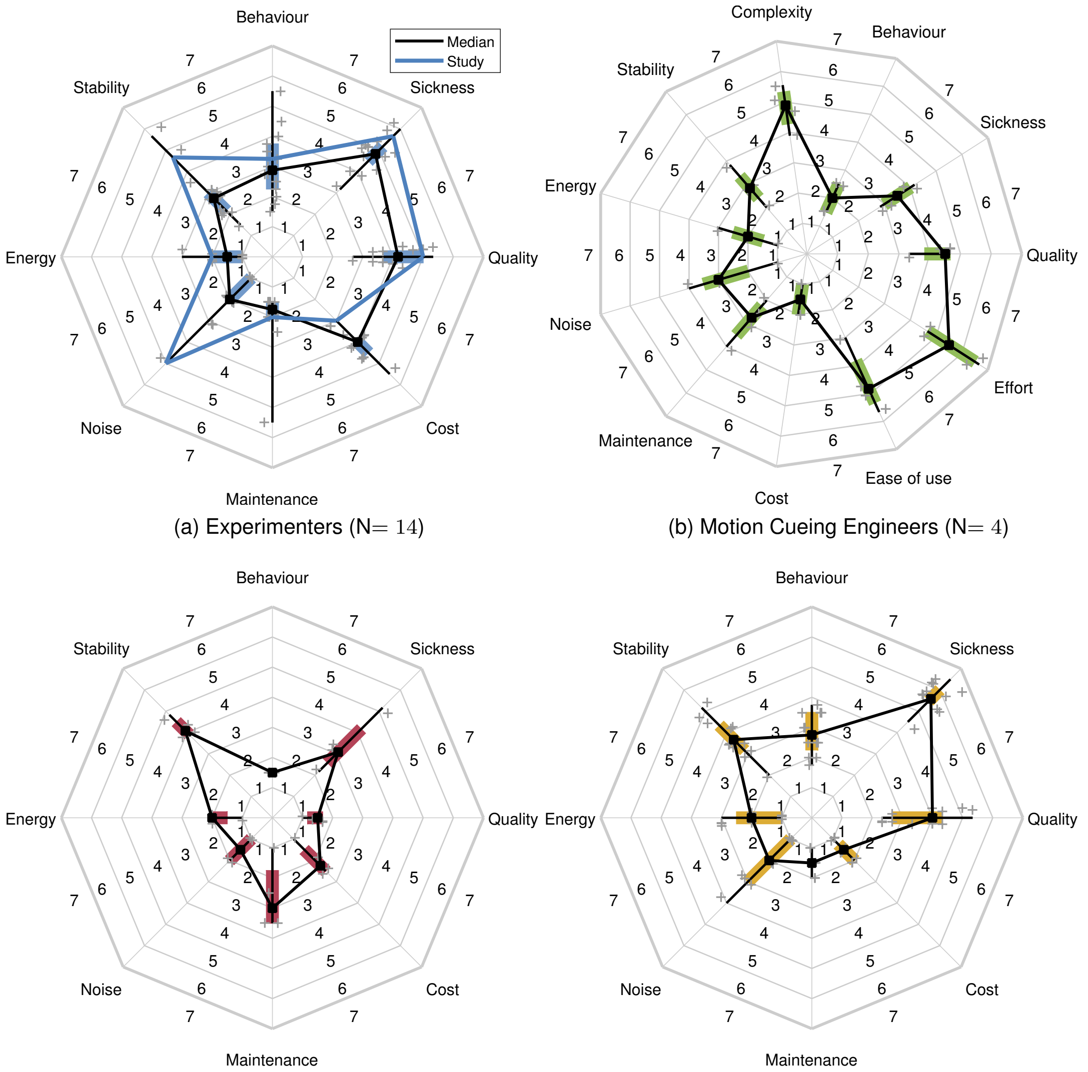
2.6. Survey

To identify *which* of the metrics identified in the previous subsections are important for different experiment stakeholders, a survey was performed. The same survey was given to all four stakeholder groups to increase the comparison between them. It consisted of 35 questions (in German) on MCA quality, divided into several categories. Answers were given on a seven-point Likert scale from fully disagree - disagree - somewhat disagree - neutral - somewhat agree - agree - fully agree. The questions of the survey are shown in Appendix A, translated to English. As the last category "Tuning metrics" typically only applies to motion cueing engineers, this category is only filled out by this group.

In total, 14 experimenters, 4 motion cueing engineers, 4 operators and 12 participants of previously performed experiments filled out the survey. The respondents filled out the survey in a document themselves after receiving a short briefing on the context of the survey. The respondents were specifically instructed to answer the questions in a way that corresponds to the interests associated to their role. For further analysis, questions belonging to the same metric (e.g., simulator sickness) are combined by taking the mean value of these questions.

3. Survey Results

Figures 2a-2d show the survey results, displayed as radar charts, similar to the representation of (Fischer, et al., 2015). For each category, a box plot is shown along each 'spoke' of the radar charts. The lines on each axis represent the data ranges, the colored areas the first and third quartiles, and the grey crosses



(a) Experimenters (N$=14$)

(b) Motion Cueing Engineers (N$=4$)

(c) Operators (N$=4$)

(d) Participants (N$=12$)

Figure 2: Radar plots showing the box plots over all survey results for the eleven defined quality, cost, and tuning aspects.

the individual data points. The medians of the data are visualized by the black dots, connected by black lines for increased readability. The values $(1-7)$ correspond to the answers of the survey (fully disagree - fully agree). Thus, lower values indicate less importance, higher values indicate higher importance.

For the "experimenter", Figure 2a, the results were collected from fourteen past or planned studies. Generally, the quality of the motion, reducing simulator sickness, and the somewhat cost of the experiment were shown to be the most important metrics for these studies (i.e., >4, so more than the answer 'neutral').

Four "motion cueing engineers" filled out the survey (Figure 2b). Note that here the data points do not represent individual studies, but represent what these engineers generally find important. Unsurprisingly, aspects of tuning complexity and tuning effort are found most important by these engineers. Ease of use (in the operations) shows somewhat of an importance. Although the engineer is not the stakeholder performing the actual experiment, an easy-to-use MCA can still be a key benefit when designing, debugging, and testing an MCA.

The "operator" (Figure 2c) has two clear metrics of importance. First, the ease-of-use, which is explained by the operator having to use the motion cueing method in an experiment. Second is stability: A stable motion cueing method avoids obstructive situations at the simulator, where valuable time is thrown away or debugging must be performed during the experiment.

Stability is a similar issue for the "participant" group (Figure 2d). Here, twelve survey respondents indicated that this metric is of medium importance, which could be in the interest of protecting their own safety in the simulator. Most importantly, perhaps unsurprisingly, is simulator sickness, which was indicated by all respondents to be of high importance (>5). Note that the results of the stakeholder "organization" are not presented here, due to data protection reasons.

4. Example Evaluation

4.1. Experiment use-case

To highlight the quality, cost, and tuning metrics of a specific experiment, a real, *future* driving simulation experiment at BMW is analyzed. Therefore, the exact details of the experiment are unknown. The experiment will investigate the role of fatigue under an Advanced Driver Assistance System (ADAS) system. First, participants will drive for 15 minutes on a German highway scenario. After that, the autonomous driving systems will be engaged until the participant reports a high level of drowsiness. This phase can take up to 90 minutes. After that, a second manual segment is performed, lasting another 15 minutes. This use-case contains mainly longitudinal motion (braking/accelerating) and some lateral motion (overtaking).

The survey results of this use-case are highlighted by the blue lines in Figure 2a. Two large outliers compared to the overall data are present. In this experiment, as specific interest is taken in fatigue while driving autonomously, the noise of the simulator is to be reduced as much as possible, as this can influence the fatigue and attentiveness of drivers. Furthermore, as it considers an experiment in a relatively long scenario (> 120 minutes) of a single drive, stability is important, as a simulator hardware or software crash can render the long experiment sessions useless.

From these results, the metrics noise, stability, simulator sickness, motion cueing quality and tuning complexity are considered, as these were indicated as most important metrics over all stakeholders (> 4). To illustrate how the identification of these metrics can be used in a trade-off, four different simulators (the same as in Kolff, et al., 2022) considered.

4.2. Simulators

- The *Vega Vector (VV)* (Cruden B.V., 2021, Figure 3a) is the smallest simulator under investigation and consists of a 6-DoF hexapod with an actuator stroke of 64 cm. Its cylindrical $220°$ LED-wall allows for high contrast visuals combined with high brightness and vivid colors.
- The *Sirius Vector (SV)* (Cruden B.V., 2021, Figure 3b) has the same hexapod as the Vega Vector, but has an additional $\pm 175°$ yaw-drive underneath, resulting in a 7-DoF system. The LED-wall is similar to the Vega Vector, but covers the full $360°$ horizontal field-of-view.
- The *Ruby Space (RS)* (VI-Grade, 2015, Figure 3c) is a 9-DoF system (hexapod on tripod). The tripod can rotate $\pm 25°$ and also has a 1.5 m workspace in both x and y-directions.
- The *Sapphire Space (SS)* (Van Halteren Technologies B.V. and AVSimulation, 2021), Figure 3d) is BMW's largest simulator (9-DoF). It includes a large 6-DoF hexapod (total stroke of 1.15 m) with a single DoF $\pm 175°$ yaw-drive on top. Its XY-drive underneath allows additional movement over an area of 19.14 m $\times$ 15.7 m. Visuals are projected by a full $360°$ projection system inside the enclosed dome.

4.3. Motion Cueing Algorithm

A typical CWA is considered (Conrad, Douvillier, and Schmidt, 1973; Reid and Nahon, 1985). Due to the worst-case tuning and the CWA's 'blindness' to future states, it cannot utilize the simulator full potential. Typical for CWA is the need for tuning its large number of parameters, a time-consuming process.

4.4. Evaluation

Motion Cueing Quality To evaluate the various combinations in terms of the motion cueing quality, predictions of their subjective ratings are made. Cleij, et al. (2018) introduced the use of continuous ratings, in which drivers continuously rate the realism of the motion through a rating interface on a scale from 0 (perfect motion) to 10 (highly unrealistic motion). This method is only possible in open-loop driving (i. e., drivers are *passengers*) and results in a continuous rating signal $R(t)$. A rating model is employed to predict the continuous ratings as function of mismatch signals, i.e., the difference in inertial motion (specific forces and rotational rates) between the vehicle motion $\widetilde{S}_{veh,n}(t)$ and the simulator motion $\widetilde{S}_{sim,n}(t)$, i. e., $\Delta\widetilde{S}_m(t)$, with $\widetilde{P}_m(t) = |\Delta\widetilde{S}_m(t)|$. Here, m represents the mismatch direction.

Kolff, et al. (2023b) proposed a linear model that predicts the continuous rating of the average participant using low-pass filter transfer functions $H_m(s)$ between the measured mismatch signals $\widetilde{P}_m(t)$ (inputs) and a modeled rating signal $\tilde{R}(t)$ (output):

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

with the cut-off frequency ω_c and $K_{\widetilde{P}_m}$ the gains of the several mismatch channels. The $\widehat{(\cdot)}$-terms indicate the Fourier transforms. Kolff, et al. (2023b) showed that the continuous ratings of a classical washout algorithm as measured in that experiment could be largely explained when considering the mismatch channels $\widetilde{P}_{f_y}$, $\widetilde{P}_{f_x}$, and $\widetilde{P}_{\omega_z}$, with respective gains of 0.93, 0.66, and 2.77, together with $\omega_c = 0.35$ rad/s.

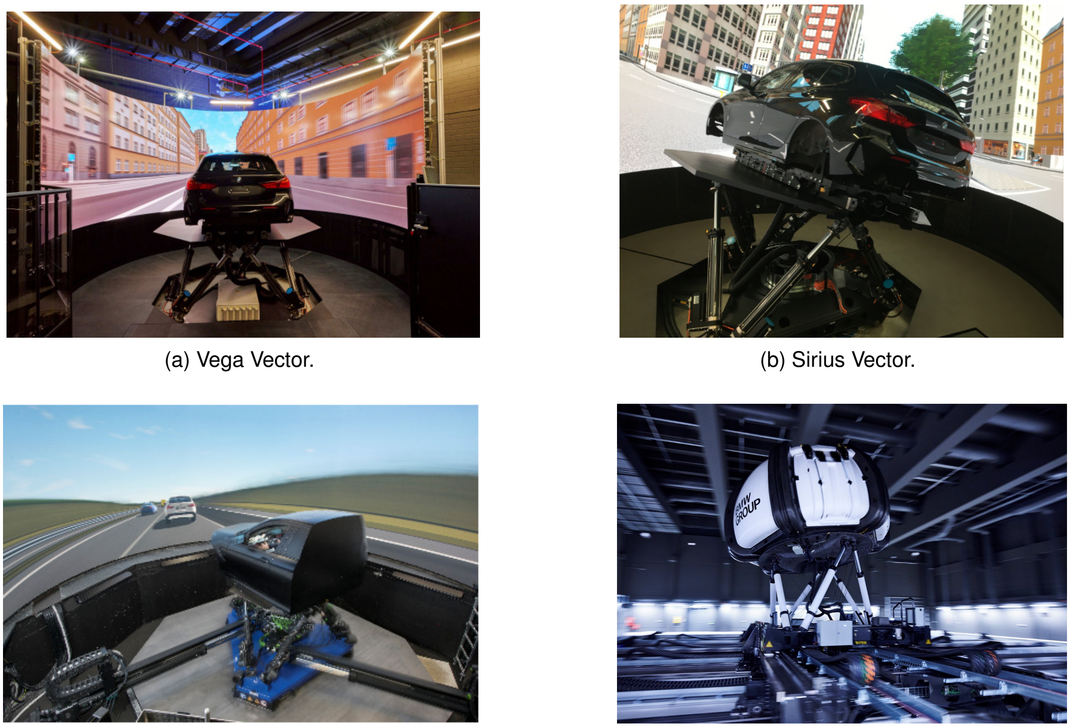
The usefulness of predicting continuous ratings lies in the fact that these correlate well to overall ratings ('OR'), which are given at the end of a drive. Kolff, et al. (2023a) found a predictive relation between $R(t)$ and OR by considering the most incongruent point, i. e., the maximum of the continuous rating:

$$OR = 0.67 \cdot \max[\widetilde{R}(t)] - 0.14 \tag{2}$$

This allows for comparing the various motion cueing methods of the experiment in an offline manner. The overall rating is especially useful, as it can be used to trade off various motion cueing methods.

Note that the rating model requires vehicle and simulator data that is normally not present *before* performing the experiment. However, in the case of the considered experiment, driving data of a highly similar experiment were available, containing drives of 24 participants. For each drive, the modeled rating and corresponding overall rating are calculated. The average over all participants is calculated to obtain a prediction of the average overall rating.

Number of Parameters: The number of parameters is considered as a measure of tuning effort and com-



(a) Vega Vector. (b) Sirius Vector.

(c) Ruby Space. (d) Sapphire Space.

Figure 3: The four considered simulators for comparison. Pictures are property of BMW Group.

plexity. For the CWA, the number of parameters depends on the motion system. For the smallest system, Vega Vector, there are a total of 26 parameters. These include gains, split frequencies, as well as washout parameters. For the Sirius Vector there two additional parameters, necessary for the yaw-drive (one cut-off frequency, one washout frequency parameter), i.e., 28 parameters. For Ruby Space and Sapphire Space, there are another six additional parameters (two cut-off frequency, four washout frequency parameters) due to the presence of an xy-drive, resulting in 34 parameters.

Noise: Currently, no models exist to predict the noise production of the various simulators. From experience, however, the large mechanical systems of the Sapphire Space can induce noise and vibrations at lower frequencies (i.e., noise is 'high'). Similarly, Ruby Space induces some noise due to the air pressure system that controls the secondary motion system (i.e., noise is 'medium'). From our experience, both the Vega Vector and Sirius Vector are indeed the most quiet (noise is 'low').

Simulator Sickness: Previous research has shown that on a highway scenario, simulator sickness is generally low on all considered systems (Himmels, et al., 2022). Although no explicit information on the Ruby Space exists, the dimensions of the motion system are in-between Sirius Vector and Sapphire Space, such that motion-induced simulator sickness is likely similarly as low.

Stability Stability is considered to be an important metric in the evaluation due to the experiment length: a stability issue would likely require restarting the long scenario, such that large parts of the data must be discarded. Considering the known, stable behaviour of an CWA, all simulators perform equally well here ('high').

Metric Evaluation: Table 1 shows the most important metrics of the considered experiment, for the four simulators. Although important for the presented experiment, the stability of a CWA is generally high and does not vary across the motion platform.

The predicted overall rating is similar between Vega Vector and Sirius Vector. Although Sirius Vector contains an additional yaw-drive, thus potentially reducing the ratings, it is likely that the presented scenario does not include large yaw rate errors, such that the yaw-drive is not of a large benefit in this considered experiment. Both Ruby Space and Sapphire Space perform better in terms of the predicted overall rating (lower is better) and expected simulator sickness occurrences. However, this comes at the cost of having more tuning parameters and more noise. As the gain in motion cueing quality rating is rather small, and the high importance of noise for this specific experiment, the Vega Vector and Sirius Vector both show to be the best-suited options.

5. Discussion

Next to aspects of motion cueing quality (behavioural or perceptual), the survey results revealed several other aspects are generally of importance: The cost and simulator sickness for the experimenter, ease of use, tuning effort, and complexity for the motion cueing engineer, stability and ease of use for the operator, and (again) the stability and the simulator sickness for the participant. In the experiment use-case

Table 1: Considered MCA quality metrics of the four motion system for the example evaluation.

Simulator	MCA	Pred. Overall Rating [0-10]	# Tuning parameters	Noise	Simulator Sickness	Stability
Vega Vector	CWA	2.8	26	low	medium	high
Sirius Vector	CWA	2.7	28	low	medium	high
Ruby Space	CWA	2.4	34	medium	low	high
Sapphire Space	CWA	2.4	34	high	low	high

example, an additional focus lies on noise, indicated by the wishes of the experimenter. It is possible that due to practical experience, existing metrics are refined or more metrics are added. We also invite the community to provide insights into their relative importance of such metrics, as these insights might differ per institution, as different use-cases are covered. This could also lead to more data points for the operators and motion cueing engineers, which is inherently limited for a single company.

Through the presented experiment example, it is shown that the survey can serve not only as a one-time identification tool of important metrics, but also as the baseline for a framework of future experiment design, where the survey is filled-out in early experiment design stages to identify the best-suited MCA for that experiment. By identifying the important metrics for a given experiment beforehand, seemingly trivial aspects, such as the noise production, could show to be of high importance, such as in the presented fatigue experiment example. Thus, here it could be preferred to select and MCA and/or simulator that produces little noise, while a limited reproduction of the motion cues is perfectly acceptable. In this presented example, the obtained result is different than when only motion cueing quality would have been considered. In that case the Sapphire Space would have been the best-suited option. This shows that the trade-off between MCAs, simulators, and tuning configuration can benefit from including these metrics as well.

Nevertheless, several limitations in the presented work should be mentioned. First, the identified metrics could benefit from a more objective measurement approach, rather than the subjectively estimated categories (such as low, medium, and high). For example, the noise production of simulators could be expressed in absolute values. Second, note that for the survey itself, a fifth stakeholder could also be defined, the *organization*. This group owns the simulator and aims to protect organizational interests, which could be governing cost, safety, and public image. In our work, we did not include this stakeholder due to data protection reasons. Including this might put more emphasis on financial cost and energy consumption. Third, the final trade-off *between* the metrics remains subjective. It is possible to make this choice by setting lower bounds to eliminate MCAs that do not meet the requirements or, if no defined lower bounds exist for the experiment, calculating the “overall quality”, by calculating a total cost. Here, a trade-off could be made more objective by taking the relative importance of the various metrics into account (e.g., by the relative importance of the considered metrics). Finally, the presented methods only apply to metrics related to the motion cueing. In practice, especially when considering different motion systems, other metrics might be of importance as well, such as the availability of specific mock-ups in a given simulator or the (type of) visual system and/or the associated synchronization between the visual scene and the motion. Future work could thus aim to include such properties as well.

6. Conclusion

The successful use of motion cueing algorithms in driving simulation does not only depend on the quality of the generated motion, but also how well the algorithm performs operationally. Both the analysis of general operational MCA qualities, as well as specific wishes of operational MCA qualities for each experiment, can be better understood through understanding the wishes of the users, i.e., the stakeholders. From the survey evaluations, it can be concluded that, next to the motion cueing quality, the total cost, simulator sickness, easy of use, tuning effort, and stability are typically of high importance in motion cueing experiments. In the provided experiment use-case example, including such operational aspects lead to a different best-suited motion cueing solution than when considering motion cueing quality alone. Improving the simulations in these aspects can therefore likely result in more successful driving simulation experiments. With the proposed method, therefore the best MCA with the highest overall quality for a given use case can be selected for future driving simulator experiments.

References

Cleij, D., Venrooij, J., Pretto, P., Pool, D. M., Mulder, M., and Bülthoff, H. H., 2018. Continuous Subjective Rating of Perceived Motion Incongruence During Driving Simulation. *IEEE Transactions on Human-Machine Systems*, 48(1), pp. 17–29.

Conrad, B., Douvillier, J. G., and Schmidt, S. F., 1973. Washout Circuit Design for Multi-Degrees-Of-Freedom Moving Base Simulators. In: *Proc.of the AIAA Visual and Motion Simulation Conference*. AIAA-1973-929. Palo Alto, United States.

Dagdelen, M., Reymond, G., Kemeny, A., Bordier, M., and Maïzi, N., Apr. 2009. Model-based predictive motion cueing strategy for vehicle driving simulators. *Control Engineering Practice*, 17(2009), pp. 995–1003.

Ellensohn, F., 2020. Urban Motion Cueing Algorithms - Trajectory Optimization for Driving Simulators. PhD Dissertation. Technische Universität München.

Fang, Z. and Kemeny, A., Oct. 2016. An efficient Model Predictive Control-based motion cueing algorithm for the driving simulator. *SIMULATION*, 92.

Fischer, M., Labusch, A., Bellmann, T., and Seehof, C., 2015. A Task-oriented Catalogue of Criteria for Driving Simulator Evaluation. In: *Proc.of the Driving Simulation Conference 2015 Europe*. Tübingen, Germany, pp. 139–150.

Himmels, C., Venrooij, J., Gmünder, M., and Riener, A., 2022. The influence of simulator and driving scenario on simulator sickness. In: *Proc.of the Driving Simulation Conference 2022 Europe*. Strasbourg, France, pp. 29–36.

Hogerbrug, M., Venrooij, J., Pool, D. M., and Mulder, M., 2020. Simulator Sickness Ratings Reduce with Simulator Motion when Driven Through Urban Environments. In: *Proc.of the Driving Simulation Conference 2019 Europe*. Antibes, France.

Kolff, M., Venrooij, J., Arcidiacono, E., Pool, D. M., and Mulder, M., 2023a. Predicting Motion Incongruence Ratings in Closed- and Open-Loop Urban Driving Simulation. *IEEE Transactions on Intelligent Transportation Systems*. Under review, pre-print available.

Kolff, M., Venrooij, J., Pool, D. M., and Mulder, M., 2020. Comparison of Quality Metrics between Motion Cueing Algorithm in a Virtual Test Environment. In: *Proc.of the Driving Simulation Conference 2020 Europe*. Antibes, France.

Kolff, M., Venrooij, J., Schwienbacher, M., Pool, D. M., and Mulder, M., 2022. Motion Cueing Quality Comparison of Driving Simulators using Oracle Motion Cueing. In: *Proc. of the Driving Simulation Conf. 2022 Europe*. Strasbourg, France, pp. 111–118.

Kolff, M., Venrooij, J., Schwienbacher, M., Pool, D. M., and Mulder, M., 2023b. Reliability and Models of Subjective Motion Incongruence Ratings in Urban Driving Simulations. *IEEE Transactions on Human-Machine Systems*. Under review, pre-print available.

Nahon, M. A. and Reid, L. D., 1990. Simulator Motion-Drive Algorithms: A Designer's Perspective. *Journal of Guidance, Control, and Dynamics*, 13(2), pp. 356–362.

Reid, L. D. and Nahon, M. A., 1985. *Flight Simulation Motion-Base Drive Algorithms: Part 1. Developing and Testing the Equations*. Tech. rep. UTIAS 296. University of Toronto, Institute for Aerospace Studies.

Romano, R., Markkula, G., Boer, E., Jamson, H., Bean, A., Tomlinson, A., Horrobin, A., and Sadraei, E., 2019. An objective assessment of the utility of a driving simulator for low mu testing. *Transportation Research Part F: Traffic Psychology and Behaviour*, 65(1), pp. 34–45.

Appendix A: Survey Questions

The appendix contains the questions as presented to the stakeholders. The questions have been translated from German to English.

Immersion - During the simulation, a feeling of being in the virtual world can arise (immersion). Please use the following questions to answer whether it is important to you that:

Q1 the perceived motion in the simulator is as realistic as possible.
Q2 participants are immersed into the virtual world.
Q3 the simulator motion does not behave unexpectedly.
Q4 the simulator is never driven into its limits.
Q5 the simulation of a drive never stops working (simulation crash).

Behavior of the driver - Using the following questions, please rate whether it is important to you that:

Q6 the same driving style in the transverse direction (steering) is induced as in the real vehicle.
Q7 the same driving style in the longitudinal direction (pedals and/or shifting gears) is induced as in the real vehicle.
Q8 the same behavior with certain secondary components (navigation, radio, ABK) is induced as in the real vehicle.
Q9 the same behavior with the surrounding traffic is induced as in the real vehicle.

Simulator sickness - Simulator sickness can occur during a ride. Sometimes the attempt has to be stopped because of this. Using the following questions, please rate whether it is important to you that:

Q10 Heavy simulator sickness symptoms (nausea, throwing up) are reduced as much as possible.
Q11 Light simulator sickness symptoms (headache, dizziness, sweating) are reduced as much as possible.
Q12 the overall failure rate of the experiment is as low as possible.
Q13 the drive can always last until the end.
Q14 the participant is still able to work after the study.
Q15 the recorded experiment data of each drive is always complete.

Driver safety - Using the following questions, please rate whether it is important to you that:

Q16 the participant always feel safe in the simulator.
Q17 the health of the participant is not harmed.

Hardware - Using the following questions, please rate whether it is important to you that:

Q18 as much electricity as possible is saved during the study.
Q19 the simulator is not damaged.
Q20 fuel is saved by performing the study virtually, rather than with a real vehicle.
Q21 simulator maintenance is minimized.
Q22 the simulator generates little noise while moving.
Q23 the simulator does not move unnecessarily.

Costs - Using the following questions, please rate whether it is important to you that:

Q24 the cost for preparing and executing the experiment are minimized as much as possible.
Q25 the costs for the organization are as low as possible.

Ease of use

Q26 no unclear errors in the simulation operation software are generated.
Q27 all options in simulation operation software requiring changes during a simulation are self-evident.

Tuning metrics - Using the following questions, please rate whether it is important to you that:

Q28 there are as few tuneable parameters as possible.
Q29 the tuneable parameters can be interpreted physically.
Q30 the MCA behaves linearly between input and output.
Q31 there is only one parameter for each axis.
Q32 The meaning of parameters is self-explanatory.
Q33 there are as few parameters as possible in each degree of freedom of the simulator.
Q34 the tuning of a previously performed study can be used.
Q35 the tuning can be tested without a simulator.
