Motion Cueing Quality in Driving Simulation

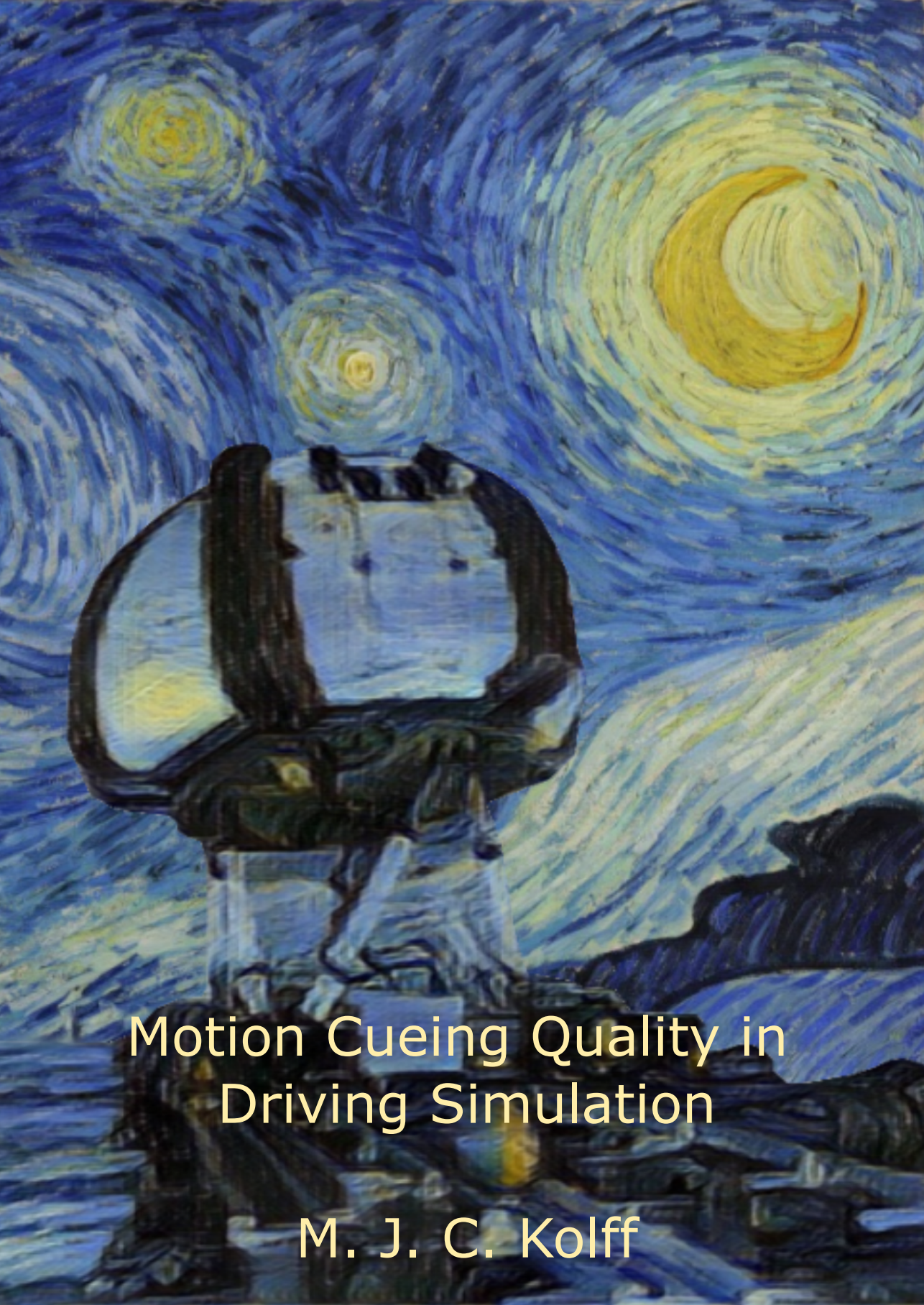
Kolff, M.J.C.

DOI
10.4233/uuid:a9c8abb2-c45d-400e-8ef1-ed1c76515b44

Publication date
2025

Document Version
Final published version

Citation (APA)
Kolff, M. J. C. (2025). *Motion Cueing Quality in Driving Simulation*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:a9c8abb2-c45d-400e-8ef1-ed1c76515b44



Motion Cueing Quality in Driving Simulation

M. J. C. Kolff

MOTION CUEING QUALITY IN DRIVING SIMULATION

MOTION CUEING QUALITY IN DRIVING SIMULATION

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T. H. J. J. van der Hagen;
Chair of the Board for Doctorates
to be defended publicly on
Monday 29, September 2025 at 15:00 o'clock

by

Marius Johannes Claus KOLFF

Master of Science in Aerospace Engineering,
born in Werkendam, The Netherlands.

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. M. Mulder	Delft University of Technology, promotor
Dr. ir. D. M. Pool	Delft University of Technology, copromotor
Independent members:	
Dr. J. L. Campos	University of Toronto
Prof. dr. ir. L. L. M. Veldhuis	Delft University of Technology
Prof. dr.-ing. J. Häcker	Baden-Wuerttemberg Cooperative State University
Prof. dr. ir. J. C. F. de Winter	Delft University of Technology
Dr. ir. D. Cleij	Institute for Road Safety Research

Dr. ir. J. Venrooij of BMW Group has contributed greatly to the preparation of this dissertation.

This research was funded by BMW Group and supported by Delft University of Technology.

Keywords: Motion cueing, driving simulation, quality comparison, operations optimization

Printed by: Ipskamp Printing

Front & Back: M.J.C. Kolff, based on Starry Night (Vincent van Gogh). Simulator image: BMW Group.

ISBN 978-94-6473-936-7

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

Table of Contents

Summary

Motion Cueing Quality in Driving Simulation

by

Marius Johannes Claus KOLFF

Driving simulators are important tools in supporting the research and development of automotive systems. By replacing real-world testing with a virtual equivalent, significant benefits can be achieved in terms of financial cost, controllability, sustainability, and safety. The role of a simulator's motion system is to reproduce the physical motion cues as they would be perceived in the real vehicle. An inaccurate motion reproduction can lead to *simulator sickness*, unwanted adaptations in the driver's control behavior, and a reduced validity of experimental results. The control of the simulator's motion system is performed by the *motion cueing algorithm* (MCA). It converts the motion of the real vehicle to motion that fits in the constrained workspace of the simulator. This nevertheless inevitably leads to imperfections, which are, when perceived by the driver, known as *incongruences*. For an effective motion cueing and to achieve the best-possible simulator motion, a complete understanding of the motion system control *and* its effect on the human perception of the motion is thus required.

For a growing number of driving simulation use-cases, an increasing variety of different MCAs and simulators is currently used. Motion systems range from small hexapods, to large and advanced configurations consisting of various motion *sub*systems (e.g., xy-drives, tripods, and yaw-drives). MCAs range from "filter-based" algorithms, consisting of causal filters, to more complex, optimization-based algorithms. While perhaps enabling a better possible motion reproduction, using a more advanced MCA and simulator can also have drawbacks, such as an increased cost or complexity of the system. Which MCA and simulator are best suited thus depends on the use-case under consideration. The MCA also requires tuning of its parameters, further increasing the scope of available options. An obvious question arising is then how it can be determined which *combination* of MCA, simulator, and tuning parameter set is best suited for a certain driving simulation use-case. This requires an evaluation of each combination's *motion cueing quality* across the multiple dimensions that the term "quality" may entail.

In existing methods and literature, motion cueing quality is generally only evaluated in terms of the motion reproduction itself. This dissertation shows that considering a set of quality, cost, and tuning aspects, important to the various stakeholders involved in an experiment, provides a broader view of the motion cueing quality. This could, for example, include the quality of the motion, together with the energy consumption, total

financial cost, and required tuning effort. The evaluation of this "total" motion cueing quality can lead to a more well-founded choice of the simulator, MCA, and its parameters for a driving simulator experiment. Another current limitation is that in the early experiment design stages, the tuning of the MCA has not yet been performed. To justify the cost and time investment required for tuning a given MCA and simulator, an evaluation of the quality is necessary, but this quality cannot be measured exactly, *because* the tuning has not yet been performed. A key innovation presented in this dissertation is a method to *predict* the *total* motion cueing quality that an MCA and simulator can *potentially* reach, to improve the selection process of driving simulator motion cueing.

Predicting the potential quality of the motion reproduction introduces a further difficulty, as a *model* of subjective assessments (i.e., ratings) on the incongruence of the motion by human test drivers is required. This is notoriously difficult due to the complexity and black-box nature of human perception and subjective rating behaviour. Recent work, however, introduced a *continuous* rating method, in which drivers are tasked with continuously evaluating the motion incongruence of a simulated scenario through a rating interface. This method is especially suitable for developing rating models due to the high temporal resolution of the ratings. Restricted by the additional workload that comes with continuously evaluating the motion, which can interfere with the task of driving, the method is only feasible in *open-loop* driving, meaning that the test driver is driven around rather than actively steering the vehicle. In most driving simulation use-cases, however, the driver needs to actively control the vehicle in a *closed-loop* setting instead (e.g., vehicle handling testing). While shown to be promising in modeling approaches for simplified scenarios, the continuous rating method thus requires the validation that it also produces reliable ratings relevant for closed-loop driving.

Through five novel experiments described in this dissertation, and the use of three pre-existing data sets, an explicit investigation on the suitability of the continuous rating method in realistic driving simulation use-cases is performed. The three main scenario types of typical driving simulations (urban, highway, and rural) are investigated through dedicated continuous rating experiments and subsequently compared, providing a direct and unique insight in differences *between* scenario types. Estimates on the reliability of the rating data show that the urban scenario produces more reliable rating data than the highway and rural scenarios. This indicates that highway and rural simulations are more difficult to rate, and therefore also more difficult to predict. Furthermore, inverse relations exist between the ratings and their reliability, showing that the *more* congruent (i.e., the better) the motion is rated to be, the *less* reliable the ratings are. Whereas congruence of the motion is generally the goal, ratings of *incongruent* motion are thus more useful for developing accurate rating models.

Based on the recorded continuous rating data of the experiments, *models* of the ratings are subsequently developed, as a function of the objective mismatch signals of the platform motion (specific forces and rotational rates). These rating models provide a transformation of quality predictions from objective mismatch signals to subjective ratings by yielding the predicted subjective rating for the *average* participant. The mismatch signals are shown to be dependent on, and derived for, the variety of existing kinematic systems. The analysis in this dissertation shows that the continuous ratings in

urban simulations can be modeled by a low-pass filtered response of the objective mismatches of the lateral and longitudinal specific forces, with only a small and situational contribution of yaw rate mismatches. The rating data of the rural and highway scenarios are then subsequently used to improve the validity of the rating model. While the balance of the mismatch channels differs per scenario type, the same motion channels are found to be dominant in all three scenario types, which shows a high generalizability and validity of the model *across* the various scenario types.

While continuous ratings are highly useful in modelling, they are less useful as metrics in direct comparisons of MCA and simulator quality. Comparative assessments between configurations are easier to perform based on single value (i.e., scalar) ratings, rather than the long time series of continuous ratings. This dissertation therefore also investigates the prediction of overall ratings (a single rating representing the whole drive), as well as maneuver-based ratings (one rating given for each maneuver). The novel experiments described in this dissertation, where continuous, overall, and maneuver-based ratings are collected in parallel, show that the most incongruent motion in a drive or maneuver (as measurable by the continuous rating) strongly correlates with the overall or maneuver-based rating, respectively. Combined with the continuous rating model, this allows for also predicting overall and maneuver-based ratings from objective mismatch signals. Tuning processes can benefit from this finding by always considering the most incongruent point as the target for motion optimization.

Most of the presented findings in this dissertation concern measuring and predicting ratings of open-loop driving, which is not necessarily representative for the hands-on closed-loop driving that is the focus of most driving simulator use-cases. A dedicated experiment described in this dissertation shows, however, that subjective ratings of closed- and open-loop driving in urban simulations are in fact equivalent. This confirms the hypothesis that continuous ratings may also be representative for closed-loop driving. Combined, this thus enables an approach in which a continuous rating model is used to predict overall and maneuver-based ratings of closed-loop driving, as a function of objective mismatch signals. This allows for unmatched predictions of the potential subjective ratings in driving simulator experiments. Furthermore, a closed-loop drive can never be reproduced in exactly the same way due to inherent variations in driving behavior. As open-loop driving allows for using recordings of a drive, it provides significant benefits in terms of reproducibility and controllability of the experienced motion mismatches. This makes open-loop testing especially practical to accelerate future tuning processes.

Engineering applications of motion cueing in driving simulators can directly benefit from the presented findings. With a valid rating model based on reliable rating data, accurate predictions on the potential quality of the motion can be made. To this end, a method to determine the potential motion cueing quality is also required. The final contribution of this dissertation is therefore a method to estimate the potential motion cueing quality, in which the prediction models of subjective ratings form a pivotal role. This potential estimation is based on the extrapolation of baseline motion mismatch signals to fully use the workspaces of the available candidate simulators. For each MCA and simulator combination, a potential total motion cueing quality can, and should be, pre-

dicted that represents the true capacity for motion reproduction. While not yet worked out to enable direct quantitative trade-offs, the findings for the two considered use cases show that the inclusion of both the potential and total motion cueing quality can significantly affect the choice of MCA and simulator.

Future research can also directly benefit from the findings described in this dissertation. MCAs can be specifically designed to tailor the subjective human perception, leading to an improved subjective experience. For example, the linear rating model can be included in the optimization scheme of MCAs, directly optimizing for the subjective experience of the driver. Alternatively, the weights of the motion channels in an MCA can be tuned to match the balance of importance found in the identified rating model weights. Finally, the knowledge on the overall and maneuver-based ratings can be used to more heavily penalize or optimize for the most incongruent point(s) in a drive, leading to a better *overall* quality of the motion. These contributions all potentially lead to a better subjective experience in driving simulator experiments.

This dissertation offers both practical guidelines and fundamental methods for improving the motion cueing of future driving simulation experiments. By providing a new view on motion cueing quality, the long lasting and complex problem of driving simulator motion cueing optimization is better understood. Apart from its direct relevance to any driving simulator, these insights are furthermore highly applicable to other research and engineering fields involving motion-base simulators, such as flight and maritime simulation.

Samenvatting

Bewegingsaansturingskwaliteit in Autorijsimulatie

door

Marius Johannes Claus KOLFF

Rijsimulatoren zijn belangrijke hulpmiddelen ter ondersteuning van onderzoek en de ontwikkeling van voertuigsystemen. Door een wegtest te vervangen door een virtueel equivalent, kunnen aanzienlijke voordelen worden behaald op het gebied van financiële kosten, bestuurbaarheid, duurzaamheid en veiligheid. De taak van het bewegingssysteem van een simulator is om de fysieke bewegingsprikkels zo weer te geven als ze in een echt voertuig zouden worden waargenomen. Een onnauwkeurige weergave van beweging kan leiden tot *simulatorziekte*, ongewenste aanpassingen in het stuurgedrag van de bestuurder en een verminderde validiteit van experimentele resultaten. De besturing van het bewegingssysteem van de simulator wordt uitgevoerd door het *Motion Cueing Algoritme* (MCA). Dit zet de beweging van het echte voertuig om in een beweging die binnen de beperkte bewegingsruimte van de simulator past. Dit leidt echter onvermijdelijk tot imperfecties die, als ze door de bestuurder worden waargenomen, *incongruenties* worden genoemd. Voor een effectieve bewegingsaansturing en een zo goed mogelijke simulatorbeweging is dus een volledig begrip nodig van zowel de bewegingsaansturing *alsook* het effect ervan op menselijke bewegingswaarneming.

Voor een groeiend aantal toepassingen van rijsimulatie wordt een steeds grotere verscheidenheid aan MCA's en simulatoren gebruikt. Bewegingssystemen variëren van kleine hexapods tot grote, complexe configuraties bestaande uit meerdere bewegings*sub*systemen (zoals XY-actuatoren, tripoden en gierplatforms). MCA's variëren van "filtergebaseerde" algoritmes met causale filters tot complexere, op optimalisatie gebaseerde methodes. Hoewel geavanceerdere MCA's en simulatoren een betere bewegingsweergave kunnen bieden, kunnen ze ook nadelen met zich meebrengen, zoals hogere kosten of systeemcomplexiteit. Welke MCA en simulator het meest geschikt zijn hangt daarom af van de specifieke toepassing. Daarnaast vereist de MCA een parametrisatie, wat het aantal mogelijke opties verder vergroot. De voor de hand liggende vraag is dan hoe kan worden bepaald welke *combinatie* van MCA, simulator en parameterinstelling het meest geschikt is voor een bepaalde rijsimulatie. Dit vereist een evaluatie van de *bewegingsaansturingskwaliteit* van elke combinatie over de verschillende dimensies die het begrip "kwaliteit" kan omvatten.

In bestaande methoden en de literatuur wordt de bewegingsaansturingskwaliteit meestal alleen beoordeeld op basis van de bewegingsweergave zelf. Dit proefschrift

laat zien dat het overwegen van meerdere kwaliteits-, kosten- en parametrisatiesaspecten die voor de verschillende belanghebbenden in een experiment relevant zijn, een completer beeld geeft van de bewegingsaansturingskwaliteit. Dit kan bijvoorbeeld de bewegingskwaliteit omvatten samen met het energieverbruik, de totale financiële kosten en de benodigde inspanning voor de parametrisatie van een MCA. Het beoordelen van deze "totale" kwaliteit kan leiden tot een beter onderbouwde keuze van simulator, MCA en parameters voor een rijsimulatie-experiment. Een bijkomende huidige beperking is dat de parametrisatie van de MCA vaak nog niet is uitgevoerd in vroege ontwerpfase van een simulatie. Om de inspanning voor deze parametrisatie te rechtvaardigen, is een evaluatie van de kwaliteit nodig, maar die kan niet nauwkeurig worden gemeten *omdat* de parametrisatie nog niet heeft plaatsgevonden. Een centrale innovatie van dit proefschrift is een methode voor het *voorspellen* van de *totale* bewegingsaansturingskwaliteit die een MCA en simulator *potentieel* kunnen bereiken, ter ondersteuning van het keuzeproces van de bewegingsaansturing in rijsimulatie.

Het voorspellen van potentiële bewegingskwaliteit introduceert nog een andere uitdaging, gezien een *model* van subjectieve beoordelingen van door menselijke testbestuurders waargenomen bewegingsincongruenties nodig is. De complexiteit en het zwarte-doos-karakter van menselijke waarneming en subjectieve beoordeling maakt dit complex. In recent onderzoek is echter een *continue* beoordelingsmethode geintroduceerd, waarbij bestuurders tijdens een gesimuleerd scenario de bewegingsincongruentie voortdurend beoordelen via een beoordelingsknop. Deze methode is bijzonder geschikt voor de ontwikkeling van beoordelingsmodellen vanwege de hoge temporele resolutie van de metingen. Wegens de extra belasting voor de testpersoon die een continue beoordeling met zich meebrengt, wat de rijtaak kan verstoren, is de methode in de praktijk enkel toepasbaar tijdens *open-lus* rijden, d.w.z. wanneer de proefpersonen passief rond worden gereden in plaats van zelf actief te sturen. In de meeste toepassingen van rijsimulatie moet echter *gesloten-lus* worden gereden (bv. voor het testen van voertuigdynamica). Hoewel de continue beoordelingsmethode veelbelovend bleek in vereenvoudigde scenario's, is de bevestiging nodig dat deze methode ook betrouwbare beoordelingen oplevert die representatief zijn voor gesloten-lus rijden.

Aan de hand van vijf nieuwe experimenten en drie bestaande datasets onderzoekt dit proefschrift expliciet hoe geschikt de continue beoordelingsmethode is voor realistische rijsimulaties. De drie hoofdscenario's van typische rijsimulaties (stad, snelweg en landweg) worden onderzocht via aparte continue beoordelingsstudies en met elkaar vergeleken, wat een uniek inzicht biedt in de verschillen *tussen* scenario's. Betrouwbaarheidsschattingen tonen aan dat stadscenario's betrouwbaardere beoordelingen opleveren dan snelweg- en landwegscenario's. Dit wijst erop dat laatstgenoemden moeilijker te beoordelen, en dus moeilijker te voorspellen zijn. Bovendien bestaat er een omgekeerd verband tussen beoordeling en betrouwbaarheid: hoe *meer* de beweging als congruent wordt beoordeeld, hoe *lager* de betrouwbaarheid van deze beoordeling. Hoewel congruente beweging normaal gesproken wordt nagestreefd, zijn beoordelingen van *incongruente* bewegingen dus nuttiger voor het ontwikkelen van beoordelingsmodellen.

Op basis van de continue beoordelingsdata uit de experimenten worden vervolgens *modellen* ontwikkeld die beoordelingen voorspellen op basis van objectieve discrepan-

ties in platformbeweging (specifieke krachten en draaisnelheden). Deze beoordelingsmodellen maken het mogelijk om objectieve discrepantiesignalen om te zetten in voorspellingen van beoordelingen, door de beoordeling van een *gemiddelde* deelnemer te modelleren. De discrepantiesignalen zijn afhankelijk van het kinematische systeem en worden daarvoor afgeleid. De analyse in dit proefschrift toont aan dat continue beoordelingen in stadscenario's kunnen worden gemodelleerd als een laagdoorlaatgefilterde respons op objectieve discrepanties in laterale en longitudinale specifieke krachten, waarbij giersnelheden slechts een kleine, situatie-afhankelijke bijdrage leveren. De data uit landweg- en snelwegscenario's worden gebruikt om de modelvaliditeit te verbeteren. Hoewel de kanaalweging scenario-afhankelijk varieert, blijken dezelfde dominante bewegingskanalen in alle drie scenariotypes terug te keren, wat wijst op een sterke generaliseerbaarheid en validiteit van het model in de verschillende scenario's.

Hoewel continue beoordelingen waardevol zijn voor modelontwikkeling, zijn ze minder geschikt voor directe vergelijking van de kwaliteit van MCA's en simulatoren. Vergelijkingen tussen configuraties zijn eenvoudiger te maken tussen enkele beoordelingen (één waarde) dan met lange tijdreeksen van continue beoordelingen. Dit proefschrift onderzoekt daarom ook de voorspelling van *algehele* beoordelingen (één cijfer voor de hele rit) en *manoeuvre-gebaseerde* beoordelingen (één cijfer per manoeuvre). Experimentele gegevens waarin continue, algehele en manoeuvre-gebaseerde beoordelingen parallel zijn verzameld, tonen aan dat de meest incongruente momenten (volgens de continue beoordeling) sterk correleren met algehele en manoeuvre-gebaseerde beoordelingen. In combinatie met het continue beoordelingsmodel maakt dit het mogelijk om algehele en manoeuvre-gebaseerde beoordelingen te voorspellen op basis van objectieve discrepanties. Parametrisatieprocessen van MCA's kunnen hiervan profiteren door zich te richten op de meest incongruente momenten als doel voor optimalisatie.

Hoewel de meeste resultaten betrekking hebben op open-lus rijden, is dit niet noodzakelijk representatief voor de gesloten-lus toepassingen van rijsimulatoren. Een speciaal experiment in dit proefschrift toont echter aan dat subjectieve beoordelingen in open- en gesloten-lus rijden in stadscenario's equivalent zijn. Dit bevestigt de hypothese dat continue beoordelingen ook representatief zijn voor gesloten-lus rijden. In combinatie maakt dit een aanpak mogelijk waarbij een continue beoordelingsmodel wordt gebruikt om algehele en manoeuvre-gebaseerde beoordelingen van gesloten-lus ritten te voorspellen, op basis van objectieve discrepanties. Daardoor worden voorspellingen van subjectieve kwaliteit mogelijk die voorheen niet mogelijk waren. Bovendien kunnen gesloten-lus ritten door inhererente variabiliteit nooit exact worden gereproduceerd. Omdat open-lus ritten het afspelen van eerder opgenomen ritten mogelijk maken, zijn ze bijzonder geschikt voor reproduceerbare, controleerbare experimenten. Dit maakt open-lus simulaties nuttig om het parametrisatieproces te kunnen versnellen.

Technische toepassingen van bewegingsaansturing in rijsimulatoren kunnen direct profiteren van de inzichten uit dit proefschrift. Met een gevalideerd beoordelingsmodel gebaseerd op betrouwbare data kunnen nauwkeurige voorspellingen worden gedaan over de potentiële bewegingskwaliteit. Daarvoor is ook een methode nodig om de potentiële bewegingsaansturingskwaliteit te schatten. Het laatste deel van dit proefschrift beschrijft daarom een methode om de potentiële totale kwaliteit te voorspellen, waarbij

de voorspellingsmodellen van subjectieve beoordelingen een centrale rol spelen. Deze voorspelling is gebaseerd op bestaande metingen van discrepantiesignalen die geextrapoleerd worden tot de volledige benutting van de bewegingsruimte van beschikbare simulatoren. Voor elke combinatie van MCA en simulator kan, en zou, een potentiële totale kwaliteit moeten worden voorspeld, die het daadwerkelijke potentieel van de bewegingsweergave reflecteert. Hoewel deze methode nog niet is uitgewerkt tot een direct kwantitatieve afweging, tonen de besproken toepassingen aan dat het overwegen van zowel potentiële als totale kwaliteit de keuze van MCA en simulator wezenlijk beïnvloedt.

Toekomstig onderzoek kan direct profiteren van de resultaten uit dit proefschrift. MCA's kunnen gericht worden ontworpen op de subjectieve menselijke waarneming, om daarmee een betere bewegingsbeleving mogelijk te maken. As voorbeeld kan het lineaire beoordelingsmodel direct worden geïntegreerd in een optimalisatie-algoritme om expliciet te optimaliseren op subjectieve kwaliteit. Als alternatief kunnen de gewichtingsfactoren van de bewegingskanalen worden aangepast volgens de gewichting geïdentificeerd door de modellen uit dit proefschrift. Tot slot kan de kennis van algehele en manoeuvre-gebaseerde beoordelingen worden gebruikt om bijzonder incongruente momenten sterker te bestraffen of te optimaliseren, om daarmee de *algehele* bewegingskwaliteit te verbeteren. Deze bijdragen kunnen allen leiden tot een betere subjectieve ervaring in rijsimulatiestudies.

Dit proefschrift biedt zowel praktische richtlijnen als fundamentele methoden voor het verbeteren van de bewegingsaansturing in toekomstige rijsimulaties. Door een nieuw perspectief te introduceren op bewegingsaansturingskwaliteit, wordt het al lang bestaande en complexe probleem van bewegingssimulatie beter begrepen. Naast directe relevantie voor elke vorm van rijsimulator zijn de inzichten ook toepasbaar op andere onderzoeks- en technologiegebieden met bewegingssimulators, zoals vlieg- of scheepssimulatie.

Zusammenfassung

Bewegungssteuerungsqualität in der Fahrsimulation

von

Marius Johannes Claus KOLFF

Fahrsimulatoren sind wichtige Werkzeuge zur Unterstützung der Forschung und Entwicklung von Fahrzeugsystemen. Durch den Ersatz von realen Tests durch ein virtuelles Äquivalent können erhebliche Vorteile in Bezug auf finanzielle Kosten, Steuerbarkeit, Nachhaltigkeit und Sicherheit erzielt werden. Die Aufgabe des Bewegungssystems eines Simulators besteht darin, die physikalischen Bewegungsreize so wiederzugeben, wie sie im realen Fahrzeug wahrgenommen würden. Eine ungenaue Bewegungswiedergabe kann zu *Simulatorkrankheit*, unerwünschten Anpassungen im Steuerverhalten des Fahrers und einer verminderten Validität experimenteller Ergebnisse führen. Die Steuerung des Bewegungssystems des Simulators erfolgt durch den *Motion Cueing Algorithmus* (MCA). Dieser wandelt die Bewegung des realen Fahrzeugs in eine Bewegung um, die in den begrenzten Bewegungsbereich des Simulators passt. Dies führt jedoch zwangsläufig zu Imperfektionen, die, wenn sie vom Fahrer wahrgenommen werden, als *Inkongruenzen* bezeichnet werden. Für eine effektive Bewegungssteuerung und eine möglichst gute Simulatorbewegung ist daher ein vollständiges Verständnis der Bewegungssteuerung *und* ihrer Wirkung auf die menschliche Bewegungswahrnehmung erforderlich.

Für eine wachsende Anzahl an Anwendungen von Fahrsimulationen wird eine zunehmende Vielfalt unterschiedlicher MCAs und Simulatoren eingesetzt. Bewegungssysteme reichen von kleinen Hexapoden bis hin zu großen, komplexen Konfigurationen, bestehend aus verschiedenen Bewegungs*sub*systemen (z.B. XY-Antriebe, Tripoden und Gierantriebe). MCAs reichen von „filterbasierten“ Algorithmen mit kausalen Filtern bis hin zu komplexeren, optimierungsbasierten Methoden. Während fortschrittlichere MCAs und Simulatoren eine bessere Bewegungswiedergabe ermöglichen, können sie auch Nachteile wie erhöhte Kosten oder Systemkomplexität mit sich bringen. Welche MCA und Simulator am besten geeignet sind, hängt somit vom jeweiligen Anwendungsfall ab. Zusätzlich erfordert der MCA eine Parametrierung, wodurch sich die Anzahl möglicher Optionen weiter erhöht. Es stellt sich daher die naheliegende Frage, wie bestimmt werden kann, welche *Kombination* aus MCA, Simulator und Parametereinstellung am besten für einen bestimmten Fahrsimulationsfall geeignet ist. Dies erfordert eine Bewertung der Bewegungssteuerungsqualität jeder Kombination über die verschiedenen Dimensionen, die der Begriff „Qualität“ umfassen kann.

In bisherigen Methoden und der Literatur wird die Bewegungssteuerungsqualität meist nur in Bezug auf die Bewegungswiedergabe selbst bewertet. Diese Dissertation

zeigt, dass die Berücksichtigung einer Vielzahl an Qualitäts-, Kosten- und Parametrierungsaspekten, die für die verschiedene Interessenvertreter eines Experiments wichtig sind, ein umfassenderes Bild der Bewegungssteuerungsqualität liefert. Dies kann beispielsweise die Qualität der Bewegung zusammen mit dem Energieverbrauch, den Gesamtkosten und dem erforderlichen Parametrierungsaufwand umfassen. Die Bewertung dieser „totalen" Bewegungssteuerungsqualität kann zu einer fundierteren Wahl des Simulators, MCAs und dessen Parametern für ein Fahrsimulatorexperiment führen. Eine weitere gegenwärtige Einschränkung besteht darin, dass die Parametrierung des MCAs in frühen Phasen der Versuchsentwicklung noch nicht erfolgt ist. Um den für die Parametrierung erforderlichen Kosten- und Zeitaufwand zu rechtfertigen, ist eine Bewertung der Qualität notwendig, diese kann jedoch nicht exakt gemessen werden, *da* die Parametrierung noch nicht erfolgt ist. Eine zentrale Innovation dieser Dissertation ist eine Methode zur *Vorhersage* der *totalen* Bewegungssteuerungsqualität, die ein MCA und ein Simulator *potenziell* erreichen könnten, um den Auswahlprozess in der Fahrsimulationssteuerung zu verbessern.

Die Vorhersage der potenziellen Qualität der Bewegungswiedergabe bringt eine weitere Schwierigkeit mit sich, da ein *Modell* subjektiver Bewertungen (d.h. Einschätzungen) der vom Menschen empfundenen Inkongruenz der Bewegung erforderlich ist. Dies ist aufgrund der Komplexität und Black-Box-Natur der menschlichen Wahrnehmung und des subjektiven Bewertungsverhaltens notorisch schwierig. Aktuelle Arbeiten führten jedoch eine *kontinuierliche* Bewertungsmethode ein, bei der Fahrer die Bewegungsinkongruenz eines simulierten Szenarios fortlaufend über eine Bewertungsschnittstelle beurteilen. Diese Methode ist besonders geeignet zur Entwicklung von Bewertungsmodellen aufgrund der hohen zeitlichen Auflösung der Bewertungen. Wegen der zusätzlichen Belastung, die eine kontinuierliche Bewertung mit sich bringt, welche die Fahraufgabe stören kann, ist die Methode nur für Fahrten mit offenem Regelkreis praktikabel, d.h., der Proband wird gefahren anstatt das Fahrzeug aktiv zu steuern. In den meisten Fahrsimulationsanwendungen muss der Fahrer jedoch aktiv in einer Umgebung mit geschlossenem Regelkreis steuern (z.B. bei Fahrdynamiktests). Obwohl sich die kontinuierliche Bewertungsmethode in vereinfachten Szenarien als vielversprechend erwies, bedarf sie daher einer Validierung dahingehend, ob sie auch zuverlässige Bewertungen liefert, die für Fahrten mit geschlossenem Regelkreis relevant sind.

Anhand fünf neuartiger Experimente und drei bereits existierender Datensätze wird in dieser Dissertation explizit untersucht, wie gut sich die kontinuierliche Bewertungsmethode für realistische Fahrsimulationsanwendungen eignet. Die drei Hauptszenarientypen typischer Fahrsimulationen (Stadt, Autobahn, Landstraße) werden mittels dedizierter kontinuierlicher Bewertungsstudien untersucht und miteinander verglichen, was einen einzigartigen Einblick in die Unterschiede *zwischen* den Szenarien liefert. Schätzungen zur Zuverlässigkeit der Bewertungsdaten zeigen, dass Stadtszenarien zuverlässigere Bewertungen liefern als Autobahn- und Landstraßenszenarien. Dies deutet darauf hin, dass letztere Szenarien schwieriger zu bewerten, und folglich auch schwerer vorherzusagen, sind. Zudem besteht ein inverser Zusammenhang zwischen Bewertung und deren Zuverlässigkeit: Je *mehr* die Bewegung als kongruent bewertet wird, desto *niedriger* ist die Zuverlässigkeit der Bewertung. Während Bewegungskongruenz grundsätzlich

angestrebt wird, sind Bewertungen *inkongruenter* Bewegungen somit nützlicher für die Entwicklung zuverlässiger Bewertungsmodelle.

Auf Basis der kontinuierlichen Bewertungsdaten aus den Experimenten werden anschließend *Modelle* der Bewertungen in Abhängigkeit von objektiven Diskrepanzen der Plattformbewegung (spezifische Kräfte und Rotationsraten) entwickelt. Diese Bewertungsmodelle ermöglichen eine Transformation von objektiven Diskrepanzsignalen in subjektive Qualitätsprognosen, indem sie die vorhergesagte Bewertung eines *durchschnittlichen* Teilnehmers liefern. Die Diskrepanzsignale hängen von der jeweiligen kinematischen Konfiguration ab und werden entsprechend der Konfiguration abgeleitet. Die Analyse in dieser Dissertation zeigt, dass kontinuierliche Bewertungen in Stadtszenarien durch eine tiefpassgefilterte Reaktion auf die objektiven Diskrepanzen in lateralen und longitudinalen spezifischen Kräften modelliert werden können, wobei Gierraten nur einen kleinen, situationsabhängigen Beitrag liefern. Die Bewertungsdaten der Landstraße und Autobahnszenarien werden zur Verbesserung der Modellvalidität verwendet. Obwohl die Gewichtung der Diskrepanzkanäle szenarienspezifisch variiert, zeigen sich in allen drei Szenarientypen dieselben dominanten Bewegungskanäle, was auf eine hohe Generalisierbarkeit und Validität des Modells *szenarienübergreifend* hinweist.

Kontinuierliche Bewertungen sind zwar sehr nützlich für die Modellentwicklung, aber weniger für direkte Vergleiche der MCA- und Simulatorqualität geeignet. Vergleichende Bewertungen zwischen Konfigurationen lassen sich einfacher anhand von Einzelwerten (Skalarwerten) durchführen als mit langen Zeitreihen kontinuierlicher Bewertungen. Diese Dissertation untersucht daher auch die Vorhersage von Gesamtbewertungen (eine Bewertung für die gesamte Fahrt) sowie von manöverbasierten Bewertungen (eine Bewertung pro Manöver). Die in dieser Arbeit beschriebenen Experimente, in denen kontinuierliche, Gesamt- und Manöverbewertungen parallel erhoben werden, zeigen, dass die am stärksten inkongruente Bewegung (gemessen über die kontinuierliche Bewertung) stark mit der Gesamt- bzw. Manöverbewertung korreliert. In Kombination mit dem kontinuierlichen Bewertungsmodell ermöglicht dies auch die Vorhersage von Gesamt- und Manöverbewertungen aus objektiven Diskrepanzsignalen. Parametrierungsprozesse können davon profitieren, indem stets der inkongruenteste Moment einer Fahrt als Zielgröße für die Bewegungsoptimierung verwendet wird.

Die meisten in dieser Dissertation vorgestellten Erkenntnisse betreffen die Bewertung und Vorhersage subjektiver Einschätzungen im Fahren mit offenem Regelkreis, welches jedoch nicht notwendigerweise das typische Fahren mit geschlossenem Regelkreis vieler Simulatoranwendungen repräsentiert. Ein dediziertes Experiment in dieser Arbeit zeigt jedoch, dass subjektive Bewertungen im Fahren mit offenem und geschlossenem Regelkreis in urbanen Szenarien äquivalent sind. Dies bestätigt die Hypothese, dass kontinuierliche Bewertungen auch für Fahrten mit geschlossenem Regelkreis repräsentativ sein können. Kombiniert ermöglicht dies einen Ansatz, bei dem ein kontinuierliches Bewertungsmodell verwendet wird, um Gesamt- und Manöverbewertungen von Fahrten mit geschlossenem Regelkreis vorherzusagen, basierend auf objektiven Diskrepanzsignalen. Dadurch werden bisher unerreichte Prognosen der potenziellen subjektiven Bewertungen in Simulatorversuchen möglich. Zudem kann eine Fahrt mit geschlossenem

Regelkreis aufgrund individueller Fahrvariabilität nie exakt reproduziert werden. Da das Fahren mit offenem Regelkreis das Abspielen zuvor aufgezeichneter Fahrten erlaubt, bietet es erhebliche Vorteile hinsichtlich Reproduzierbarkeit und Steuerbarkeit der erlebten Diskrepanzen, was Tests mit offenem Regelkreis besonders praktikabel zur Beschleunigung zukünftiger Parametrierungen macht.

Ingenieurtechnische Anwendungen von Bewegungssteuerung in Fahrsimulatoren können direkt von den vorgestellten Erkenntnissen aus dieser Dissertation profitieren. Mit einem validierten Bewertungsmodell auf Basis zuverlässiger Daten können präzise Vorhersagen über die potenzielle Bewegungsqualität gemacht werden. Dazu ist auch eine Methode zur Bestimmung der potenziellen Bewegungssteuerungsqualität erforderlich. Der abschließende Beitrag dieser Dissertation ist daher eine Methode zur Schätzung der potenziellen Bewegungssteuerungsqualität, wobei die Vorhersagemodelle subjektiver Bewertungen eine zentrale Rolle spielen. Diese Schätzung basiert auf der Extrapolation von existierenden Diskrepanzsignalen zur vollen Ausnutzung der Bewegungsräume verfügbarer Simulatoren. Für jede Kombination aus MCA und Simulator kann, und sollte, eine potenzielle und totale Qualität vorhergesagt werden, die das tatsächliche Bewegungspotenzial widerspiegelt. Auch wenn diese Methode noch nicht zur direkten quantitativen Abwägung ausgearbeitet ist, zeigen die betrachteten Anwendungsfälle, dass die Berücksichtigung sowohl der potenziellen als auch der totalen Qualität die Wahl von MCA und Simulator wesentlich beeinflussen kann.

Zukünftige Forschung kann direkt von den Ergebnissen dieser Dissertation profitieren. MCAs können gezielt auf die subjektive menschliche Wahrnehmung abgestimmt werden, um ein verbessertes Bewegungserlebnis zu ermöglichen. So kann z.B. das lineare Bewertungsmodell direkt in das Optimierungsschema eines MCAs integriert werden, um explizit auf subjektive Qualität zu optimieren. Alternativ können die Gewichtungen der Bewegungskanäle so angepasst werden, dass sie den im Modell identifizierten Wichtigkeiten entsprechen. Schließlich kann das Wissen über Gesamt- und Manöverbewertungen dazu verwendet werden, besonders inkongruente Punkte stärker zu bestrafen bzw. zu optimieren, um die *gesamt*e Bewegungsqualität zu verbessern. Diese Beiträge können alle zu einem besseren subjektiven Erlebnis in Fahrsimulationsstudien führen.

Diese Dissertation bietet sowohl praktische Leitlinien als auch grundlegende Methoden zur Verbesserung der Bewegungssteuerung zukünftiger Fahrsimulationsversuche. Durch die Einführung eines neuen Blickwinkels auf die Bewegungssteuerungsqualität wird das lang bestehende und komplexe Problem der Bewegungssimulation besser verstanden. Neben ihrer direkten Relevanz für jede Art von Fahrsimulator sind die Erkenntnisse auch auf andere Forschungs- und Technikbereiche mit bewegungsbasierten Simulatoren übertragbar, wie z.B. Flug- oder Schiffssimulationen.

Nomenclature

Acronyms

ADAS Advanced Driver Assistance System

AIC Akaike's Information Criterium

ANOVA Analysis of Variance

ARX AutoRegressive eXogenous

AWA Adaptive Washout Algorithm

CID Central Information Display

CRG Curved Regular Grid

CRP Cueing Reference Point

CWA Classical Washout Algorithm

DoF Degree of Freedom

FR Forward Regression

FSZ Fahrsimulationszentrum

HMI Human-Machine Interaction

ICC Intraclass Correlation Coefficient

IMU Inertial Measurement Unit

MCA Motion Cueing Algorithm

MCS Motion Control System

MIR Motion Incongruence Rating

MISC Motion Illness Symptoms Classification

MISO Multiple-Input-Single-Output

MPC Model-Predictive Control

MRP Motion Reference Point

MS Motion System

MSHQ Motion Sickness History Questionnaire

MTP Model Transfer Parameter

OR Overall Rating

OWA Optimal Washout Algorithm

PH Post-Hoc

PMI Perceived Motion Incongruence

PS Perceptual System

RMSE Root Mean Square Error

RS Response System

SIS Simulator Inertial System

SPH Section-Wise Post-Hoc

SSQ Simulator Sickness Questionnaire

TIS Tripod Inertial System

VAF Variance-Accounted-For

VIS Vehicle Inertial System

VRP Vehicle Reference Point

TOST Two One-Sided Tests

1 Introduction

AUTOMOTIVE companies currently face several challenges. Global warming requires a rapid change towards emission-free vehicles. The surge in automation leads to the development of partially or fully autonomous vehicles. Strong urbanization, where it is expected that 68% of the global population will live in urban areas by 2050 [United Nations, 2018], will change the way in which vehicles are owned or shared, controlled, and communicate. As a result, fundamental changes have to be made in vehicle design, requiring rapid systematic testing abilities for car manufacturers.

1.1 The Need for Driving Simulation

Driving simulation is an essential tool that provides the ability to contribute towards achieving all of these changes. Here, dedicated hardware is used to create a virtual environment that resembles a real-life driving situation. Through simulator experiments, research and development can be performed simulation-based, rather than through real-life testing, providing significant benefits in terms of cost, safety, and practicality. Even early prototype or hypothetical vehicles can be tested in any (virtual) situation [Freeman et al., 1995], without being restricted to dedicated test tracks. Its high controllability compared to real-life vehicle testing furthermore allows for specific conditions to be repeated multiple times in exactly the same way [de Winter et al., 2012].

In the real world, drivers mainly rely on a combination of visual [Sivak, 1996], auditory [Zhao et al., 2024], and motion cues [Markkula et al., 2019] to operate the vehicle. Whereas some simulators only reproduce and provide the visual and auditory cues to the driver, others have the additional ability to generate motion cues by physically moving the simulator cabin through inertial space. These are referred to as *motion-base* simulators. This inertial motion provides the driver with the vestibular [Markkula et al., 2019], somatosensory [Bruschetta et al., 2021], and proprioceptive [Hlavačka et al., 1992] cues of the simulation. In most cases, all three translational motions (x, y, and z), as well as the three rotations (roll, pitch, and yaw), can be manipulated, resulting in six motion axes.

Due to the constrained excitation of the motion systems, a simulator only has a limited workspace in which it can move. This workspace is typically smaller than is required to fully reproduce the motion of a real vehicle [Ellensohn, 2020]. Therefore, even with

the largest motion-base simulators, the full vehicle motion can often not be fully reproduced. Especially the motion sensations one would experience when performing large excitations, such as sustained cornering, are difficult to fully reproduce, as the simulator can only accelerate for a limited period of time inside its workspace. This is problematic, as inaccurate motion reproduction can decrease the realism of the simulation [Cleij et al., 2018], affect the task performance and control behavior [Romano et al., 2019], and lead to participants getting sick [Himmels et al., 2022b], which all negatively affect the outcomes of experiments and prototype tests. An accurate motion reproduction is thus paramount for a successful driving simulator experiment.

1.2 Driving Simulation Use-cases

The contents of an experiment are described by its *use-case*. This includes the (type of) simulated vehicle, the scenario (e.g., urban, rural, or highway scenarios [Chanmas et al., 2023]), and the goal of the experiment, including the task participants have to perform. Classical examples of use-cases include the evaluation of driving dynamics of new vehicles [Brems et al., 2015] and tire testing [Baldoni et al., 2011], as well as the development and use of Human-Machine Interaction (HMI) systems (e.g., navigation and entertainment systems [Jeong et al., 2013]). In an attempt to move much of the current real-life testing to the simulated world, the number of use-cases to be handled in simulation has further increased in recent years [Bruck et al., 2021]. Specifically for the recent developments in automated driving and Advanced Driver Assistance System (ADAS) technologies, their potential effect on the driver's behavior is a prime focus in studies [Rossi et al., 2020]. This includes the effect of drivers' fatigue [Schmidt et al., 2016; Xu et al., 2018], distraction levels [Ezzati Amini et al., 2023; Wijayaratna et al., 2019], and shared control between the vehicle and its user [Terken and Pfleging, 2020]. Furthermore, driving simulators are widely used in comparative comfort evaluations, such as to evaluate different driving strategies for electric vehicles [Xue et al., 2024].

1.3 Motion Systems

Given the large variety of driving simulation use-cases, the required motion can be different as well. For example, a use-case investigating braking behavior of drivers will require the ability to reproduce strong longitudinal motion. In contrast, urban environments involving many tight corners require a combination of lateral and yaw motion [Ellensohn, 2020]. It can therefore depend on the use-case which motion-base simulator is best suited, and whether a motion-base is needed in the first place, as specific qualities of a given simulator might work well in combination with a certain use-case. If a large variety of use-cases is to be tested, a range of simulators might be required, such as at the *Fahrsimulationszentrum (FSZ)* (English: Driving Simulation Center) at BMW Group, see Figure 1.1. It operates a total of ten motion-base simulators and several static ones, of which a selection is shown in Figure 1.2. In general, prominent differences can be found in the dimensions and kinematic configuration of motion-base simulators [Mohajer et al., 2015], and thus in their capacity to reproduce motion, as is also visible in Figure 1.2.

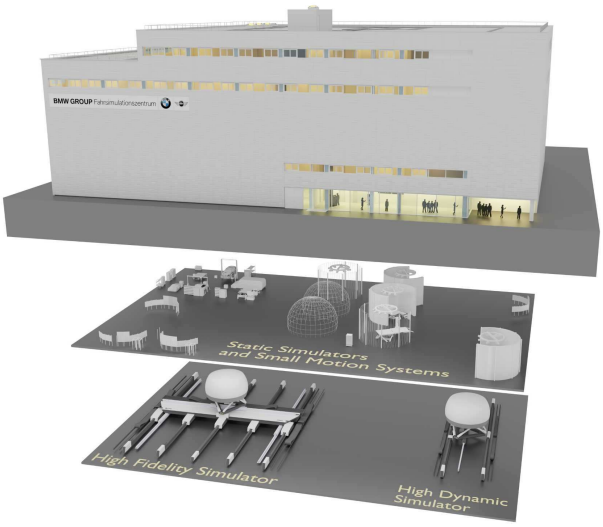


Figure 1.1: BMW's driving simulation center in Munich, Germany. Image courtesy: BMW Group [2018].

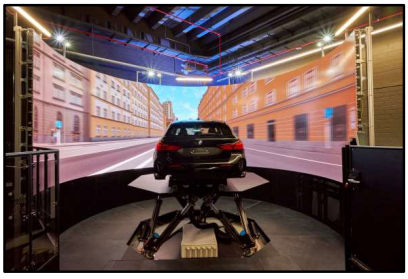
1.3.1 Hexapod

The most common simulator motion system is the hexapod, a six-legged configuration that is able to drive six *Degrees of Freedom (DoFs)*. The hexapod exists since the 1960s and is often used in flight simulation [Huang et al., 2016b]. It is otherwise known as a *Stewart* [Stewart, 1965] or *Stewart-Gough* [Gough, 1962] platform, named after its first inventors. An example for flight simulation is TU Delft's SIMONA Research Simulator [Stroosma et al., 2003] and, for driving simulation, BMW's Vega Vector (Figure 1.2c). The six DoFs of the hexapod allow the manipulation of the three translational and three rotational axes, thus enabling it to move and generate motion similar to a real vehicle's motion, even with a relatively small system. The system is controlled by its six actuators that act in parallel configuration to each other [Liu et al., 1993], i.e., they operate together to move a single platform, resulting in an inherently rigid and simple design [Huang et al., 2016a].

The hexapod, however, also has several disadvantages. First, the six DoFs of the simulator are coupled, such that moving the simulator in one DoF reduces the available workspace of the other DoFs. This is problematic especially in use-cases that require simultaneous motion in multiple axes. Tight cornering in urban simulations, for example, requires relatively high excitations in both the lateral and the yaw direction [Ellensohn, 2020]. Second, hexapods inherently have a limited motion space, which means that sustained accelerations cannot be reproduced. A common solution is tilt-coordination, where sustained motion is generated through platform tilt [Conrad et al., 1973; Stratulat et al., 2011]. Through the roll angle φ and pitch angle θ, part of the gravity vector is perceived by humans as translational motion, as long as the rotational motion due to the tilting remains below the human perceptual threshold. In flight simulation, a combination of translational and rotational hexapod motion can be used to reproduce the relatively slow translational motion of an aircraft. Driving, however, induces more dy-

(a) Green Stage (static).

(b) Ruby Space (tripod, hexapod).



(c) Vega Vector (hexapod).

(d) Sirius Vector (yaw-drive, hexapod).

(e) Diamond Space (x- or y-drive, hexapod).

(f) Sapphire Space (xy-drive, hexapod, yaw-drive).

Figure 1.2: Four simulators of different kinematic configurations operated as part of the driving simulation center (*FSZ*). Image courtesy: BMW Group.

namic and asymmetric motion [Ellensohn, 2020], which makes the application of tilt-coordination less suitable. This results in a larger dependence on simulator translational motion, which the hexapod can only produce to a limited extent.

1.3.2 xy-drive

To increase the translational workspace, the hexapod can be extended with additional motion subsystems. An example is adding (a) linear actuator(s) below the hexapod, extending the translational workspace. Examples are BMW's Diamond Space (Figure 1.2e), IKA's Hochdynamischer Fahrsimulator [Wagener et al., 2023], and the Daimler-Benz Driving Simulator [Cleij et al., 2019]. Some simulators have a two-DoF linear actuator system, i.e., an *xy-drive*, which allows the cabin to move horizontally in both the x- and y-direction, resulting in a total of eight DoFs. Examples are the FKFS's Stuttgarter Fahrsimulator [Baumann et al., 2014], Renault's ULTIMATE [Fang et al., 2017], and VTI's Sim IV [Jansson et al., 2014]. The benefit of an xy-drive is that it can greatly increase the translational workspace of the simulator. For example, the Diamond Space can move over a rail of nearly twenty meters. A downside lies in the required physical space and high cost.

1.3.3 Yaw-drive

Another system used to extend the motion workspace is the yaw-drive, a system which can rotate the simulator to generate additional yaw motion. This motion is required to reproduce the turning motion of the car. An inaccurate yaw motion reproduction negatively affects the control behavior [Lakerveld et al., 2016] and can increase the subjective realism [Kusachov et al., 2015], although the latter is disputed [Hogema et al., 2012].

Because a yaw-drive is a separate motion subsystem and uncoupled from the other hexapod, it can typically move close to the full 360°, which is valuable for, e.g., sustained corners and roundabouts. Yaw drives can be placed on top of the hexapod structure [Fang et al., 2017] or below, such as BMW's Sirius Vector (Figure 1.2d). Combining an xy-drive, hexapod, and a yaw-drive, yields a nine-DoF system. Examples are BMW's Sapphire Space (Figure 1.2f), Renault's ROADS [Fang et al., 2022], and Toyota's Driving Simulator [Nguyen Van and Ito, 2016].

1.4 Motion Cueing Algorithms

Regardless of the simulator's dimensions and configuration, a one-to-one inertial motion reproduction is often impossible due to workspace restrictions [Ellensohn, 2020]. Therefore, a Motion Cueing Algorithm (MCA), which converts the vehicle motion into motion that fits in the workspace of the simulator, is required. An MCA takes the output of a simulated vehicle model, typically specific forces and rotational rates, and converts these to the commanded simulator platform movement. The challenge of choosing the right motion cueing is widely applicable, also if only one simulator system is available. Several MCAs exist, each with their own (dis)advantages.

1.4.1 Classical Washout Algorithm

The Classical Washout Algorithm (CWA) is an MCA that uses linear scaling and filters between the vehicle and commanded simulator movement [Conrad et al., 1973; Schmidt and Conrad, 1970] and is therefore a "filter-based algorithm". A schematic representation of a CWA for a nine-DoF simulator is shown in Figure 1.3. It illustrates the various typical steps from the simulated vehicle motion (specific forces $\mathbf{f}$ and rotational rates $\boldsymbol{\omega}$) to the hexapod position $\mathbf{p}_h$, xy-drive position $\mathbf{p}_d$, hexapod rotation angles $\boldsymbol{\beta}_h$, and the yaw-drive rotation angle $\boldsymbol{\psi}_d$.

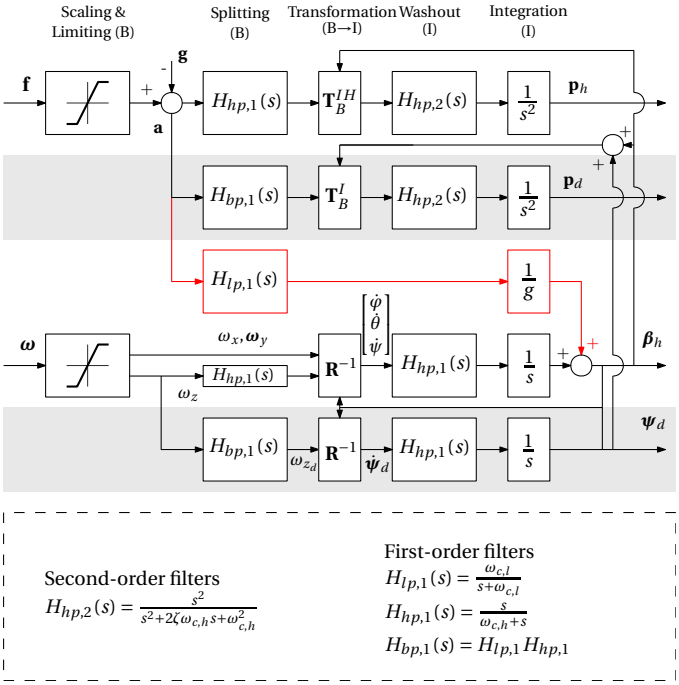


Figure 1.3: Block diagram of a Classical Washout Algorithm (CWA) a nine-Degree of Freedom (DoF) simulator. The top grey bar represents the xy-drive channel, the bottom grey bar the yaw-drive channel.

First, the magnitude of the motion is reduced by scaling and limiting. Scaling is a linear operation through a gain, such that all amplitudes are decreased by the same factor. In limiting, values above a selected threshold are attenuated [Reid and Nahon, 1985]. Second, the motion is split into low- and high-pass components, for both the specific forces and rotational rates, using the body-frame filters $H_{lp,1}(s)$ and $H_{hp,1}(s)$, respectively. Typically, only the high-frequency components are reproduced by the translational and rotational channels of the simulator. The sustained low-frequency components are then reproduced using tilt-coordination [Conrad et al., 1973; Stratulat et al., 2011]. In case an xy-drive is present, an additional channel with medium-frequency components (through a band-pass filter) can be used [Fang et al., 2017]. Similarly, a band-pass filter $H_{bp,1}(s)$ can be used to determine the yaw-drive motion. The channels for both motion subsystems are indicated by the grey bars in Figure 1.3.

Third, the motion of the six channels is high-pass filtered in the inertial frame, such that the simulator is always brought back to its neutral state ('washout'), preferably without the human subject noticing. This is required as the simulator might move close to a workspace limit when the next maneuver commences, which might result in reaching the limits. The washout is indicated in Figure 1.3 by the filters $H_{hp,2}(s)$ (for translational motion) and $H_{hp,1}(s)$ (for rotational motion). To washout the motion under a constant acceleration, the former washout component requires a second-order filter [Reid and Nahon, 1985]. To ensure that the simulator is brought back to its neutral state, also when the cabin is rotated, the filtering must be performed in the simulator's inertial frame. Therefore, the motion is first transformed using the matrices $\mathbf{T}_B^{IH}$ (hexapod translational motion), $\mathbf{T}_B^I$ (xy-drive), and $\mathbf{R}^{-1}$ (rotational motion). The content of these matrices depend on the kinematic configuration of the simulator, whereas its values depend on the simulator's attitude, as will be discussed in depth in Chapter 3.

1.4.1.1 Strengths

The CWA remains a popular choice with its relatively simple structure [Asadi et al., 2015], ensuring a high degree of transparency during the tuning process. The simulation of the linear transfer functions and lack of optimization furthermore gives the benefit of a low computational load and high stability [Qazani et al., 2024]. Finally, as the CWA consists of (linear) transfer functions, it is deterministic, such that a given input always yields the same output [Jamson, 2010]. This adds to the transparency of the algorithm and makes it easier for drivers to adapt to the motion.

1.4.1.2 Drawbacks

The CWA also has several inherent weaknesses, however. Despite its transparency, tuning the CWA can be a labor-intensive process, due to a variety of factors. First, the CWA has a large number of parameters that need to be tuned. This scales with the number of DoFs, as each DoF requires several parameters, such as the scaling factors and cut-off frequencies. For example, if each DoF has a gain, a split frequency, and two washout parameters, a total of 36 parameters need to be tuned for a nine-DoF system, compared to 24 for a hexapod. Another issue is that the tuning procedure is a largely heuristic procedure, of which the outputs depend on the tuning procedure, such that there is no guaranteed consistency in stimuli between virtual vehicle and simulator motion [Beghi et al., 2012]. In practice, this also implies that the tuning often requires subjective testing and must therefore be performed on-site by experts, a time-intensive and costly procedure.

Apart from the tuning procedure, other inherent drawbacks exist, of which a comprehensive overview can be found in Beghi et al. [2012]. Inherent for the CWA is the lack of state feedback, as the simulator state is not compared to the constraints of the platform. As a result, hard workspace constraints cannot be accounted for, but must be accounted for in the tuning process. As the algorithm uses linear filters, the parameters of the washout filters must be tuned considering the worst-case scenario, such that the simulator just remains in its workspace for even the most extreme maneuver [Aminzadeh et al., 2012]. Often this leads to conservatively-tuned washout parameters for regular maneuvers, decreasing the motion magnitude over the whole drive. Finally, as low-pass and high-pass filters produce lag and lead, respectively, there is an inherent phase shift of the simulator motion with respect to the vehicle reference motion. Espe-

cially when both high- and low-frequency motions are present in a maneuver (e.g., a roundabout), this can create artifacts in the motion. These cannot be corrected for, as the CWA only consists of linear filters, and can thus not consider future states.

1.4.2 Variations on the CWA

1.4.2.1 Adaptive Washout Algorithm

The Adaptive Washout Algorithm (AWA) [Ariel and Sivan, 1984; Parrish et al., 1975] aims to solve two drawbacks of the CWA: (1) the lack of simulator state feedback, and (2) worst-case tuning. In the AWA, the parameters of the filters can be changed in real-time during the simulation. The MCA still relies on filters that washout the motion towards the neutral state of the simulator, but it can change its filter parameters depending on the workspace position of the simulator. Its cost function is typically minimized for (i) the current simulator motion state, i.e., excitations from the simulator neutral position, (ii) errors between the vehicle and simulation, and (iii) differences between adaptive parameters and their nominal values.

AWA potentially increases the quality of the motion [Aykent et al., 2011] compared to CWA and has been successfully implemented in driving simulation [Aykent et al., 2011; Nehaoua et al., 2005]. The relative weights between these components might be more intuitive to tune [Garrett and Best, 2010], but also add more parameters to be tuned. Another drawback is that AWAs typically suffer from stability issues [Asadi et al., 2019].

1.4.2.2 Optimal Washout Algorithm

To mitigate the problem of complex and time-consuming tuning of CWA and AWA in a heuristic, mostly trial and error approach, the Optimal Washout Algorithm (OWA) has been developed [Sivan et al., 1982; Telban et al., 2005]. In this method, the MCA controlling the simulator is typically still a CWA. However, the filter parameters (i.e., gains and cut-off frequencies) are chosen not through a manual tuning process, but through an optimization problem beforehand, i.e., offline, using a cost function, that incorporates the human perception through a vestibular model. The main benefit of an OWA is that, through the inclusion of a vestibular model, a better consistency between the stimuli can be achieved. Furthermore, its tuning process can arguably be easier than the tuning process of a CWA, as only the balance between the motion cue reproduction and the workspace adherence must be selected.

Its structure also gives two inherent disadvantages, however. First, as its functionality relies on a model of the vestibular system, its advantage over CWA is limited due to the lack of proper vestibular models with visual integration developed specifically for driving simulation [Jamson, 2010]. Second, in the CWA, a worst-case tuning must still be employed [Aminzadeh et al., 2012]. An attempt at solving this issue employs adaptive optimal control [Naseri and Grant, 2005; Telban et al., 2005] in the OWA, which allows the parameters of the filters to be varied throughout the drive. However, as mentioned by Aminzadeh et al. [2012], this algorithm still suffers from the same disadvantages as the AWA, such as a high computational load and stability issues.

1.4.3 Model-Predictive Control

The previously discussed drawbacks of filter-based cueing algorithms have resulted in a push towards exploring other MCA solutions. The MCA architecture that has received the most focus in driving simulation is Model-Predictive Control (MPC) [Dagdelen et al., 2004; Dagdelen et al., 2009]. MPC works by minimizing a cost function at each time step of the simulation in real-time. Differences between reference and simulator motion are minimized with the goal of providing accurate motion. To explicitly include future states, a model of the simulator's motion system is used to find the sequence of inputs that leads to the best control tracking over a given prediction horizon, typically with the smallest inputs required. Excursions of the simulator state with respect to its equilibrium condition can also be penalized, ensuring that the simulator is brought back to its neutral position, similar to the washout characteristic in filter-based algorithms. This term is also often added to ensure convexity of the cost function [Katliar, 2020]. Furthermore, MPC can incorporate explicit platform constraints in its optimization, which guarantees that the hardware limits are not reached during the simulation [Beghi et al., 2012], even on the actuator level [Ellensohn et al., 2019b; Garrett and Best, 2013; Khusro et al., 2020]. As hard constraints can lead to infeasible solutions [Rengifo et al., 2019], applying soft constraints [Fang et al., 2017], which penalize nearing the workspace limits by increasing the cost function value, have become the standard.

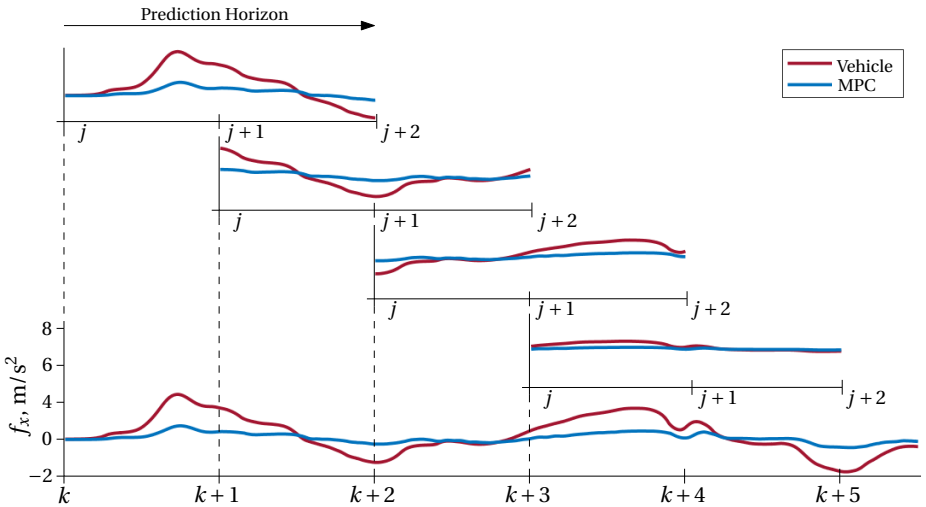


Figure 1.4: The optimization process of a basic Model-Predictive Control (MPC) algorithm, here shown for the longitudinal specific force. At each time step k in the simulation, an optimization with a limited prediction horizon is solved (smaller segments at the top).

The working principle of the MPC's primary function, the reference tracking, is shown in Figure 1.4, where the vehicle is currently at point k. The reference tracking term of the MPC is optimized over several points J in the future, up to the *prediction horizon* (here: 2 steps). Thus, MPC calculates the sequence of future inputs that brings the MPC closest to the reference over the prediction horizon. However, the eventual

future states of the simulator cannot be known exactly beforehand, as they depend on the inputs of the driver. Thus, by the time the simulation has reached point $k+1$, the reference signal can have changed, such that the optimization output of point k is not necessarily accurate anymore. For this reason, the optimization procedure is performed again at point $k+1$, of which the output of that reference tracking optimization is now taken. As the actual future states can thus deviate from the predicted future states, only the first element of the sequence of inputs (i.e., the next time steps) is effectively used at a given time-step.

1.4.3.1 Strengths

Unique for MPC is that it can incorporate future states in the cost function. The further the MPC looks ahead and the better the quality of the prediction, the better it can potentially anticipate vehicle motion (of upcoming maneuvers) and system constraints (such as approaching the simulator workspace limits). Larger prediction horizons generally lead to an improvement in motion cueing quality [Katliar et al., 2015], but also increase the complexity of the system, hampering the real-time performance.

Because of its optimization-based nature, it also attempts to find the optimal solution for that given simulator state without explicit manual tuning beforehand. Although the components of the cost function can be weighted relatively to each other, the tuning process is intuitive and not necessary to ensure good motion cueing quality. Furthermore, MPC can use the full motion workspace in every separate maneuver, in contrast with the worse-case tuning of the CWA, resulting in overall better motion cueing than is possible with filter-based algorithms [Cleij et al., 2019].

1.4.3.2 Drawbacks

Even though MPC generally provides a higher motion cueing quality than filter-based algorithms, the former have not fully replaced the latter. Arguably, there are three main issues with MPC. First, the real-time requirement puts stringent requirements on the calculation time, as the optimization procedure must be repeated at each time step. It is common practice to have an MCA run at 10 ms at the slowest. This often restricts the length of the prediction horizon and the complexity of the system [Munir et al., 2017]. Second, both the complexity and the calculation time also scale with the DoFs, which makes the algorithm harder to use for more advanced driving simulators. Third, the algorithm has a lower stability than a CWA due to its optimization nature, such that infeasibilities can occur and need to be actively avoided [Fang and Kemeny, 2012].

1.5 Selecting the Motion Cueing

1.5.1 Motion Cueing Quality

To make well-informed decisions in selecting the best possible simulator and MCA for each use-case, being able to better *predict* and systematically compare how available options perform relative to each other is paramount. An important part of expressing the quality of the motion cueing is the quality of the reproduced motion itself, sometimes also referred to as *fidelity* [Pool, 2012]. Mismatches between the reference vehicle and the simulator can reduce the quality/fidelity of the simulation. Since motion mis-

matches are often inevitable in driving simulations, an important question is which of the mismatches are the most crucial to avoid.

In *behavioral fidelity/quality*, the interest lies in replicating the behavior of the driver. Mainly in the aviation domain, where training for real-life situations is the main goal, the focus has been on behavioral fidelity [Pool, 2012]. Here, the most important goal of a simulation is to have pilots behave in the same or similar manner as they would do in its real aircraft equivalent. This even allows or promotes discrepancies between the aircraft and simulator, as long as the induced behavior is consistent.

In driving simulation, the goal is often to make the simulation feel as realistic as possible. In this case, achieving *perceptual fidelity/quality* is the main goal. This increases the sense of presence [Himmels et al., 2022a] and reduces simulator sickness [Himmels et al., 2022b], which both have a positive effect on the outcome of experiments. Thus, the importance lies not in the behavior of the driver, but in the sensory perception of the motion. Smaller differences mean a higher fidelity, and vice versa. Perceptual fidelity can be evaluated through both subjective and objective methods [Casas et al., 2015].

1.5.2 Quality, Cost, and Tuning Metrics

To save costs and reduce unnecessary complexity, each experiment is ideally performed on the smallest and cheapest simulator that still meets the use-case's requirements. Only use-cases that require large-excitation motion in all axes (e.g., urban driving) might justify the use of the largest simulators or a complex MCA, even if these options induce higher costs. Depending on the use-case, low-quality motion could be perfectly acceptable and the significant time and money investment in providing high-quality motion would not be preferred. It thus seems a natural consideration that apart from the motion cueing quality, other metrics affect the simulator and MCA selection process as well. Nahon and Reid [1990] were the first to point out that several other metrics regarding the tuning complexity can impact the choice of a simulator and MCA as well.

The reality is that the objective analysis and evaluation of the combination of use-cases, simulators, and MCAs is commonly exclusively presented in terms of the quality of the motion [Biemelt et al., 2021; Cleij et al., 2019; Lamprecht et al., 2019]. From the three categories of metrics (quality, cost, and tuning) that potentially describe the "total" quality, it is furthermore an open question *which* metrics would be important to consider. This selection of metrics might furthermore vary between experiments, depending in its goal and the resources available. Identifying which of such metrics matter for each experiment is thus an important task, and an experiment's total quality should hence drive the selection of the MCA and simulator. The first research goal to be addressed in this dissertation is therefore stated as:

Research Goal 1

To develop a framework to evaluate the total motion cueing quality of an experiment, explicitly including its relevant quality, cost, and tuning metrics.

1.5.3 Subjective Motion Cueing Quality

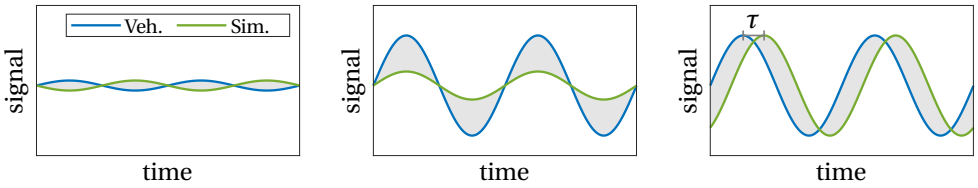
To express the quality of the motion reproduction itself, subjective evaluations are the standard. Here, the motion cueing quality is evaluated by a test driver, who then gives their subjective opinion and/or a rating. When a test driver perceives a mismatch between the perceived and the motion they would expect to feel from the real vehicle, the motion is known as *incongruent* [Cleij et al., 2018]. As perceptual fidelity describes that the coherence of the stimulus (the ground truth) matters most, subjective ratings aim to describe this ground truth, for example by using a rating scale or survey containing questions on the subjective opinion of the test driver. This implies that the obtained subjective evaluations are merely a *representation* of this ground truth, and can be biased through several effects, such as through interaction with the experiment leader, known as the Pygmalion effect [Rosenthal and Jacobson, 1968].

There are several methods to acquire such subjective ratings. One method is to ask the driver several questions regarding different aspects (cornering, accelerating, braking, etc.) of the motion [Valente Pais et al., 2009; Zhao et al., 2024]. It is also possible to let participants rate all aspects of the motion using a single rating, for example, directly after the drive [Biemelt et al., 2021; Rengifo et al., 2021]. The test drivers evaluate the motion of the drive *after* having experienced the drive, such that the rating is known as *post-hoc*. Although easy to implement, as this method does not interfere with the driving task(s), the information obtained from a single post-hoc rating is limited: for instance, the effects of particular maneuvers and/or aspects cannot be derived. Alternatively, it is possible to divide a drive into smaller segments that are driven and rated separately [Cleij et al., 2018], or by extracting ratings for each maneuver [Ellensohn et al., 2020].

By far the most detail-rich extraction method are continuous ratings [Cleij et al., 2018], in which drivers continuously evaluate the motion throughout the whole drive using a rating interface. Due to the high workload of this method, it is only possible in *open-loop* driving, i.e., where test drivers are driven around passively rather than driving themselves. Due to its relative novelty, however, much is still unknown about the value of continuous ratings. For example, although the ratings are obtained continuously, it is not necessarily the case that the obtained ratings accurately represent the continuous subjective rating at each point in time. Rather, it could be the case that participants effectively only rate specific maneuvers, while not actively rating other segments. Another challenge lies in the fact that the continuous rating method can only be used in open-loop driving. This might alter the perception of motion [Nesti et al., 2016; Valente Pais et al., 2012] compared to closed-loop driving, such that it is currently unknown whether the obtained ratings are also representative for closed-loop driving experiments.

1.5.4 Objective Motion Cueing Quality

In the selection and tuning of simulators and MCAs, subjective ratings are generally an expensive method, as they require extensive simulator testing with a large number of participants. Especially when a larger number of configurations is at hand (simulators and/or MCA configurations), it rapidly becomes impractical to test each possible configuration. Even subjectively testing a selection of these configurations is generally considered time-inefficient and costly. Ideally, the human as the evaluator can be replaced by *objective* measures, mathematical metrics of the motion cueing quality based



(a) Case 1: good absolute difference, poor cross-correlation.

(b) Case 2: poor absolute difference, good cross-correlation.

(c) Case 3: Time delay of τ seconds.

Figure 1.5: Absolute difference, cross-correlation, and time delay objective metrics between the vehicle and simulator for a synthetic signal. The shaded area represents the absolute difference between the signals.

on the difference between the vehicle and simulator motion. This enables testing and development without on-site testing using simulator and human subjects, resulting in a faster and a lower cost tuning and development [Fischer and Werneke, 2008; Qaisi and Trächtler, 2012] and allow for a more systematic evaluation.

Examples of objective measures are absolute difference, cross-correlation, and time delay metrics [Grottoli et al., 2019], as shown in Figure 1.5. Note that objective measures do not necessarily imply the use of objective *data*, such as in behavioral fidelity analysis of control behavior [Pool, 2012]. For objective measures, several qualities exist [Casas-Yrurzum et al., 2020], of which the key challenge is *validity*: finding objective measures that are valid for describing quality that can be validated through subjective measures.

Objective metrics are also extremely useful for the development of new MCAs that are in early development stages and cannot be tested at the simulator yet. Ideally, the objective measures can be used in an (semi-)automatic tuning process, where the values of MCA design parameters are selected by optimizing the output of the objective metric(s). For optimization-based algorithms, such as MPC, such objective measures are also directly used in the cost function [Cleij et al., 2020].

1.5.5 Predicting Subjective Motion Cueing Quality

Objective measures of absolute difference have been successfully used to predict continuous subjective ratings [Cleij et al., 2018]. Such continuous rating models represent one of the few options for bridging the goal between the objective and subjective measures. If based on valid and reliable rating measurements, such models might combine the best of both worlds by allowing for making unparalleled predictive judgments on the motion cueing quality *without* the need for expensive on-site simulator testing.

Although the models of Cleij et al. [2018] might be of direct use, no predictive models for realistic driving simulation use-cases (e.g., urban, rural, and highway scenarios) have been developed so far. The assumption on the equivalence between open-loop and closed-loop driving also remains untested. Furthermore, questions regarding continuous rating models remain regarding their systematic value, as it is currently unknown whether such models accurately predict ratings *between* experiments. The second research goal of this dissertation is therefore stated as:

Research Goal 2

To objectively and systematically predict subjective motion incongruence ratings of closed-loop driving simulator experiments in urban, rural, and highway driving scenarios.

1.5.6 Motion Cueing Method Selection

Ultimately, the goal for any driving simulation center is to provide the best-suited driving simulation experience among all experiments. Arguably, the best possible motion cueing is only obtained when the best combination MCA and simulator (i.e., the "motion cueing method") is identified and selected. The inter-dependency of the use-case, simulator, and MCA opens up a multi-dimensional selection problem, as, for example, certain characteristics of a simulator might only be relevant for certain use-cases or in combination with certain MCAs.

Another problem pointed out by Nahon and Reid [1990] is that in the early stages of the design process of experiments the justification for a specific MCA requires exact information on how that MCA will perform. How it will perform, however, is unknown, as the tuning will only be performed *if the MCA is selected*. In other words, the justification for tuning an MCA is given by the information on the tuned MCA, leading to a causality problem. Similarly, exact information on the given use-case, such as driving data to evaluate the motion cueing quality, may not even be available in the early selection stages. Therefore, an *estimate* of the *potential* motion cueing quality must be made. Such a multi-dimensional selection process currently does not exist in literature. Therefore, the third research goal can be stated as:

Research Goal 3

To develop a methodology to select the potentially best-suited motion cueing method (simulator and MCA) of a driving simulator experiment to improve its motion cueing.

1.6 Overall Research Goal and Outline

Combining the presented research goals, the overall research goal of this dissertation is:

Overall Research Goal

To develop a validated method to **objectively** assess and improve the **total potential** quality of motion cueing for **closed-loop** driving simulator experiments.

The dissertation is structured into three parts, see Figure 1.6. Each part aims to address one of the three research goals. The main findings and implications of each chapter and part are summarized and discussed in Chapter 10. Together, they aim to fulfill the Overall Research Goal of this dissertation.

Part I - Fundamentals of Motion Cueing Quality

This part aims to identify currently existing methods and the state-of-the-art of quantifying motion cueing quality and to extend the concept of motion cueing quality through operational metrics.

Chapter 2 introduces the concepts of quality, cost, and tuning metrics that can affect, and should affect, the selection of a motion cueing method. To identify which metrics are of importance for driving simulator experiments, it describes the results of a survey, in which four stakeholder groups (experimenters, motion cueing engineers, participants, and operators) were interviewed on metrics important to them in the preparation for a driving simulator experiment.

Chapter 3 derives the kinematic relations of frequently-used motion system architectures and evaluates the importance of accounting for such differences. The results will be shown to be highly dependent on the motion system configuration and thus need to be implemented correctly to evaluate and predict the quality of the motion. Results from this chapter form the baseline for all later chapters involving simulator experiments.

Part II - Predicting Motion Cueing Quality

This part addresses the second research goal, by developing rating models to predict the subjective ratings of motion cueing quality based on objective motion mismatches.

Chapter 4 describes an open-loop driving simulator experiment in a realistic urban scenario, in which participants evaluated the quality of the motion using the continuous rating method. Subsequently, a model is developed for the prediction of the measured subjective ratings. This chapter, from a methodological perspective, forms the basis of the other four chapters in Part II.

Chapter 5 then applies the urban model of Chapter 4 to rating data gathered in a highway experiment, with an otherwise identical experiment set-up. The effect on the ratings between the scenarios and necessary adaptations to be made in the predictive model are investigated. To complete the triangle of common scenarios, **Chapter 6** uses the continuous rating data of three rating experiments in a rural scenario from literature.

The last two chapters of Part II return to the urban scenario. **Chapter 7** investigates the equivalence of ratings in closed- and open-loop driving by means of a dedicated experiment. It is furthermore investigated whether the developed rating models are applicable for closed-loop driving simulator experiments, and not only for open-loop settings. Lastly, **Chapter 8** investigates how the findings regarding motion incongruences can be extended towards simulator sickness under variations of axis and error type effects. Based on these results, further improvements to the rating model are proposed.

Part III - Improving Motion Cueing Quality

This part addresses the third research goal, by developing a methodology that evaluates the total quality of various available motion cueing methods.

Chapter 9 contains an analysis of two real driving simulator use-cases. An analysis is made of the relevant quality and operation metrics of these use-cases, and a prediction of the potential motion cueing quality is made. Together, these form the basis of a trade-off to select the best-suited motion cueing method for both use-cases.

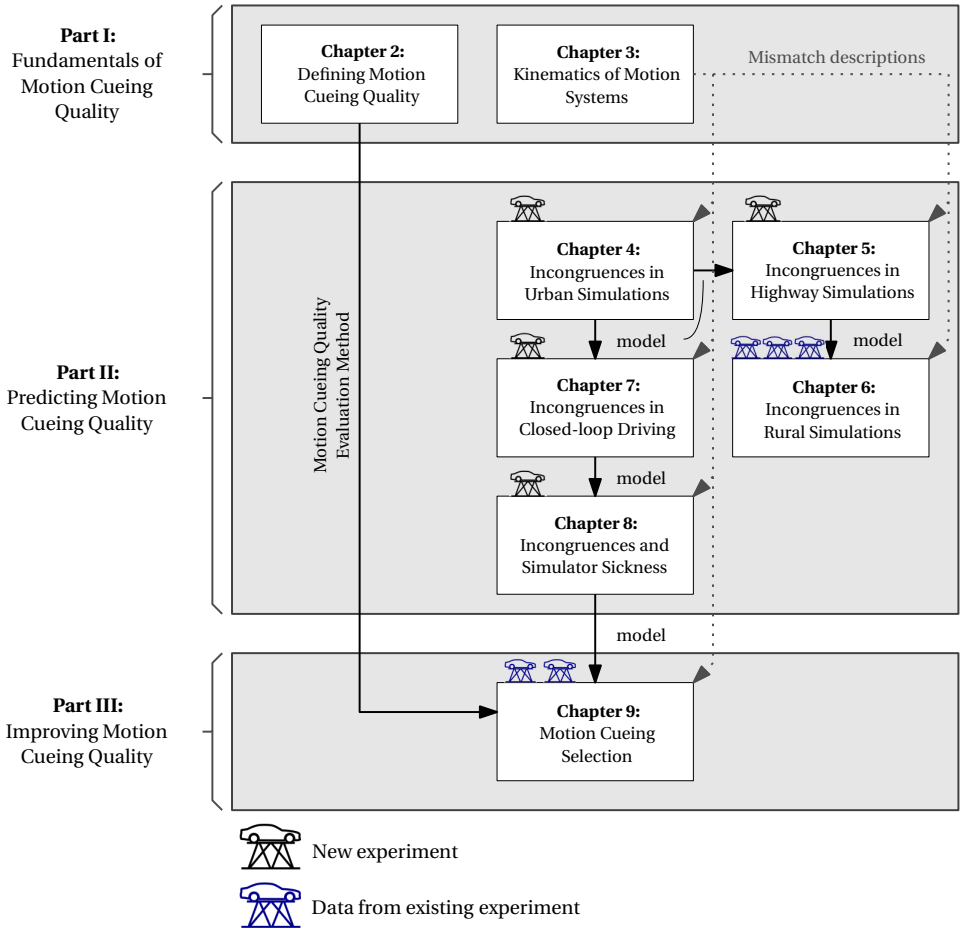


Figure 1.6: The outline of this dissertation, with the parts, chapters, and their dependencies highlighted.

Appendices

This dissertation includes four appendices which present supplementary material. **Appendix A** contains the survey that was described in Chapter 2. **Appendix B** contains the written briefing given to the participants before the experiment phase of Chapter 4, which formed the basis of the briefings given in the other experiments described in this dissertation. **Appendix C** describes a continuous rating experiment performed on a race track scenario. Due to technical difficulties the experiment had only a limited sample size and the data are thus not used in answering the dissertation's research goals. However, as the experiment contained both closed- and open-loop driving, the experiences gained formed the foundations of the explicit closed- and open-loop driving rating experiment of Chapter 7. Finally, **Appendix D** shows all individual ratings given by the participants in all experiments described in this dissertation. These ratings may provide additional insight to the reader how individual rating strategies differ.

1.7 Scope

Driving simulator experiments become increasingly more complex, operating at the interface between understanding the human operator on one side, and control theory on the other. For this reason, several choices were made to set the scope of this dissertation.

First, the trade-off methodology presented in the dissertation always contains a single experiment. Aspects regarding the availability of the various motion systems are also not considered. When multiple experiments are to be distributed among several motion systems, a logistical distribution of the experiments would be required. This, however, is not considered in this dissertation, to limit the scope of the already complex problem.

Second, it is assumed that differences in the perceived motion only result from differences in the reproduction of the motion cueing itself. The proposed findings and solutions are thus assumed to be *independent* of the simulation architecture (e.g., simulator and mock-up). Therefore, also only simulators with a high physical fidelity are considered, i.e., with full mock-ups of the vehicle, providing full 360° view to the driver. Differences arising due to the *type* of visual system (e.g., the monitor displays of the "Stage" simulators, the LED displays of the "Vector" simulators, and projector systems of the "Space" simulators, see Figure 1.2) are not considered in the evaluation. The methods also do not explicitly consider *interactions* between visual and inertial motion perception, e.g., through visual-vestibular integration [Markkula et al., 2019], to reduce the complexity of the research problem.

A third scope limitation lies in the reproduction of the inertial motion itself. Many modern simulators provide the ability of simulating tire- or road-induced vibrations (such as by using Curved Regular Grid (CRG) technologies) [Romano et al., 2019]. These can either be provided by the motion system directly, or through dedicated "shaker systems" in the vehicle mock-up. CRG technologies often require road surface scans and add significantly to the complexity of the simulation, such that they are not (yet) applied as a standard in driving simulations. In this dissertation, road rumble is therefore not considered, in an attempt to isolate the effect of inertial motion reproduction as much as possible. For this reason, the road rumble was turned off in all experiments. For considerations and recommendations on the effect of including road rumble systems, the

reader is referred to the work of Parduzi [2021].

Lastly, to increase the generalizability, the dissertation is limited to findings that concern the "average" driver. Given that all participants of the experiments are employees of BMW Group, differences in the degree of familiarity with vehicles may arise. Although the experiment described in Appendix C specifically only considered "expert drivers", the experiments in the main chapters of this dissertation did not have a pre-selection of participants based on vehicle familiarity, experience, or driving skill, and only estimations for the average driver are considered.

1.8 Guidelines for the Reader

The chapters of this dissertation are written as separate papers, in most cases submitted or published as either journal or conference papers. For published papers, it was attempted to remain as close as possible to the original publications, as these works have been peer-reviewed. A description of the publication status is given on the first page of each chapter, together with a summary of the most important findings of that chapter. In case two publications were merged to form one chapter, both publications are indicated (see e.g., Chapter 4). As the methodology of many papers was similar, some information might also repeat itself in various chapters. It was attempted, while writing the publications, to make the nomenclature as consistent as possible between the chapters.

Fundamentals of Motion Cueing Quality

2

Defining Motion Cueing Quality

Key findings

- In a trade-off between various MCAs and motion systems for a driving simulator experiment, a fairer comparison can be made if an estimation of their full motion cueing potential is considered.
- Apart from predictions of motion incongruence ratings, including quality, cost, and tuning metrics yields a more effective comparison and trade-off between the MCA and simulator options.

This chapter is based on the following publication:

Title:	Driving Simulator Experiment Stakeholder Perspectives on Motion Cueing Algorithm Quality.
Authors:	M. Kolff, J. Venrooij, D.M. Pool, and M. Mulder.
Conference:	Driving Simulation Conference 2023 Europe, Antibes, France, pp. 99–106.

Abstract

In driving simulation, the trade-off between different motion cueing methods (the motion cueing algorithm and the simulator) is typically made with an exclusive focus on the quality of the motion itself. In practice, however, many other metrics could affect the trade-off as well, such as the resulting financial cost and the tuning complexity. Furthermore, in a trade-off, sub-optimal tuning sets are commonly used, leading to possibly unfair comparisons between motion cueing methods. This chapter provides a framework for improved motion cueing method trade-off and selection based on two new standards. First, the results of an expert survey show that, next to the quality of the motion, the total cost, simulator sickness, ease of use, tuning effort, tuning complexity, and stability are typically of high importance. Second, it describes how subjective ratings on the quality of the motion can be obtained and how predictions of these ratings can be used to estimate the full potential of motion cueing methods. The presented methods allow for a more accurate, multi-faceted trade-off of motion cueing algorithms and simulators, improving the quality of future driving simulation experiments.

2.1 Introduction

EXPERIMENTS in driving simulation are often performed on motion-base simulators to increase the realism and reduce simulator sickness. Because of the inherently restricted size of a motion-base, the motion of the real vehicle can generally not be fully reproduced. Therefore, a Motion Cueing Algorithm (MCA) is required, which converts the motion of the real vehicle to motion that fits within the simulator workspace. A wide range of simulators and MCAs exists for a growing number of driving simulation use-cases. Selecting the best-suited combination of a simulator and an MCA, for a given use-case, is critical for an effective and valid experiment.

Evaluations of motion cueing can be given either objectively or subjectively. In objective metrics, signal differences between the vehicle reference and simulator are evaluated. Examples are absolute difference [Casas et al., 2015], cross-correlation [Grottoli et al., 2019], and delay metrics. Most literature on comparing motion cueing, however, focuses on evaluations through subjective Motion Incongruence Ratings (MIRs) instead [Cleij et al., 2018], where the quality of the motion is evaluated by a human test driver. Although there is an active discussion whether objective or subjective evaluations hold more value [Casas et al., 2015], it is the latter that have shown to be the backbone in evaluating and selecting motion cueing for driving simulation, e.g., Cleij et al. [2018], Cleij et al. [2019], Lamprecht et al. [2019], Ellensohn et al. [2019], Ellensohn et al. [2019], Ellensohn et al. [2020], and Valente Pais et al. [2009].

When a large selection of simulators and MCAs is available to choose from, it becomes practically infeasible to tune and subjectively evaluate each available combination. In a selection process, it is thus necessary to predict the subjective MIRs that a given combination will achieve, which is a challenging endeavor [Cleij et al., 2018]. In the trade-off and selection of simulators and MCAs, their quality is most often based on their current state of tuning, and not on the potential they can reach, *precisely because* they are still untuned. For a fair comparison, it is thus also necessary to estimate the *potential* that each combination can reach without actually performing their tuning. Furthermore, in practice, other metrics concerning the experiment cost or the effort required to tune a given MCA influence the MCA and simulator choice as well. Which metrics are of importance in a trade-off is likely to be dependent on the experiment use-case. The only authors that have explicitly recognized the importance of other metrics that determine the quality of MCAs were Nahon and Reid [1990]. To be able to provide fair comparisons between MCAs and simulators, a paradigm shift towards predicting and evaluating the *total potential* motion cueing for quality evaluations is thus needed.

The goal of this chapter is to propose a basis for such quality evaluations to improve the motion cueing selection for future driving simulator experiments. This basis consists of two components. First, it provides a framework for the design process of motion cueing experiments that explicitly includes the identification of all relevant quality, cost, and tuning metrics, together defining the 'total' potential motion cueing quality. To this end, an expert survey is performed among the stakeholders involved in motion cueing experiments to identify the most important metrics of motion cueing experiments. Second, it describes the state-of-the-art methods of measuring subjective ratings on the motion cueing and how these ratings can be predicted to make accurate evaluations on the motion cueing of future driving simulator experiments.

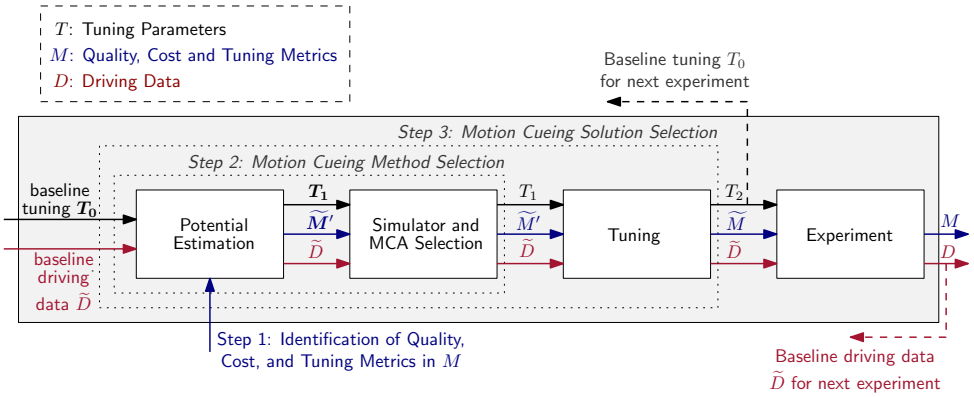


Figure 2.1: Proposed steps in the selection process of motion cueing, highlighting the role of tuning sets T, quality metrics M, and driving data D.

2

This chapter is structured as follows. First, Section 2.2 describes the framework for motion cueing selection, including a description of the survey and its results. Section 2.3 then describes the methods in measuring motion cueing quality. A discussion is presented in Chapter 2.4. The conclusions are given in Chapter 2.5.

2.2 Framework for Motion Cueing Selection

2.2.1 Selection Process

The central aspect in a driving simulation experiment is the use-case, describing the type of vehicle, the scenario, and the goal for which the experiment is performed (see Chapter 1). For a successful experiment, the simulator and MCA are chosen and tuned according to the requirements of the use-case to achieve the best-suited motion cueing, given a set of constraints (e.g., cost, available time, man-power). Second, this choice should always be towards the best and "cheapest" combination that still fulfills the use-case requirements. For some experiments, a relatively low motion cueing quality may be perfectly acceptable and a time and money investment in providing a high quality solution is not needed [Fischer et al., 2015]. The proposed further process of selecting the simulator, MCA, and tuning parameters consists of three steps, as shown in Figure 2.1. First, it must be identified which metrics are of importance to the experiment, described by the vector M. It must then be determined how well these metrics can potentially perform, on which the simulator and MCA (the method) can be chosen. Finally, the tuning of the MCA is performed and the experiment can commence.

2.2.1.1 Step 1: Quality, Cost, and Tuning Metric Identification

As explained in Chapter 1, the analysis and evaluation of MCAs is currently almost exclusively performed in terms of the quality of the motion. Nahon and Reid [1990] noted that other qualities can, and should, impact the choice of the motion cueing as well. These qualities can be similar to the quality of the motion itself (such as the occurrence of

simulator sickness as a result of bad motion cueing), but can also relate to several other metrics considering the cost aspects of an experiment. This consideration seems natural, as the total financial cost will be a practical consideration in the design of most experiments. More broadly speaking, cost metrics can also refer to other "negative" contributors, such as energy consumption [Meike and Ribickis, 2011], noise production [Schwarz et al., 2003], or required maintenance.

In the analysis of Nahon and Reid [1990], an Adaptive Washout Algorithm (AWA) and an Optimal Washout Algorithm (OWA) were compared as viable alternatives to a CWA. Even though the AWA provided a higher motion cueing quality compared to CWA, the high tuning complexity, lower transparency in the tuning process, and inherent instability of the AWA made it less favorable for practical application in a simulator. This also uncovers a third category of "tuning metrics", which describe the difficulty and practicality of implementing a specific MCA (in combination with a simulator and a specific use-case) in terms of its tuning.

When considering the quality, cost, and tuning metrics, *which* of the metrics are actually of importance will depend on the use-case and are thus different per experiment. The important metrics in the vector M are ideally identified early in the experiment design process ("Step 1" in Figure 2.1), such as through a standardized survey or interview that is presented to those involved in the experiment. Different *stakeholders* exist in the design of an experiment, each with their own wishes for the experiment:

- The **experimenter** wants to perform an experiment to answer a research question. Therefore, experimenters will generally want an experiment design (and associated motion cueing) that is most suitable for answering this research question.
- The **motion cueing engineer** designs and/or tunes the MCA according to the wishes of the other stakeholders to achieve the best-suited motion cueing solution. The motion cueing engineer is the expert on the design of the motion cueing.
- During the experiment, the **simulator operator** operates the simulator from the control room and monitors the safety and well-being of the participant. This stakeholder could for example be interested in a stable experiment without the occurrence of simulator sickness.
- Central in each experiment is the **participant**, as they participate in the experiment and thus experience what the motion cueing solution is designed for. Although their personal wishes might have the smallest impact on the design of the experiment, they can still have clearly defined wishes during the experiment that affect the success of the experiment, such as a low level of simulator sickness.
- Finally, the **organization** that owns the simulator. Their interest could for example lie in the overall safety and keeping the cost, working hours, and energy consumption as low as possible.

In some cases, the same individual may have multiple stakeholder roles. For example, someone could be the experimenter and the simulator operator at the same time.

2.2.1.2 Step 2: Motion Cueing Method Selection

The simulator and MCA choice together define the *motion cueing method* of the motion cueing for which the tuning and the later experiment are performed (see Figure 2.1). In

practice, it is typically the simulator that has to be chosen first, as simulators often must be reserved well beforehand. For the MCA choice, it must be determined which type of algorithm is used. Different MCA options exist (e.g., CWA, AWA, and MPC), each with their own (dis)advantages, as discussed in Chapter 1. The quality of an MCA can also be strongly related to the simulator choice. For example, whereas for a hexapod system a small angle approximation can be used [Ellensohn, 2020] to simplify the motion control problem, this simplification is no longer valid on a system with a yaw-drive (allowing large yaw angles), such as BMW's Sapphire Space simulator. Vice versa, a complex MCA might only be able to exploit the simulator workspace better if it is given sufficient workspace. The simulator choice can thus not be made independently of the (later) MCA choice. The same holds for the use-case: depending on the type of motion, different qualities of the simulator and MCA can become more or less pronounced. For example, a yaw-drive might only be of added value in tight corners of an urban scenario, and might only be controlled effectively using an MPC algorithm. Selecting the best-suited simulator when various simulators are available thus requires understanding how well a simulator will perform in *combination* with the MCA for a given use-case.

An inherent process uncertainty lies in the quality of the input data on which the motion cueing method selection is performed. Because the experiment has not yet commenced, driving data on how the participants will drive during the experiment are generally unavailable at this point. Once the scenario and its route are fixed, "baseline driving data" ($\widetilde{D}$) can be gathered from a small group of test drivers or from a similar previous experiment to form a first reference. The notation $\widetilde{(\cdot)}$ denotes the uncertainty caused by not having the final driving data of the experiment yet. Another uncertainty is that although "baseline tuning" parameter sets $\boldsymbol{T_0}$ (e.g., from previous experiments) might provide a good starting point, these tuning sets at this point may not be fully optimized for the given use-case. A comparison between available options is only fair when these have been tuned to their full potential [Ellensohn et al., 2020; Nahon and Reid, 1990]. 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. This would require a large number of motion cueing methods to be tuned, which is expensive and impractical. It is therefore difficult to make early and informed decisions on which motion cueing method suits a specific use-case best.

A potential estimation for each of the available options must thus be made to provide a fair comparison between the available motion cueing methods. This can be based, for example, on altering the gains of the baseline tuning sets $\boldsymbol{T_0}$ by analyzing the remaining workspace of the simulator [Grottoli et al., 2019], giving an estimate on how the tuning may be improved given that the workspace is fully utilized. The more accurate these prediction tools, the better the process of motion cueing method selection can be performed. Based on the potential estimation, already improved tuning sets $\boldsymbol{T_1}$ (Figure 2.1) can be extracted. After the simulator and MCA are selected, only a single tuning set T_1 is then used for the succeeding steps.

The quality vector $\widetilde{\boldsymbol{M}}'$ contains an estimate on the potential of the relevant quality, cost, and tuning metrics. The notation $(\cdot)'$ indicates the estimation of the potential. It thus also contains an estimate of the potential on the quality of the motion. A single value for each option representing the quality during the whole experiment would likely

be the easiest method for comparing between the various options, such as post-hoc overall ratings (OR_{PH}) [Cleij et al., 2018], as will be discussed in Section 2.3.

2.2.1.3 Step 3: Motion Cueing Solution Selection

Once a motion cueing method (simulator and MCA) is selected, the MCA tuning process starts. In this step, the parameters of the method are changed to provide the highest possible quality, while still fitting the motion (of most participants) within the workspace of the simulator. This can be done either through offline simulation (objectively) or through on-site testing (subjectively) at the simulator. The parameters of the method are then iteratively changed until a satisfactory level of motion cueing is achieved, resulting in the final tuning set T_2 and estimated quality metrics $\widetilde{M}$. Similar to the motion cueing method, the combination of the method and the eventual set of tuning parameters as selected for an experiment can be defined as the *motion cueing solution*. The resulting tuning parameters can be stored as a baseline tuning for a next experiment. When the tuning process is completed, the final quality of the motion cueing method is still uncertain, however, due to the unavailability of the experiment's driving data.

2.2.1.4 Step 4: Experiment Phase

Once an adequate tuning is obtained, the experiment can commence. To improve the overall process in each of the three preceding steps, it is highly beneficial if a set of representative outputs from the experiment is acquired. A non-invasive validation of the motion cueing can be performed through overall post-hoc ratings (See Section 2.3). These subjective ratings can be asked after a drive and therefore do not affect the driving itself. They can therefore be obtained from practically any experiment and can thus further validate rating predictions that were made beforehand. Furthermore, the actual vehicle driving data D for each drive should be recorded as well, as this can serve as the representative data on which the motion cueing of future experiments is designed.

2.2.2 Identification of Quality, Cost, and Tuning Metrics

To identify which of the metrics are of the highest importance to BMW's driving simulator experiments, a survey was performed. This survey relies on identifying the key metrics from a pre-defined catalogue of quality, cost, and tuning metrics.

2.2.2.1 Quality Metrics

The following metrics are proposed to present the quality of a motion cueing experiment, which would all be ideally maximized as much as possible.

Perceptual Quality

Perceptual motion quality is often considered in driving simulation, as described in Section 2.3. When the presented motion is perceived as unrealistic, perceptual quality is low [Cleij et al., 2018]. To increase perceptual quality, more accurate MCAs and/or larger motion simulators are generally required. The importance of perceptual quality can differ per experiment, as some experiments might induce so little motion (e.g., when mostly driving straight) that perceptual quality is not of high importance.

Behavioral Quality

The behavioral quality of the simulation is determined through the degree of similarity in driving behavior compared to the behavior 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 behavior [Romano et al., 2019] or the interaction with traffic, as well as handling secondary driving tasks (such as operating a navigation system).

Reducing 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]. In terms of the motion cueing, it is thus a function of the perceptual fidelity. Simulator sickness is characterized by physical symptoms (e.g., dizziness, headache, nausea [Reason and Brand, 1975]) and can lead to participants dropping out of the experiment (increasing their cost), provoke unwanted changes in behavior [Cobb et al., 1999; Igoshina et al., 2022], and lead to a decreased sense of presence [Almallah et al., 2021].

The level of simulator sickness has also been shown to depend on the use-case. For example, urban driving generally results in more simulator sickness than rural driving [Himmels et al., 2022b], which in turns induces more simulator sickness than highway driving [Klüver et al., 2015; Mourant and Thattacherry, 2000]. For these reasons, simulator sickness is generally to be avoided and motion cueing solutions for use-cases that are expected to induce a high degree of sickness are ideally identified beforehand.

Stability

Unstable motion cueing solutions can crash, such that a drive cannot be continued. Apart from loss of valuable experiment data, a motion cueing solution that is unstable might lead to unexpected and/or dangerous behavior of the system. This can affect the quality of the motion, but also decrease the system safety, resulting in a higher risk of incidents. Stability can also be considered to be an important metric especially for long drives, as restarting a drive would lead to the loss of valuable experiment time, increasing the chance that an experiment session cannot be finished. Especially for MPC MCAs, the associated inherent stability has shown to be an issue [Fang and Kemeny, 2016]. In contrast, CWA MCAs are known for their inherent high stability [Nahon and Reid, 1990].

2.2.2.2 Cost Metrics

In contrast to quality metrics, which are preferably maximized to the extent possible, cost metrics should be minimized as much as possible.

Financial cost

The financial cost of an experiment is a driving factor in the design of motion cueing [Allen et al., 1995; Huang and Chihsiuh, 2003]. It is a function of the purchasing cost of the simulator, the cost of preparing and tuning the motion cueing, and the cost of running the simulator (energy consumption, maintenance, etc.) [Bennett, 1995], which generally all 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. It can be argued, however, that the most expensive simulator or MCA in terms of purchasing

cost are those that are *not* used for experiments, as the purchasing cost per hour used goes down with increasing system usage [Kappe and Van Winsum, 2002].

Energy Consumption

Energy consumption is a metric of increasing importance [Meike and Ribickis, 2011] due to costs, environmental impact, and hardware limitations. Although the energy consumption might not be important for those involved in designing the motion cueing, it could be important for the organization 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 quality. Larger, heavy systems (e.g., BMW's Sapphire Space and Diamond Space) tend to generate both more noise and vibrations due to the combination of a high weight and large platform movements [Schwarz et al., 2003].

Maintenance

Apart from increasing the cost, maintenance also 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.

2.2.2.3 Tuning Metrics

Finally, tuning metrics are generally only relevant to the motion cueing engineer and describe properties of finding suitable tuning parameters for the motion cueing method.

Total Number of parameters

The total number of parameters can be considered as a measure of tuning effort and complexity [Nahon and Reid, 1990]. Ideally, the number of tuneable parameters would be as low as possible, as each parameter requires time and effort to tune. As an example, a CWA requires tuning of its several gains, split frequencies, and washout parameters. This results in 23 parameters for a typical hexapod system (e.g., Vega Vector, see Chapter 1). Including a yaw-drive results in an additional split frequency and a first-order washout parameter. An xy-drive adds another split frequency and two washout parameters. The total number of parameters can thus be expressed as $23 + 2b_{\psi_d} + 3b_x + 3b_y$, with the Booleans b_{ψ_d}, b_x, and b_y indicating the presence of a yaw-drive, x-drive, and a y-drive, respectively. This results in 25 for additionally including a yaw-drive (Sirius Vector), and 31 for also including an xy-drive (Sapphire Space and Ruby Space). Whereas additional motion subsystems thus potentially increase the quality of the motion, they also increase the number of parameters to be tuned and therefore likely the tuning effort.

Note that the "cost" due to the number of parameters can also be reduced by only considering the effective number of parameters. For example, depending on the use-case, it is possible that not all vehicle axes require tuning. This can, for example, occur when a use-case with only lateral motion 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. Furthermore, it can be chosen to use the same parameters along various axes (e.g., longitudinal and lateral motion) to make the motion

in these axes as similar as possible, reducing the number of parameters to be tuned.

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 [Casas et al., 2016; Nahon and Reid, 1990]. Thus, the clearer it is which effect a parameter has, the better. One benefit of a CWA, for example, is that changing parameters often has a direct and predictable effect on the motion. In contrast, changing parameters for MPC is generally less transparent, as the optimization nature of the algorithm does not necessarily lead to a different MCA output.

Determinism of System

Some MCA types will always give the same output when the same input is provided. An example of this is the linear filter approach of the CWA. In contrast, MPC is based on optimization, such that the output 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

An MCA that can be tested outside of the physical simulator allows for additional debugging and testing capabilities. Thus, having no offline testing option can also lead to a higher tuning complexity.

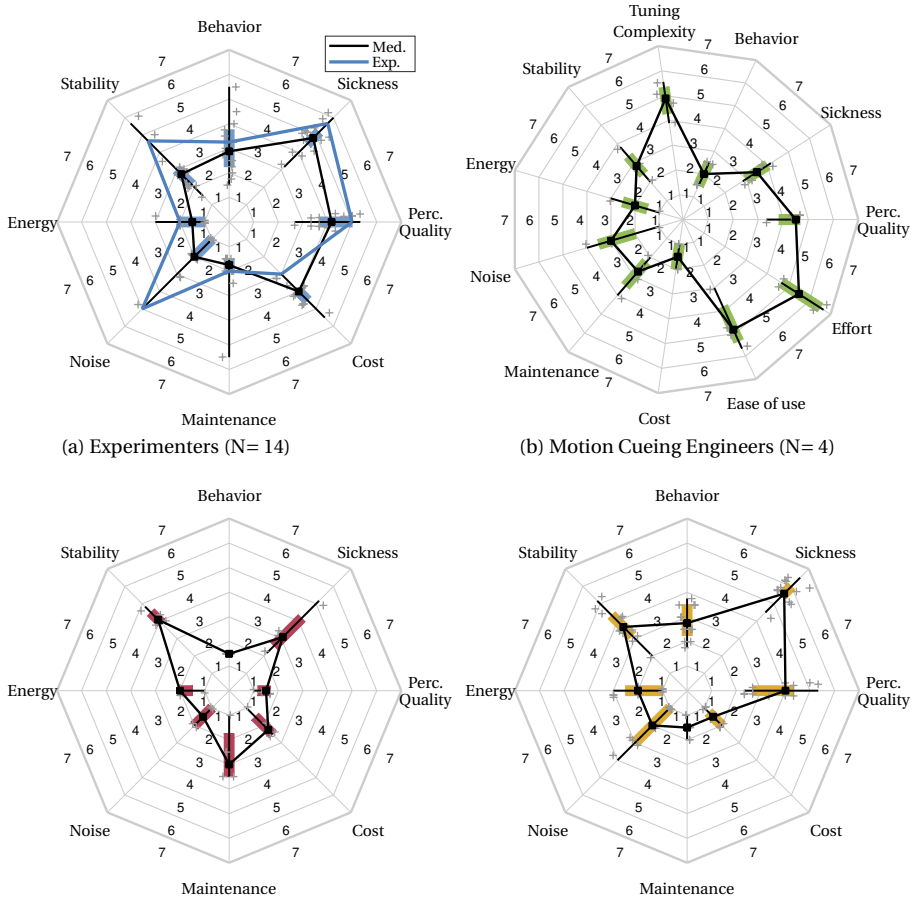
2.2.3 Survey

A survey was developed to identify the metrics of highest importance to driving simulator use-cases at BMW. The same survey was presented to the considered stakeholder groups to enable a direct comparison between them. The results of the stakeholder "organization" are not presented due to data protection reasons.

2.2.3.1 Methodology

The survey consisted of 35 questions (in German) on quality, cost, and tuning metrics, divided into several categories. Answers were given on a seven-point Likert scale ranging from fully disagree (score of 1) to fully agree (score of 7). The survey questions are shown in Appendix A. As the last three categories (A.2.7 - A.2.9) only apply to motion cueing engineers, these were only filled out by this group.

In total, 14 experimenters, 4 motion cueing engineers, 4 operators and 12 participants filled out the survey. The respondents filled out the survey in a document, after receiving a short briefing on the context of the survey. They 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 to obtain an insight in the general importance of each category.



(a) Experimenters (N= 14)

(b) Motion Cueing Engineers (N= 4)

(c) Operators (N= 4)

(d) Participants (N= 12)

Figure 2.2: Radar plots showing the box plots over all survey results for the quality, cost, and tuning aspects. The black line ("Med.") connects the medians of the distributions for increased visibility. The blue line ("Exp.") highlights the specific results of the highway driving experiment.

2.2.3.2 Results

Figures 2.2a-2.2d show the survey results, displayed as radar charts, similar to the simulator requirement analysis of Fischer et al. [2015]. For each category, a box plot is shown along each 'spoke' of the radar chart. The lines on each axis represent the data ranges, the colored areas the first and third quartiles, and the grey crosses 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, higher values indicate higher importance.

For the "experimenter", Figure 2.2a, who requests the experiment to be performed, the results were collected from fourteen past or planned studies. Generally, the quality of the motion, reducing simulator sickness, and the cost of the experiment were shown to be the most important metrics for these studies (i.e., > 4, above the answer 'neutral'). The large range in answers given to stability can be explained by differences in drive lengths, ranging from 2 to 45 minutes per drive. For short drives stability is likely not so important, as they can be easily repeated.

Four "motion cueing engineers" filled out the survey (Figure 2.2b). Note that here the data points do not represent individual studies, but represent what these engineers generally find important. Aspects of tuning complexity and tuning effort are found most important. Ease of use (in the operations) shows to be somewhat of importance. Although the engineer is not the stakeholder performing the actual experiment, an easy-to-use MCA can still be beneficial when designing, debugging, or testing an MCA.

The metrics of highest importance to the "operator" (Figure 2.2c) are the ease-of-use, which is explained by the operator having to use the motion cueing method in an experiment. The 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.

Finally, stability is a similar issue for the "participant" group (Figure 2.2d). Here, twelve survey respondents indicated that this metric is of medium importance, which could be in the interest of protecting their own well-being in the simulator. Most importantly, perhaps unsurprisingly, is simulator sickness, which was indicated by all respondents to be of importance (> 4).

2.2.3.3 Example Experiment: Automated Highway Driving

To highlight what important quality, cost, and tuning metrics can look like for a specific experiment, the survey results of a real driving simulation experiment at BMW are highlighted. The experiment investigated the role of fatigue under an ADAS system. Participants first drove for 15 minutes on a German highway scenario. After that, the autonomous driving systems were engaged until the participant reported a high level of drowsiness. This phase could take up to 90 minutes. After that, a second manual segment was performed, lasting another 90 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 line in Figure 2.2a. 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 was to be reduced as much as possible, as this could influence the fatigue and attentiveness of

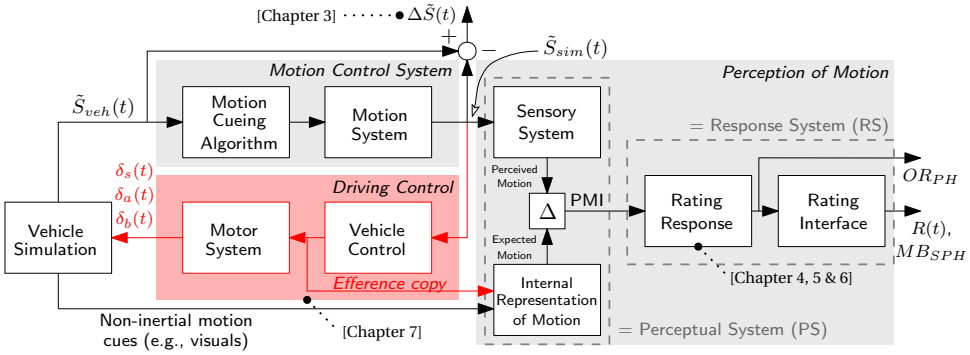


Figure 2.3: A model of the motion cueing rating process in closed-loop driving simulations.

drivers. Furthermore, as it considers an experiment in a relatively long scenario (> 120 minutes) of a single drive, stability was important, as a simulator hardware or software crash could render the long experiment sessions useless. Concluding from these results, the metrics stability, simulator sickness, motion cueing quality, tuning effort, and tuning complexity are determined to be the most important (> 4) metrics across all stakeholders, which led to the selection of the relatively small Vega Vector simulator and a CWA as a highly highly stable motion cueing method.

2.3 Predicting Motion Incongruence Ratings

The survey results showed that several metrics define the "total" motion cueing quality, including the quality of the motion itself. When the potential of a motion cueing method is to be estimated prior to its selection and implementation, subjective ratings are not yet available, because the motion cueing cannot be tested without the actual tuning. This metric thus differs from the other metrics, as it requires a *prediction* of the subjective rating to be able to express its potential. Accurate prediction methods of the subjective ratings are thus paramount. This section therefore focuses on current standards for measuring and predicting subjective motion cueing quality.

2.3.1 Driving Task

In most driving simulation use-cases, participants will drive themselves, i.e., *closed-loop* scenarios. As illustrated in Figure 2.3 by the red paths, the driver then actively controls the steering wheel $\delta_s(t)$, the accelerator $\delta_a(t)$, and brake $\delta_b(t)$ pedals. These act as inputs to the vehicle simulation, which calculates the motion states of the simulated vehicle $\tilde{S}_{veh}(t)$. The motion states consist of the specific forces $[f_x,\ f_y,\ f_z]$ and the rotational rates $[\omega_x,\ \omega_y,\ \omega_z]$. As $\tilde{S}_{veh}(t)$ comes from a vehicle model, it is an approximation of the real vehicle motion $S_{veh}(t)$.

The simulated vehicle motion states are sent to the Motion Control System (MCS), which consists of the MCA and the Motion System (MS). Whereas the MCA converts the vehicle motion states to commanded platform motion, the MS (the physical simulator and its control software) provides the platform motion signals $\tilde{S}_{sim}(t)$ to the driver.

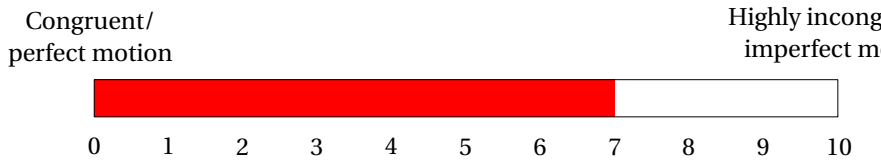


Figure 2.4: Semantic differential rating scale, ranging from 0 (Congruent motion) to 10 (Highly incongruent motion). It currently displays a rating value of 7.

These can differ from the commanded platform motion due to a variety of factors, such as the motion system latency. Differences between the vehicle reference and simulator motion are then the objective mismatches, i.e., $\Delta\tilde{S}(t) = \tilde{S}_{veh}(t) - \tilde{S}_{sim}(t)$. The human driver perceives the platform motion through their sensory system. The relation between the inertial platform motion and the output signal $\tilde{S}_{sim}(t)$, in terms of specific forces and rotational rates, depends on the kinematic configuration of the simulator and is thus crucial in the analysis of motion incongruences. How $\Delta\tilde{S}(t)$ is properly expressed for each motion platform will be further described in Chapter 3. Based on the inertial motion, as well as other, non-inertial motion cues (e.g., the visuals) the driver controls the vehicle. The motor system of the human body produces the control actions $[\delta_s(t), \delta_a(t)$ and $\delta_b(t)]$, closing the loop.

2.3.2 Rating Task

To be able to evaluate the motion cueing quality, participants must also perform a rating task by evaluating how well the inertial motion they perceive in the simulator matches to what they would expect to feel from the real vehicle. This difference is defined as their Perceived Motion Incongruence (PMI) [Cleij et al., 2018], see Figure 2.3. Drivers must use an *internal representation* [Stassen et al., 1990] of the vehicle motion based on non-motion cues (e.g., visuals) of the simulation as they do not exactly know what the vehicle motion would feel like in a particular situation. Differences can thus arise based on memory, as well as familiarity and experience with the simulated vehicle. Note that thus both the expected motion and as the vehicle model motion $\tilde{S}_{veh}(t)$ thus likely differ from the real vehicle motion $S_{veh}(t)$. In closed-loop control, an *efference copy* [Mulder et al., 2022] of the intended control actions is fed to the internal representation of motion. It describes what drivers expect to feel, based on their own control actions.

As the PMI is an internal signal to the human, a Motion Incongruence Rating (MIR) must be asked from the participants that represents the PMI. Several options exist to extract MIRs. These refer to which parts of the drive the ratings represent, when they are obtained, how they are measured, on which scale, and what part of the motion they represent. A proper choice of these variables is crucial, because biases in the MIRs can occur depending on them, leading to MIRs that are not representative of the PMI.

A typical choice for MIRs is an ordinal rating scale [Cleij et al., 2018]. Such rating scales can be expressed with wording (congruent motion to highly incongruent motion) or using associated values, e.g., ranging between values of 0 (representing congruent) and 10 (representing highly incongruent), with steps of 1 [Kolff et al., 2024b] (Figure 2.4).

As shown in Figure 2.3, participants use a “Rating Response” to determine their MIR. Based on practical experience with subjective ratings, it is expected that this rating re-

sponse (and therefore the MIR) can be affected by the following effects:

1. *Task motivation* describes the willingness to focus (on the motion) and actively perform the (rating) task [McRuer and Jex, 1967; Mulder et al., 2018].
2. *Cueing reference* refers to what values drivers apply for the given incongruences, which depends on which PMI-level they associate with the maximum (10) MIR score. For example, in Cleij et al. [2018], participants were shown the full range of the incongruences before the experiment for reference.
3. *Anticipation* can occur when incongruences of upcoming maneuvers are expected based on previous drives, or from recognizing that a certain MCA setting is active.
4. *Task understanding* of the participant that only the PMI is to be evaluated, and no other motion-related phenomena (e.g., visuals, engine sound or vibrations).

2.3.2.1 Rating Representation

One choice in the rating method concerns which part of a drive a rating represents. Three categories of rating representations can be used, as shown in Table 2.1.

Overall Ratings

An Overall Rating (OR) aims to capture the impression of the whole drive using a single rating. These ratings have been used extensively [Biemelt et al., 2021; Rengifo et al., 2021], as the method is arguably easiest for participants to understand. The single-value overall ratings also allow for a direct comparison and one-off evaluation between different motion cueing settings. Inherently, overall ratings are limited in the details on how different parts of the drive compare. A single rating for each drive provides no direct information on what this rating is mostly based on. For example, a "medium" overall rating can be caused by a drive with a medium level of incongruences. However, it can also be caused by a highly contrasting drive, in which moments of highly incongruent motion are compensated by moments of congruent motion.

Depending on when the overall rating is extracted, there can also be significant time between obtaining the rating and earlier parts of the drive. This can lead to biases caused by short-term memory effects, such as the serial position effect [Murdock Jr, 1962] and the peak-end-rule [Fredrickson and Kahneman, 1993].

Maneuver-Based Ratings

To obtain more detailed and informative ratings, a drive can be divided into smaller parts. Participants then give a single rating for each of those maneuvers or sections [Ellensohn et al., 2020], i.e., "maneuver-based" ratings (MB). This leads to more detailed ratings for different parts of the drive. As a consequence, however, the workload for the participants also increases, making closed-loop driving at the same time more challenging.

Current-Time Ratings

The third method, first used by Cleij et al. [2018], does not ask for a summarizing rating of a specific part of the drive at all, but rather measures the rating that represents the *current* impression of the participant at that point in time, denoted R. Although yielding high detail, this method requires a focus on the evaluation throughout the entire drive, adding to the workload.

Table 2.1: Matrix of possible rating representations and timings. Dashes indicate non-viable rating methods.

		Representation		
		Overall (OR)	Maneuver (MB)	Current-time (R)
Timing	Post-Hoc ($_{PH}$)	OR_{PH}	MB_{PH}	-
	Section-Wise Post-Hoc ($_{SPH}$)	OR_{SPH}	MB_{SPH}	-
	Continuously (t)	$OR(t)$	$MB(t)$	$R(t)$

2.3.2.2 Rating Timing

A second choice concerns *when* the rating is obtained from the participants, as also indicated in Table 2.1.

Post-Hoc Ratings

In a Post-Hoc (PH) rating method, denoted with subscript $_{PH}$, the rating is extracted after the drive [Ellensohn et al., 2020]. This has the benefit of being non-intrusive, such that the rating does not affect any of the closed-loop driving tasks during a drive. Post-hoc ratings can be collected as overall (OR_{PH}) [Kolff et al., 2024b] or maneuver-based ratings (MB_{PH}) [Cleij et al., 2018; Eppink et al., 2023]. As the post-hoc ratings are extracted after a drive, participants will have the possibility to ask for clarification if questions arise. Furthermore, as the rating method does not interfere with the driving task(s), post-hoc ratings can be used in experiments with closed- or open-loop driving.

Section-Wise Post-Hoc Ratings

Section-Wise Post-Hoc (SPH) ratings, denoted with subscript $_{SPH}$, refer to ratings acquired *during* the drive, but after a specific maneuver or event. These ratings can provide a more "in the moment" extraction than the PH ratings, reducing the risk of memory biases. Like PH ratings, SPH ratings can be acquired from experiments involving closed-loop or open-loop driving [Ellensohn et al., 2020]. However, different than for PH ratings is that the participants typically have limited time for providing their rating, because they are expected to continue with the rest of the drive (see Appendix C). A too high frequency of expected ratings could thus hinder the ability to also drive closed-loop. SPH ratings can be combined with a maneuver-based rating [Ellensohn et al., 2020], leading to measurements MB_{SPH}. Furthermore, although not yet performed in literature, an overall rating OR_{SPH} could be obtained, in which the overall impression is obtained at several points in the drive, representing the overall impression of the drive up until each point.

Continuous Ratings

Lastly, drivers can rate the motion continuously [Cleij et al., 2019] throughout the drive reflecting their current PMI at each point in time, resulting in a rating $R(t)$. This method provides unmatched insight into *where* in the simulation (in)congruent motion occurs while recording at a high temporal resolution. Because they are expected to rate continuously, participants will not have the chance to ask for clarifications on the method. This thus requires full task understanding and in-depth briefing and training *before* the experiment. Continuous ratings can also be recorded in distinct maneuvers only [Cleij et al., 2018], yielding ratings $MB(t)$. Similarly, continuous ratings $OR(t)$, representing

Figure 2.5: A rotary rating knob to extract rating values. Adapted from Ellensohn et al. [2019].

the whole drive up until the point of measurement, can be recorded.

To be able to give the rating continuously, a dedicated rating interface is necessary [Cleij et al., 2018], such as a rotary knob (Figure 2.5). Additional dynamics (e.g., a hardware delay) will be present in the recorded ratings, indicated by "Rating Interface" in Figure 2.3. Since the introduction by Cleij et al. [2018], the continuous rating method has been applied in several studies on curve driving [Cleij et al., 2019; van der Ploeg et al., 2020], as well as realistic rural driving scenarios [Ellensohn et al., 2019a; Ellensohn et al., 2020; Ellensohn et al., 2019c].

Because the drivers are expected to continuously evaluate their PMI use the rating knob with the right hand, they cannot be expected to safely operate the steering wheel with the left hand at the same time. Therefore, the continuous rating method can only be employed in open-loop simulations, i.e., using a recording playback. While this allows participants to fully focus on their subjective impression of the motion, the lack of driving control (i.e., the red elements in Figure 2.3) implies that no efference copy can be present. The internal representation must then be generated from the non-inertial motion cues, such as the visuals. Apart from perceptual differences, [Nesti et al., 2016; Valente Pais et al., 2012], the lack of an efference copy could affect the PMI of the driver, because they know less well what motion to expect. An explicit comparison between ratings of closed- and open-loop driving is discussed in Chapter 7, as shown in Figure 2.3.

2.3.2.3 Comparing Between Experiments

Whereas perfect motion can be defined and fixed to a rating scale value, there is no absolute comparison for "imperfect" motion. This is problematic, as participants can have different interpretations of what refers to as the "highly incongruent" rating (Figure 2.4). To avoid this problem of differences in cueing reference, the ratings can be normalized with respected to the highest value that a participant has reached in the whole experiment, such that the data of all participants is always on the same scale [Cleij et al., 2018]. However, it is also possible that participants have the same cueing reference but simply give different ratings, if their internal representation of the vehicle motion is different. In that case, the PMI can also differ.

Comparing between experiments can also reveal relative differences between par-

ticipant *groups*. What participants associate with "highly incongruent motion" might depend on the presented motion, and therefore the capabilities of the motion simulator. Cleij et al. [2019] showed that correcting for these relative experiment differences can be done using a Model Transfer Parameter (MTP), a scaling factor that describes the relative PMI between experiments. More research is necessary to investigate whether a linear MTP can also be used to predict between MIRs between scenario types. For this, Chapters 4, 5, and 6 will compare rating measurements between urban, highway, and rural scenarios, respectively.

2.3.3 Motion Cueing Models

The final step required for the large-scale evaluation of motion cueing methods and solutions *without* subjective simulator testing requires *predicting* their MIRs. Because of the high temporal resolution, the relation between the continuous ratings and objectively calculated mismatch signals between vehicle and simulated motion can be captured in mathematical models, which in turn allow for *predicting* continuous ratings [Cleij et al., 2018; Ellensohn et al., 2019c]. Overall and maneuver-based ratings are of such lower resolution that they are much less suitable to be used in a modeling approach.

Whereas Cleij et al. [2018] successfully predicted continuous ratings of their simplified scenario using a moving average model between the objective mismatches $\Delta\widetilde{S}(t)$ and the measured ratings, this has not (yet) been successfully repeated for more realistic urban, highway, and rural scenario's, a critical lack in understanding that is covered in Chapters 4 (urban), Chapter 5 (highway), and Chapter 6 (rural) of this dissertation.

While predictions of continuous ratings can thus be made, they are more difficult to guide decisions on motion cueing, as they do not provide a single rating value that can be used for decision. Therefore, it is also necessary to *combine* predictive power of the continuous ratings with the simplicity of overall or maneuver-based ratings. This thus requires investigating how these latter ratings are formed based on the continuous impression of motion. For example, Cleij et al. [2018] showed that participants let the most incongruent moment (as visible in the continuous ratings) determine the overall rating. These critical relationships between OR_{PH}, MB_{SPH} and the continuous $R(t)$ ratings are investigated per scenario in Chapter 4 (urban), Chapter 5 (highway), and Chapter 6 (rural) of this dissertation.

2.4 Discussion

2.4.1 Metric Identification

Next to metrics of motion cueing quality (behavioral or perceptual), the survey further revealed several other metrics that are generally of importance. the cost and the participants' 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 as well as simulator sickness for the participant. In the experiment use-case example, an additional focus lies on noise and stability, indicated by the wishes of the experimenter. It is possible that due to practical experience, existing metrics are refined or more metrics are added. The presented methods and results thus aim to serve as a broader framework rather than a solution that applies for all institutes or companies

operating driving simulators, as also different use-cases might be covered. Other institutions or companies are thus invited to share their experiences, as this could also lead to more data for the operators and motion cueing engineers. The current results are inherently limited due to being obtained from a single simulation center (in this case BMW).

Through the presented methods, 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 a future specific 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 and stability, could show to in fact 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 final chosen simulator was indeed a smaller system (BMW's Vega Vector).

In assessing a single survey for a given experiment, it would be recommended, however, to let the individual questions of the survey have more impact on the trade-off (e.g., offline testing capabilities), rather than only considering the average value of a given metric (i.e., "tuning effort") and as a summary for each group.

2.4.2 Limitations and Next Steps

Several limitations in the presented work should be mentioned. First, the trade-off between the identified metrics would benefit from a more objective measurement approach, rather than applying the ordinal scale (1-7) directly from the survey results. Several metrics would be suitable for this due to their quantifiable nature, such as noise, energy consumption, and cost. Others, such as determinism of the system or stability, are much harder to quantify. Here, qualitative categorization of the metrics (i.e., low-medium-high) may be an effective approach.

A trade-off *between* the metrics will always be subjective. It is possible to make this choice by setting lower bounds to eliminate MCAs that do not meet certain key requirements or, if no defined lower bounds exist for the experiment, calculating the "overall quality", by calculating a total cost. This trade-off can also be made more objective by taking the relative importance of the various metrics into account. 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.

As a further outline for future work, several points have been discussed, which will be dealt with in Part II of this dissertation. These include the measuring and predictive modeling across urban, highway, and rural scenarios, which will be discussed in Chapter 4, Chapter 5, and Chapter 6, respectively. The assumption that (models of) ratings of open-loop driving are representative of closed-loop driving will be verified in Chapter 7. As the basis for the models that will be discussed in Part II, Chapter 3 will derive the analytical mismatch relationships for the kinematic configurations available at BMW.

The second part of this work (Chapter 9) will deal with a key open question that has

not been covered so far, which is *how* predictions of MIRs contribute to assessing the total quality of a motion cueing method and how the various other metrics can be quantified. This requires calculating the potential of a motion cueing method *without* actually tuning it. This can be done, for example, by using an extrapolation of the MIRs based on the remaining available workspace [Grottoli et al., 2019] of currently available tuning. Similarly, quantifiable metrics, such as energy consumption or cost, are then also expressed in terms of their potentials. Whereas the present chapter presented the theoretical concept, Chapter 9 will thus cover the practical use of the proposed method as an example trade-off between motion cueing solutions.

2.5 Conclusion

This chapter described new standards to describe the MCA and simulator selection (the "motion cueing method") for future driving simulation experiments with motion cueing. This includes always considering a prediction of the motion cueing *potential* of a method, rather than using existing sub-optimal tuning sets in comparisons. The results of the stakeholder survey further showed that, next to the quality of the motion itself, the total cost, simulator sickness, easy of use, tuning effort, complexity, and stability are typically of high importance in motion cueing experiments. Better motion cueing can thus be obtained when these specific wishes of the various experiment stakeholders are taken into account. Combined, these new standards provide a fairer and more effective method to properly select an MCA and simulator. This ensures that the best motion cueing can be selected for future driving simulator experiments.

3

Kinematics of Motion Systems

Key findings

- The specific forces and rotational rates relations depend on the relative position of the motion subsystems. A correct implementation of the kinematic relations is key for correctly expressing the mismatches between vehicle and simulator motion.
- A yaw-drive located above the hexapod greatly complicates the motion control.
- For large-excitation yaw-drives ($> 20°$), the offset between the Motion Reference Point (MRP) (the point where the motion applies) and the Cueing Reference Point (CRP) (the position where the driver is located) must be corrected for in the motion control to avoid erroneous simulator motion.

This chapter is based on the following publication:

Title: The Importance of Kinematic Configurations for Motion Control of Driving Simulators.
Authors: Maurice Kolff, Joost Venrooij, Markus Schwienbacher, Daan. M. Pool, and Max Mulder.
Proceedings: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Spain, pp. 1000-1006.

Abstract

This chapter describes how the kinematic configuration of a driving simulator's motion system affects the rendered inertial motion. The specific force and rotational rate equations between the point where the motion is applied [Motion Reference Point (MRP)], and the point in which the driver perceives the motion [Cueing Reference Point (CRP)], are derived for three kinematic configurations: (i) a hexapod, (ii) a hexapod with an xy-drive and a yaw-drive below, and (iii) the same system as (ii), but with the yaw-drive on top. The rotational rate equations show that having a yaw-drive on top greatly complicates the motion control. Furthermore, simulation results show that, regardless of the yaw-drive location, the difference between MRP and CRP becomes noticeable for large yaw-drive excitations. For such driving simulators, the positional offset between MRP and CRP can therefore not be ignored, complicating the motion control.

3.1 Introduction

DRIVING simulators are important tools for automotive research as they provide a safe, cost-efficient, and controlled-environment alternative to real vehicle testing. Typical research examples are human-machine interaction [Lindner et al., 2022], human-in-the-loop testing of vehicle safety systems [Chai et al., 2020], and human factors research on automated driving and traffic safety [Bazilinskyy et al., 2020]. When employed with a motion system, its goal is often to mimic the inertial motion of the real vehicle as closely as possible to make the simulation more realistic. This conversion is performed by the MCA [Stahl et al., 2014], which converts the simulated vehicle motion (typically, specific forces and rotational rates) to motion that retains the physical limits of the simulator. Poor motion reproduction can lead to decreased realism [Casas-Yrurzum et al., 2020], simulator sickness [Himmels et al., 2022b], and unwanted adaptation of control behavior [Romano et al., 2019]. All of these consequences can negatively affect simulator testing.

Many motion systems consist of a six DoF hexapod [Ghafarian et al., 2023], also known as *Stewart* [Stewart, 1965] or *Stewart-Gough* [Gough, 1962] platform. To improve the potential motion reproduction, simulators are increasingly equipped with additional motion subsystems to extend the available workspace. This is commonly done using linear xy-drives (e.g., FKFS's simulator [Baumann et al., 2014]), rotational yaw-drives (e.g., BMW's Sirius Vector [Kolff et al., 2022]), or a combination of both (e.g., Toyota's Driving Simulator [Chiew et al., 2008], Renault's ROADS [Fang et al., 2022], and BMW's Sapphire Space [Kolff et al., 2022]). Although the additional workspace can improve the motion reproduction, the complexity of the motion control also inherently increases, as the MCA needs to correctly incorporate the kinematics of the additional motion subsystems.

The kinematic relations between the various simulator motion subsystems depend on where these subsystems are located (e.g., yaw-drive above or below the yaw-drive [Fang et al., 2017]), and affect the specific forces and rotational rates as presented in the simulator cabin. Although the kinematic relations for hexapods are known (e.g., [Bruschetta et al., 2018a]), to the best of our knowledge, a complete, formal derivation of the kinematics of driving simulators with additional motion systems has not been presented yet. Furthermore, there can be a difference in motion between the point where the simulator motion is applied (Motion Reference Point (MRP)) and where the driver is actually seated and perceives the motion (Cueing Reference Point (CRP)), which is often neglected to simplify the control problem. However, if simulators with additional subsystems are used, it is possible that this positional offset cannot be neglected. If this offset is not accounted for through the correct kinematic relations, unwanted perceivable differences could occur between the MRP and the CRP. Thus, the difference between MRP and CRP as a function of the simulator configuration must be investigated.

The contribution of this chapter is threefold: First, the kinematic relations of the specific forces and rotational rates are derived for the MRP. This is done for three general simulator configurations: (i) a hexapod, (ii) a hexapod on an xy-drive and yaw-drive combination, and (iii) the same configuration as (ii), but with the yaw-drive *above* the hexapod. Using these results, the kinematic relations of many other motion simulators can be derived analogously. Second, the equations of motion relating the MRP and the CRP are derived. Third, by simulating the kinematic configurations using sinusoids, the differences in motion between the MRP and the CRP are quantified.

The structure of this chapter is as follows. Section 3.2 describes the motion system kinematic relations. Section 3.3 describes the simulation results, followed by the discussion in Section 3.4, and the conclusion in Section 3.5.

3.2 Kinematics of Motion Systems

3.2.1 Vehicle System

When driving in a simulator, ideally the motion of the simulator comes as close as possible to the real vehicle. The human perception of motion primarily occurs through the vestibular system [Houck et al., 2005], located near the inner ears, defining the Vehicle Reference Point (VRP), see Figure 3.1. The specific forces and rotational rates occurring at this point in the real vehicle are the motion that the simulator should reproduce. An Inertial Measurement Unit (IMU) measures specific forces and rotational rates [Nogueira et al., 2021], which can be in an arbitrary position. The specific forces and rotational rates must be transformed from the position of the IMU to the VRP. The vector $\mathbf{r}_{IV}$ relates these points, see Figure 3.1.

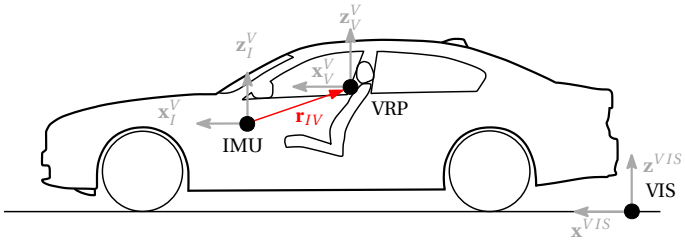


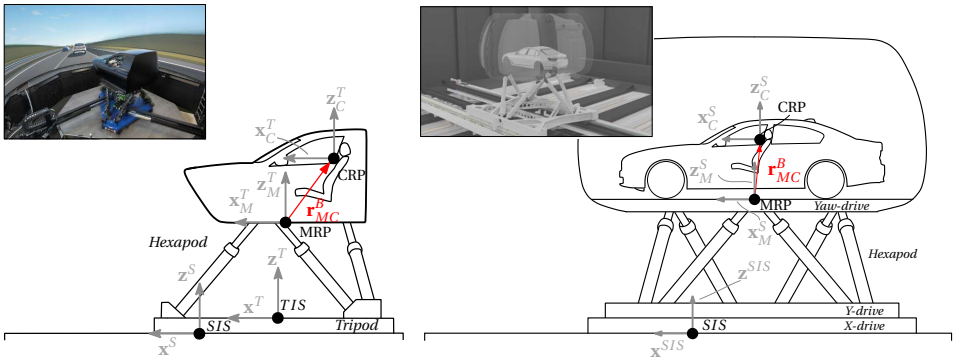
Figure 3.1: The vehicle system, indicating the position of the Vehicle Reference Point (VRP) with respect to the Inertial Measurement Unit (IMU). The IMU measures the vehicle's specific forces and rotational rates. The vehicle's position and attitude are defined in the Vehicle Inertial System (VIS).

3.2.2 Simulator Systems

The aim of the MCA is to convert the vehicle model specific forces and rotational rates, as acting on the VRP, to generate platform motion acting on the human as close as possible to those experienced in a real vehicle. The specific forces and rotational rates are therefore required as inputs of the MCA and are generated by a vehicle model. Therefore, similar to the VRP in the real vehicle, for the simulator the CRP exists, in which the motion is perceived, e.g., see Figure 3.2a. Thus, the motion applied in the MRP must aim at producing the proper motion in the CRP.

The three kinematic configurations studied in this chapter are based on two systems present at BMW Group. First, the Ruby Space (RS, Figure 3.2a) consists of a hexapod on top of a tripod. The latter acts as a combination of an xy-drive (by moving in x and y) and a yaw-drive (by rotating around the z-axis) [Kolff et al., 2021; Kolff et al., 2024b]. Second, the Sapphire Space (SS, Figure 3.2b), consisting of a large xy-drive at the bottom, a hexapod, and a yaw-drive. Both systems therefore have nine DoFs to manipulate the three translations and three rotations of the vehicle mockup. The platform workspace limits are shown in Table 3.1. Next to differences in size, the fundamental difference

between the systems is where the additional yaw rotations are applied. To calculate the resulting forces and rates acting on the vehicle mockup body frame B, as a function of the motion applied in the Simulator Inertial System (SIS), (conveniently located in the geometric center of the lowest simulator component), the contributions of the different DoFs must be combined. Because the hexapod of the Ruby Space is rotated through the tripod rotation, a local Tripod Inertial System (TIS) fixed to the tripod is defined, see Figure 3.2a. In practice, the vector $\mathbf{r}_{MC}^{B}$ between MRP and CRP can also be different per simulator and vehicle mockup combination. To allow for a fair comparison, it is assumed that $\mathbf{r}_{MC}^{B} = [x,\ y,\ z] = [-0.185,\ 0.40,\ 1.4]^{\top}$ for both systems.



(a) The Ruby Space. Tripod actuators are not displayed.

(b) The Sapphire Space.

Figure 3.2: Side views of the motion systems, showing the Motion Reference Point (MRP) and Cueing Reference Point (CRP). These are defined in the Tripod Inertial System (TIS) (superscript T) and Simulator Inertial System (SIS) (superscript S).

Table 3.1: Uni-directional workspace limits of the two motion systems, RS = Ruby Space, SS = Sapphire Space.

	hexapod						xy-drive		yaw-drive
	x_h	y_h	z_h	φ_h	θ_h	ψ_h	x_d	y_d	ψ_d
	m	m	m	deg	deg	deg	m	m	deg
RS	0.28	0.25	0.2	20	20	20	0.75	0.75	20
SS	1.4	1.2	0.8	25	25	35	9.57	7.85	180

3.2.3 Kinematic Chains

3.2.3.1 Six Degrees of Freedom

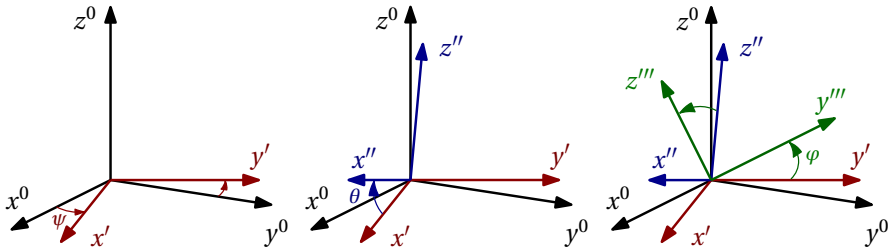
For a six DoF simulator, the attitude of the system with respect to the inertial reference system SIS is defined by the set of three Euler angles φ, θ, and ψ, describing the roll, pitch, and yaw angles, around the x-, y-, and z- axes, respectively. In driving kinematics, the x-axis typically points forwards, the y-axis to the left (seen from the drivers perspective), and the z-axis upwards. The subsequent rotations over each angle are based intrinsically on the newly created coordinate system of the object by the previous rotation,

such that the rotation sequence matters, see Figure 3.3. The most common sequence is $z-y'-x''$ [Diebel, 2006], i.e., the yaw rotation is applied first, followed by the pitch rotation, and then the roll rotation. The transformations due to these angles are calculated using the Euler transformation matrices, where the rotation angles are defined as counter-clockwise positive:

$$\mathbf{R}_x(\varphi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{bmatrix}, \mathbf{R}_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, \tag{3.1}$$

$$\text{and } \mathbf{R}_z(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.2}$$

A clockwise system would result in switched signs of the sin-terms. Commands to the



(a) Rotation 1: ψ around z.

(b) Rotation 2: θ around y'.

(c) Rotation 3: φ around x''.

Figure 3.3: Three consecutive rotations using the order $z-y'-x''$ ($\psi-\theta-\psi$).

motion system are generally defined in the inertial reference system SIS, or relative to its predecessor. To calculate the specific forces (including the gravity) on the cabin in its body-fixed system, one has to rotate the accelerations over the angles ψ, θ, and φ:

$$\begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix}_M^B = \mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta)\mathbf{R}_z^\top(\psi)\left(\begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}_{hex} + \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix}\right)^{SIS}. \tag{3.3}$$

Note that, in order to express the specific forces in terms of the motion in the inertial system, the *inverse* of the rotation sequence must be taken. This equals its transpose, because the rotational matrices are orthogonal (i.e., $\mathbf{R}^{-1} = \mathbf{R}^\top$). Because of the rotations, the specific forces in the body system of the simulator depend on the simulator orientation. This shows how sustained specific forces can be generated: rotations over the angles φ and θ induce sustained lateral or longitudinal specific force, respectively, due to the gravity component ("tilt-coordination", see Stratulat et al. [2011]).

For the rotational rates, a conversion from the Euler to body rates is necessary. Due to the importance of the rotation order, the body rates depend on any previously made rotation. The roll rate does not require a transformation, as it is the last rotation applied.

The pitch rate is next in the sequence and must be transformed over the roll angle. Finally, the yaw rate must be transformed over the roll and pitch angles:

$$\begin{aligned}\begin{bmatrix}\omega_x\\ \omega_y\\ \omega_z\end{bmatrix}_M^B &= \begin{bmatrix}\dot{\varphi}\\ 0\\ 0\end{bmatrix} + \mathbf{R}_x^\top(\varphi)\begin{bmatrix}0\\ \dot{\theta}\\ 0\end{bmatrix} + \mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta)\begin{bmatrix}0\\ 0\\ \dot{\psi}\end{bmatrix}\\ &= \begin{bmatrix}1 & 0 & -\sin\theta\\ 0 & \cos\varphi & \sin\varphi\cos\theta\\ 0 & -\sin\varphi & \cos\varphi\cos\theta\end{bmatrix}\begin{bmatrix}\dot{\varphi}\\ \dot{\theta}\\ \dot{\psi}\end{bmatrix}.\end{aligned} \tag{3.4}$$

The resulting rotational rate vector $\boldsymbol{\omega}^B$ contains the rotational rates of the cabin, i.e., the *body rates* as a function of the *Euler rates*. (3.4) furthermore shows that only for small angles ($< 5°$) the matrix can be linearized to an identity matrix, such that the body rates equal the Euler rates.

3.2.3.2 Nine Degrees of Freedom, Yaw-drive Below Hexapod

Second, a configuration is investigated with an xy-drive and a yaw-drive at the below the hexapod. This configuration corresponds to the BMW's Ruby Space simulator, see Figure 3.2a. As mentioned, it allows for additional translation in x_d and y_d, as well as additional rotation over an angle ψ_d. This latter, secondary rotation thus occurs below the hexapod. As the hexapod base moves with the tripod, the rotation sequence of the whole system is $\psi_d - \psi - \theta - \varphi$. As the two yaw contributions are consecutive, these contributions to the total yaw angle are additive, as they rotate around the same axis. The contribution of the hexapod accelerations do not rotate with the angle ψ_d, as they are defined in the local frame of the tripod TIS, on which the hexapod is located. Therefore, the total specific forces in the body frame become:

$$\begin{bmatrix}f_x\\ f_y\\ f_z\end{bmatrix}_M^B = \mathbf{R}_x^\top(\varphi)\,\mathbf{R}_y^\top(\theta)\,\mathbf{R}_z^\top(\psi)\left(\begin{bmatrix}a_x\\ a_y\\ a_z\end{bmatrix}_{hex}^{TIS} + \mathbf{R}_z^\top(\psi_d)\left(\begin{bmatrix}a_x\\ a_y\\ 0\end{bmatrix}_{tri} + \begin{bmatrix}0\\ 0\\ g\end{bmatrix}\right)^{SIS}\right). \tag{3.5}$$

And similar to (3.4), the rotational rates are:

$$\begin{aligned}\begin{bmatrix}\omega_x\\ \omega_y\\ \omega_z\end{bmatrix}_M^B &= \begin{bmatrix}\dot{\varphi}\\ 0\\ 0\end{bmatrix} + \mathbf{R}_x^\top(\varphi)\begin{bmatrix}0\\ \dot{\theta}\\ 0\end{bmatrix} + \mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta)\begin{bmatrix}0\\ 0\\ \dot{\psi}+\dot{\psi}_d\end{bmatrix}\\ &= \begin{bmatrix}1 & 0 & -\sin\theta\\ 0 & \cos\varphi & \sin\varphi\cos\theta\\ 0 & -\sin\varphi & \cos\varphi\cos\theta\end{bmatrix}\begin{bmatrix}\dot{\varphi}\\ \dot{\theta}\\ \dot{\psi}+\dot{\psi}_d\end{bmatrix}\end{aligned} \tag{3.6}$$

The resulting rotational rate vector $\boldsymbol{\omega}$ contains the rotational rates of the cabin. Note that the presence of a yaw-drive at the bottom thus does not affect the effect of the linearization, i.e., it is identical to (3.4). Thus, for small angles φ and θ, the body rates might approximate the Euler rates.

3.2.3.3 Nine Degrees of Freedom, Yaw-Drive above Hexapod

Finally, the kinematic chain is considered for the same system, but with the yaw-drive above the hexapod (corresponding to BMW's Sapphire Space, see Figure 3.2b). Again, its rotation sequence is fixed, although the yaw-drive rotation is here always the *last* in the chain, i.e., the order is $\psi - \theta - \varphi - \psi_d$. A benefit of this is that the yaw-drive rotation can be fully used to generate yaw motion, as it is unaffected by the rotations of the other DoF. This is opposed to the case with the yaw-drive below, where the yaw motion is always entangled with the roll and pitch motion. An important drawback, however, is that the yaw rotation contributions of the yaw-drive and the hexapod are *not* additive anymore, as these do not rotate around the same axis if φ and θ are non-zero. Therefore, an additional rotation matrix $\mathbf{R}_z(\psi_d)$ is required, such that the specific forces are:

$$\begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix}_M^B = \mathbf{R}_z^\top(\psi_d)\,\mathbf{R}_x^\top(\varphi)\,\mathbf{R}_y^\top(\theta)\,\mathbf{R}_z^\top(\psi)\left(\begin{bmatrix} a_x \\ a_y \\ 0 \end{bmatrix}_{xy} + \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}_{hex} + \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix}\right)^{SIS}. \tag{3.7}$$

Furthermore, the rotational rates become:

$$\begin{aligned} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}_M^B &= \begin{bmatrix} 0 \\ 0 \\ \dot{\psi}_d \end{bmatrix} + \mathbf{R}_z^\top(\psi_d) \begin{bmatrix} \dot{\varphi} \\ 0 \\ 0 \end{bmatrix} + \mathbf{R}_z^\top(\psi_d)\mathbf{R}_x^\top(\varphi) \begin{bmatrix} 0 \\ \dot{\theta} \\ 0 \end{bmatrix} + \mathbf{R}_z^\top(\psi_d)\mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta) \begin{bmatrix} 0 \\ 0 \\ \dot{\psi} \end{bmatrix} \\ &= \begin{bmatrix} 0 & \cos\psi_d & \sin\psi_d\cos\varphi & -\cos\psi_d\sin\theta + \cos\theta\sin\varphi\sin\psi_d \\ 0 & -\sin\psi_d & \cos\psi_d\cos\varphi & \sin\psi_d\sin\theta + \cos\psi_d\sin\varphi\cos\theta \\ 1 & 0 & -\sin\varphi & \cos\varphi\cos\theta \end{bmatrix} \begin{bmatrix} \dot{\psi}_d \\ \dot{\varphi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} \end{aligned} \tag{3.8}$$

This shows a key difference compared to (3.4) and (3.6), because here the transformation matrix includes terms of the yaw-drive as well. Due to the larger allowable rotations of this device, the small angle approximation is not valid anymore. Therefore, it cannot be assumed that the body rates equal the Euler rates. This complicates the motion control, as it must include the non-linear conversion matrix of (3.8). Note that the kinematic relations of other simulator systems can be derived analogously, see Table 3.2.

3.2.4 Reference Point Shift

The previously derived specific force and rotational rate definitions apply in the MRP. As the motion of the simulator is to be calculated in the CRP, the positional offset $\mathbf{r}_{MC}^B$ between these two points should be accounted for. When assuming the simulator is a rigid body, the body rates do not depend on location of the reference point ($\boldsymbol{\omega}^B = \boldsymbol{\omega}_M^B = \boldsymbol{\omega}_C^B$) [Diebel, 2006]. The specific forces require a transformation, however. As vector $\mathbf{r}_{MC}^B$ points from the MRP to the CRP:

$$\mathbf{r}_C^I = \mathbf{r}_M^I + \mathbf{r}_{MC}^I = \mathbf{r}_M^I + \mathbf{T}_{IB}\mathbf{r}_{MC}^B, \tag{3.9}$$

Table 3.2: Overview of typical kinematic structures and their expressions for the specific forces and rotational rates (for counter-clockwise rotations).

Configuration	Specific forces $\mathbf{f}^B$	Rotational rates $\boldsymbol{\omega}^B$
xy-hex-yaw *(Sapphire Space)*	$\mathbf{R}_z^\top(\psi_d)\mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta)\mathbf{R}_z^\top(\psi)\left(\begin{bmatrix}a_x\\a_y\\0\end{bmatrix}_{xy}+\begin{bmatrix}a_x\\a_y\\a_z\end{bmatrix}_{hex}+\begin{bmatrix}0\\0\\g\end{bmatrix}\right)^{SIS}$	$\begin{bmatrix}0 & \cos\psi_d & \sin\psi_d\cos\varphi & -\cos\psi_d\sin\theta+\cos\theta\sin\varphi\sin\psi_d\\ 0 & -\sin\psi_d & \cos\psi_d\cos\varphi & \sin\psi_d\sin\theta+\cos\psi_d\sin\varphi\cos\theta\\ 1 & 0 & -\sin\varphi & \cos\varphi\cos\theta\end{bmatrix}\begin{bmatrix}\dot\psi_d\\\dot\varphi\\\dot\theta\\\dot\psi\end{bmatrix}$
x/y-hex *(Diamond Space)*	$\mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta)\mathbf{R}_z^\top(\psi)\left(\begin{bmatrix}a_x/0\\0/a_y\\0\end{bmatrix}_{xy}+\begin{bmatrix}a_x\\a_y\\a_z\end{bmatrix}_{hex}+\begin{bmatrix}0\\0\\g\end{bmatrix}\right)^{SIS}$	$\begin{bmatrix}1 & 0 & -\sin\theta\\ 0 & \cos\varphi & \sin\varphi\cos\theta\\ 0 & -\sin\varphi & \cos\varphi\cos\theta\end{bmatrix}\begin{bmatrix}\dot\varphi\\\dot\theta\\\dot\psi\end{bmatrix}$
tri-hex (Ruby Space)	$\mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta)\mathbf{R}_z^\top(\psi)\left(\begin{bmatrix}a_x\\a_y\\a_z\end{bmatrix}_{hex}^{TIS}+\mathbf{R}_z^\top(\psi_d)\left(\begin{bmatrix}a_x\\a_y\\0\end{bmatrix}_{tri}+\begin{bmatrix}0\\0\\g\end{bmatrix}\right)^{SIS}\right)$	$\begin{bmatrix}1 & 0 & -\sin\theta\\ 0 & \cos\varphi & \sin\varphi\cos\theta\\ 0 & -\sin\varphi & \cos\varphi\cos\theta\end{bmatrix}\begin{bmatrix}\dot\varphi\\\dot\theta\\\dot\psi+\dot\psi_d\end{bmatrix}$
hex *(Vega Vector)*	$\mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta)\mathbf{R}_z^\top(\psi)\left(\begin{bmatrix}a_x\\a_y\\a_z\end{bmatrix}_{hex}+\begin{bmatrix}0\\0\\g\end{bmatrix}\right)^{SIS}$	$\begin{bmatrix}1 & 0 & -\sin\theta\\ 0 & \cos\varphi & \sin\varphi\cos\theta\\ 0 & -\sin\varphi & \cos\varphi\cos\theta\end{bmatrix}\begin{bmatrix}\dot\varphi\\\dot\theta\\\dot\psi\end{bmatrix}$
yaw-hex *(Sirius Vector)*	$\mathbf{R}_x^\top(\varphi)\mathbf{R}_y^\top(\theta)\mathbf{R}_z^\top(\psi)\mathbf{R}_z^\top(\psi_d)\left(\begin{bmatrix}a_x\\a_y\\a_z\end{bmatrix}_{hex}+\begin{bmatrix}0\\0\\g\end{bmatrix}\right)^{SIS}$	$\begin{bmatrix}1 & 0 & -\sin\theta\\ 0 & \cos\varphi & \sin\varphi\cos\theta\\ 0 & -\sin\varphi & \cos\varphi\cos\theta\end{bmatrix}\begin{bmatrix}\dot\varphi\\\dot\theta\\\dot\psi+\dot\psi_d\end{bmatrix}$

with I short for the inertial SIS frame and B for the body-fixed frame. When taking the time derivative to obtain the velocity:

$$\begin{aligned}\mathbf{v}_C^I &= \frac{\mathrm{d}\mathbf{r}_C^I}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{r}_M^I}{\mathrm{d}t} + \frac{\mathrm{d}(\mathbf{T}_{IB}\mathbf{r}_{MC}^B)}{\mathrm{d}t} \\ &= \mathbf{v}_M^I + \dot{\mathbf{T}}_{IB}\mathbf{r}_{MC}^B + \mathbf{T}_{IB}\dot{\mathbf{r}}_{MC}^B,\end{aligned} \tag{3.10}$$

Note that $(\mathbf{T}_{BI} = \mathbf{R}_x(\varphi)\mathbf{R}_y(\theta)\mathbf{R}_z(\psi))^\top$ for a hexapod, $(\mathbf{T}_{BI} = \mathbf{R}_x(\varphi)\mathbf{R}_y(\theta)\mathbf{R}_z(\psi+\psi_d))^\top$ for Ruby Space, and $(\mathbf{T}_{BI} = \mathbf{R}_z(\psi_d)\mathbf{R}_x(\varphi)\mathbf{R}_y(\theta)\mathbf{R}_z(\psi))^\top$ for Sapphire Space, as derived earlier. Taking the time derivative to obtain the acceleration and transforming to the body system:

$$\begin{aligned}\mathbf{a}_C^B &= \mathbf{T}_{BI}\left(\frac{\mathrm{d}\mathbf{v}_M^I}{\mathrm{d}t} + \frac{\mathrm{d}(\dot{\mathbf{T}}_{IB}\mathbf{r}_{MC}^B)}{\mathrm{d}t} + \frac{\mathrm{d}(\mathbf{T}_{IB}\dot{\mathbf{r}}_{MC}^B)}{\mathrm{d}t} + \mathbf{T}_{IB}\ddot{\mathbf{r}}_{MC}^B\right) \\ &= \mathbf{T}_{BI}\left(\mathbf{a}_M^I + \ddot{\mathbf{T}}_{IB}\mathbf{r}_{MC}^B + \dot{\mathbf{T}}_{IB}\dot{\mathbf{r}}_{MC}^B + \dot{\mathbf{T}}_{IB}\dot{\mathbf{r}}_{MC}^B + \mathbf{T}_{IB}\ddot{\mathbf{r}}_{MC}^B\right) \\ &= \mathbf{a}_M^B + \mathbf{T}_{BI}\ddot{\mathbf{T}}_{IB}\mathbf{r}_{MC}^B + 2\mathbf{T}_{BI}\dot{\mathbf{T}}_{IB}\dot{\mathbf{r}}_{MC}^B + \mathbf{a}_{MC}^B.\end{aligned} \tag{3.11}$$

Because the rotational matrices are orthogonal, the transformation matrix $\mathbf{T}_{IB}$, which is a product of the rotational matrices, is also orthogonal, such that $\mathbf{T}_{IB}^{-1} = \mathbf{T}_{IB}^T$ and therefore $\mathbf{T}_{IB}\mathbf{T}_{IB}^\top = \mathbf{I}$. The derivative of the transformation matrix $\mathbf{T}_{IB}$ can then be expressed through a tensor matrix [Buschmann, 2014; Diebel, 2006; Olivari et al., 2019]:

$$\mathbf{S}(\boldsymbol{\omega}) = \dot{\mathbf{T}}_{IB}\mathbf{T}_{IB}^\top \rightarrow \dot{\mathbf{T}}_{IB} = \frac{\mathbf{S}(\boldsymbol{\omega})}{\mathbf{T}_{IB}^\top} = \mathbf{S}(\boldsymbol{\omega})\mathbf{T}_{IB} \tag{3.12}$$

with:

$$\mathbf{S}(\boldsymbol{\omega}) = \mathbf{S}(\omega_x) + \mathbf{S}(\omega_y) + \mathbf{S}(\omega_z) = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{3.13}$$

The derivative of $\dot{\mathbf{T}}_{IB}$ can be found using the chain rule:

$$\begin{aligned}\ddot{\mathbf{T}}_{IB} = \frac{\mathrm{d}(\mathbf{S}(\boldsymbol{\omega})\mathbf{T}_{IB})}{\mathrm{d}t} &= \dot{\mathbf{S}}(\boldsymbol{\omega})\mathbf{T}_{IB} + \mathbf{S}(\boldsymbol{\omega})\dot{\mathbf{T}}_{IB} \\ &= \dot{\mathbf{S}}(\boldsymbol{\omega})\mathbf{T}_{IB} + \mathbf{S}^2(\boldsymbol{\omega})\mathbf{T}_{IB}.\end{aligned} \tag{3.14}$$

Subsequently, (3.12) and (3.14) can be substituted into (3.11):

$$\mathbf{a}_C^B = \mathbf{a}_M^B + \dot{\mathbf{S}}(\boldsymbol{\omega})\mathbf{r}_{MC}^B + \mathbf{S}^2(\boldsymbol{\omega})\mathbf{r}_{MC}^B + 2\mathbf{S}(\boldsymbol{\omega})\dot{\mathbf{r}}_{MC}^B + \mathbf{a}_{MC}^B, \tag{3.15}$$

where $\dot{\mathbf{S}}(\boldsymbol{\omega}_C)$ contains the time derivatives of the elements in $\mathbf{S}(\boldsymbol{\omega}_C)$, i.e., the rotational accelerations. This expresses the total acceleration in the CRP, which is converted to the total specific force in the CRP by substituting $\mathbf{a}_M^B$ with $\mathbf{f}_M^B$:

$$\mathbf{f}_C^B = \underbrace{\mathbf{f}_M^B}_{\text{(i)}} + \underbrace{\dot{\mathbf{S}}(\boldsymbol{\omega})\mathbf{r}_{MC}^B}_{\text{(ii)}} + \underbrace{\mathbf{S}^2(\boldsymbol{\omega})\mathbf{r}_{MC}^B}_{\text{(iii)}} + \underbrace{2\mathbf{S}(\boldsymbol{\omega})\dot{\mathbf{r}}_{MC}^B}_{\text{(iv)}} + \underbrace{\mathbf{a}_{MC}^B}_{\text{(v)}}. \tag{3.16}$$

This is the main kinematic relation for rigid bodies. It consists of (i) the acceleration of the MRP in the body system, (ii) the tangential acceleration, (iii) the centripetal acceleration, (iv) the Coriolis acceleration, and (v) the relative acceleration between the CRP and MRP [Ellensohn, 2020]. The relation itself does not depend on the kinematic configuration of the motion system. However, the centripetal, tangential, and Coriolis acceleration depend on the rotational rates $\boldsymbol{\omega}_C^B$ of the rigid body, corresponding to (3.4), (3.6), or (3.8), and are thus implicitly configuration-specific. The element $\mathbf{f}_M^B$ contain the specific forces in the MRP in the body frame, which corresponds to (3.3), (3.5), or (3.7).

3.3 Simulations

3.3.1 Testing Procedure

A testing procedure was used to compare the specific forces between the MRP and the CRP, based on (3.16):

$$\begin{aligned}\Delta\mathbf{f}_{CM}^B &= \mathbf{f}_C^B - \mathbf{f}_M^B \\ &= \dot{\mathbf{S}}(\omega)\mathbf{r}_{MC}^B + \mathbf{S}^2(\omega)\mathbf{r}_{MC}^B + 2\mathbf{S}(\omega)\dot{\mathbf{r}}_{MC}^B + \mathbf{a}_{MC}^B.\end{aligned} \tag{3.17}$$

The following motion configurations are tested, based on the three system introduced in Section 3.2.2:

C1: An xy-drive and a hexapod.

C2: An xy-drive, a yaw-drive ($\psi_d = \pm 20°$), and a hexapod.

C3: An xy-drive, a yaw-drive ($\psi_d = \pm 180°$), and a hexapod.

C4: An xy-drive, a hexapod, and a yaw-drive ($\psi_d = \pm 20°$).

C5: An xy-drive, a hexapod, and a yaw-drive ($\psi_d = \pm 180°$).

Comparing conditions C1, C2, and C4 allows for investigating the role of the presence and location of the yaw-drive for an excitation angle similar to that of Ruby Space. C3 and C5 show the extent to which additional yaw-drive excitation affects the difference between MRP and CRP.

Note that the size of the xy-drive and the translational channels of the hexapod are in fact irrelevant: as we look at the difference between MRP and CRP, both scale up equivalently in terms of acceleration if the size of the xy-drive changes. Thus, this has no effect on any further analyses on differences between the points.

A sinusoidal input was applied to the position signal of the various DoF of the considered motion configurations of the shape:

$$p(t) = A\sin(\tilde{\omega}t + \phi), \tag{3.18}$$

with its derivative as the velocity signal:

$$v(t) = \tilde{\omega}A\cos(\tilde{\omega}t + \phi), \tag{3.19}$$

and the double derivative as the acceleration signal:

$$a(t) = -\tilde{\omega}^2 A \sin(\tilde{\omega} t + \phi), \tag{3.20}$$

with the frequency $\tilde{\omega}$ in rad/s, and the phase ϕ in rad. The amplitudes A were chosen to fit the workspace of the simulator in each DoF, i.e., the positional workspace limits of Table 3.1. The amplitudes of the rotational channels were set smaller ($\pm 10°$, a typical angle used for tilt-coordination [Stratulat et al., 2011]) than the rotational limits of the motion systems (See Table 3.1), as hexapods cannot fully rotate in all three rotational DoFs at the same time [Ellensohn, 2020] due to their coupled DoFs. To represent representative motion of a vehicle, such as a 90° corner taken in a 3 s time span [Kolff et al., 2024b], the frequency was set at $\tilde{\omega} = \frac{\pi \cdot 90°}{180 \cdot 3} = 0.5$ rad/s in all motion channels.

To avoid situations where contributions of the sinusoids may (partially) cancel each other out, the phases were set different in each DoF: $\phi = 2\pi \cdot [\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{4}{9}, \frac{5}{9}, \frac{6}{9}, \frac{7}{9}, \frac{8}{9}, \frac{9}{9}]$ for $[x_h, y_h, z_h, \phi_h, \theta_h, \psi_h, x_d, y_d, \psi_d]$. The sinusoids were simulated for 20 s (Figure 3.4).

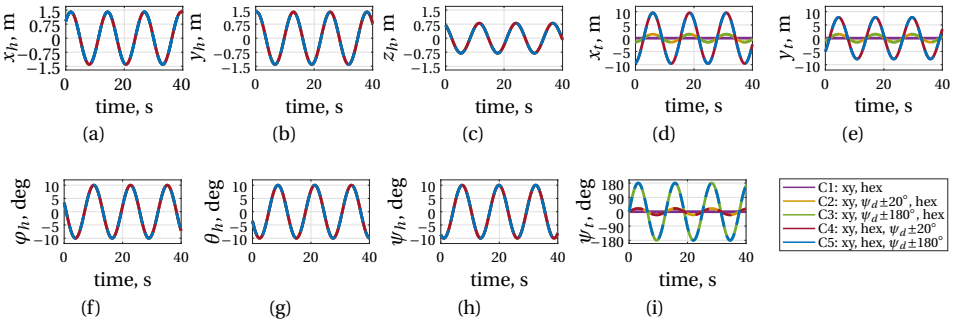
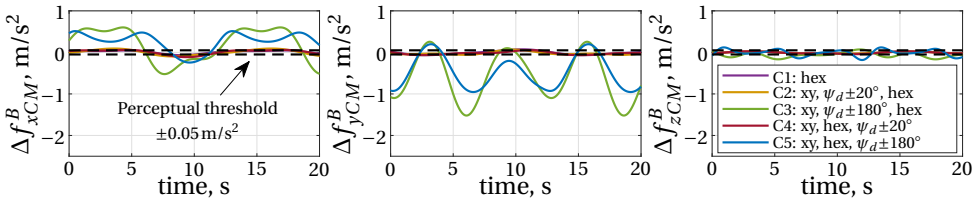


Figure 3.4: Forcing functions of the nine Degrees of Freedom (DoFs), simulated for $\omega = 0.5$ rad/s.

3.3.2 Results

Figures 3.5a-3.5c show the difference in simulated specific forces for f_x, f_y, and f_z, respectively, between the MRP and the CRP. The rotational rates ω are not shown, as for a rigid body these are unaffected by a reference point shift, see Section 3.2.4, and thus the difference between the MRP and the CRP is zero. The positive and negative perceptual threshold are plotted as dashed lines, corresponding to values of $\pm 0.05\,\mathrm{m/s^2}$ [Reymond and Kemeny, 2000]. Differences above these thresholds can be expected to lead to perceivable differences between the MRP and the CRP. For conditions C1, C2, and C4, differences between the reference points are present, but only sometimes marginally exceed the perceptual threshold. For conditions C3 and C5, corresponding to the large yaw-drive excursions of $\pm 180°$, the difference becomes noticeable and reaches its largest difference in f_y ($-1.5\,\mathrm{m/s^2}$). Some care should be taken with the interpretation, as a higher amplitude of the sinusoidal forcing functions between the conditions also implies that the Euler angle derivatives are different between the conditions. This subsequently results in higher rotational rates. The forcing functions applied to the different motion



(a) f_x (b) f_y (c) f_z

Figure 3.5: The difference between the specific forces simulated in the MRP ($\mathbf{r}^B_{MC} = [0,\ 0,\ 0]^T$) and the CRP ($\mathbf{r}^B_{MC} = [-0.185,\ 0.40,\ 1.4]^T$), for the five considered motion system configurations with $\omega = 0.5$ rad/s. The colors correspond to the conditions defined in Figure 3.4, the black dashed lines are the perceptual threshold of $\pm 0.05\,\mathrm{m/s^2}$ from Reymond and Kemeny [2000].

systems are therefore not the same. However, this shows a inherent property of larger yaw-drives, as these allow for, and practically will have larger rotational rates.

To generalize the results of the yaw-drive excitation, the results are generated for the complete range of yaw-drive amplitude A_{ψ_d}, corresponding to the maximum excitation of the yaw-drive, between 0° and ±180°. To quantify the overall difference between the MRP and the CRP specific force signals in x, y, and z direction, the norm [Pouliot et al., 1998] is used:

$$\Delta f_n(t) = \sqrt{(\Delta f^B_{xCM})^2 + (\Delta f^B_{yCM})^2 + (\Delta f^B_{zCM})^2}. \tag{3.21}$$

For each yaw-drive angle, the maximum absolute values of the norm values ($\max[|\Delta f_n(t)|]$) are calculated, i.e., the largest occurring mismatch, see Figure 3.6. These are plotted as function of the yaw-drive amplitude A_{ψ_d} for the configurations with the yaw-drive below (grey) and above (black). The perceptual threshold is again plotted as a dashed line, of which also the norm of the three directions is calculated, such that the line corresponds to $0.087\,\mathrm{m/s^2}$. The figure shows that for C1 (without a

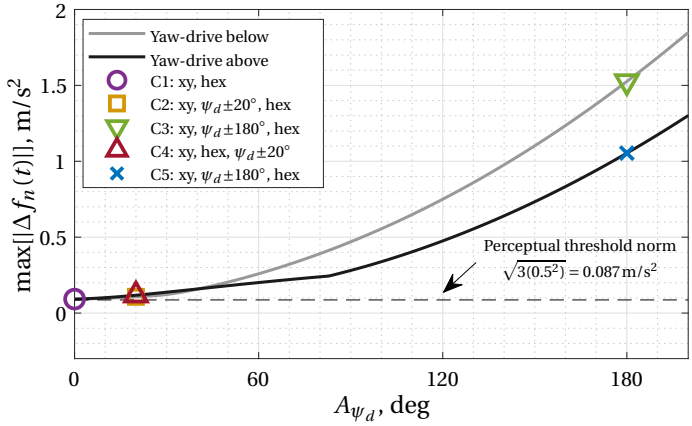


Figure 3.6: Maximum norm of the difference in specific force between the Motion Reference Point (MRP) and the Cueing Reference Point (CRP), as a function of yaw-drive forcing function amplitude.

yaw-drive), the difference between the MRP and the CRP can already be noticed, as it

lies above the perceptual threshold at $0.09\,\mathrm{m/s^2}$. When a yaw-drive is added and excites to $\pm 20°$, corresponding to the conditions C2 and C4, the differences are still limited at $0.11\,\mathrm{m/s^2}$ and $0.12\,\mathrm{m/s^2}$. At larger yaw-drive angles, the norm difference increases more than linearly. At the conditions C3 and C5, with a $\pm 180°$ yaw-drive, the norm differences are $1.05\,\mathrm{m/s^2}$ and $1.53\,\mathrm{m/s^2}$. This will lead to erroneous accelerations perceived in the simulator.

3.4 Discussion

The simulation results show that the configurations with the large excitation yaw-drive ($\pm 180°$), i.e., C3 and C5, lead to large differences in the specific forces between the MRP and CRP. For simulators with extended yaw-drive capabilities, these results emphasize the crucial importance of utilizing the correct reference point in an MCA. Incorrect reference point selection will directly lead to significant spurious specific forces, which can affect driving simulator experiments and increase the occurrence of simulator sickness. For smaller yaw-drive angles ($\pm 20°$, C2 and C4) or no yaw-drive at all (C1) the differences are smaller ($0.11\,\mathrm{m/s^2}$, $0.12\,\mathrm{m/s^2}$, and $0.09\,\mathrm{m/s^2}$, respectively), but still slightly above the perception threshold ($0.087\,\mathrm{m/s^2}$). This shows that even for such smaller systems, the difference cannot be ignored in the current simulator configurations.

Although condition C5, with the $\pm 180°$ yaw-drive above leads to smaller differences than C3, *both* show large deviations between the MRP and the CRP. A yaw-drive at the bottom shows an another, important benefit: for small roll and pitch angles ($5 - 10°$ [Diebel, 2006]) of the hexapod, see (3.6), the system can be linearized. This reduces the complexity compared to the yaw-drive on top, as here the additional contribution of the yaw-drive (see (3.8)) implies that the same linearization procedure cannot be made. This could be of great impact on MPC MCAs [Lamprecht et al., 2021], as the linearization can lead to a simplified MCA. For example, in an MPC MCA, the non-linear structure of (3.16) could be avoided, resulting in an easier control problem. A benefit of a yaw-drive at the top is that its contribution to the yaw rate does not require a transformation between ψ_d and ω_z. Which configuration is superior might thus depend on the properties of the MCA under consideration.

As a note on limitations, it must be stressed that the presented results on the deviations between the MRP and the CRP are a function of several variables. First, note that the perceptual thresholds used in this chapter by Reymond and Kemeny [2000] were measured unidirectionally, which are likely higher when measured under motion in all directions. In practice the deviation for the smaller systems (C1, C2 and C4) might therefore be of no practical meaning. The specific forces are also directly a function of the distance between the MRP and the CRP ($\mathbf{r}^B_{MC}$), see (3.16). Most simulator systems and mock-up combinations at BMW (See Kolff et al. [2022]) are similar in this regard, such that the values of $\mathbf{r}^B_{MC}$ in this chapter are representative. If a simulator is considered that is smaller than the one presented in this chapter, the degree of erroneous specific forces also decreases. Therefore, the results and interpretations presented in this chapter should be used as guidelines, rather than a one-fits-all answer.

3.5 Conclusion

This chapter presented the derivation of the kinematic relations (specific forces and rotational rates) of three types of driving simulators as currently in use at BMW: a hexapod, a hexapod and an xy-drive with a yaw-drive below, and the same system, but with the yaw-drive on top. The results show that increasing the excitation of a yaw-drive increases the difference between the Motion Reference Point (MRP) (the point where the motion is applied) and the Cueing Reference Point (CRP) (where the driver perceives the motion). For a hexapod, with or without a $\pm 20°$ yaw-drive, this difference is close to the perceptual threshold, such that the motion control of the simulator might be simplified by not correcting for the difference between MRP and CRP. For systems with a $\pm 180°$ yaw-drive, such large differences between the motion in the MRP and the CRP are present, that erroneous specific forces in the order of $1\,\mathrm{m/s^2}$ are present. Here, the motion control should account for the difference in MRP and CRP, complicating the motion control. This finding is irrespective of whether the yaw-drive on top or below the hexapod, although the difference between MRP and CRP are larger when the yaw-drive is below the hexapod. However, when the yaw-drive is on top, the resulting non-linear relations to calculate the body rates greatly increase kinematic complexity, which is a crucial factor for state-of-the-art MPC-based MCAs. With increasing kinematic complexity of state-of-the-art driving simulators, this increased understanding of the kinematic relations and the effects of the MCA reference point position is crucial knowledge for engineers and scientists testing and developing intelligent vehicles.

Predicting Motion Cueing Quality

II

4

Incongruences in Urban Simulations

Key findings

- In urban simulations, subjective ratings of the motion cueing are dominated by the lateral and longitudinal specific force mismatches.
- A linear rating model can predict a high degree (90%) of the measured continuous rating variance.
- The better the motion cueing is rated, the less reliable these ratings are, and vice versa.
- The overall rating of motion cueing in an urban drive is dominated by the worst rating *during* that drive.

This chapter is based on the following publications:

Title:	Quality Comparison of Motion Cueing Algorithms for Urban Driving Simulations.
Authors:	M. Kolff, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder.
Proceedings:	Driving Simulation Conference 2021 Europe, Munich, Germany, pp. 141–148.

Title:	Reliability and Models of Subjective Motion Incongruence Ratings in Urban Driving Simulations.
Authors:	M. Kolff, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder.
Journal:	IEEE Transactions on Human-Machine Systems, 54(6), pp. 634-645, 2024.

Abstract

In moving-base driving simulators, the sensation of the inertial car motion provided by the motion system is controlled by the Motion Cueing Algorithm (MCA). Due to the difficulty of reproducing the inertial motion in urban simulations, accurate prediction tools for subjective evaluation of the simulator's inertial motion are required. In this chapter, an open-loop driving experiment in an urban scenario is discussed, in which 60 participants evaluated the motion cueing through an overall rating and a continuous rating method. Three MCAs were tested that represent different levels of motion cueing quality. It is investigated under which conditions the continuous rating method provides reliable data in urban scenarios through the estimation of Cronbach's Alpha and McDonald's Omega. Results show that the better the motion cueing is rated, the lower the reliability of that rating data is and the less the continuous rating and overall rating correlate. This suggests that subjective ratings for motion quality are dominated by (moments of) incongruent motion while congruent motion is less important. Furthermore, through a forward regression approach it is shown that participants' rating behavior can be described by a first-order low-pass filtered response to the lateral specific force mismatch (66.0%), as well as a similar response to the longitudinal specific force mismatch (34.0%). By this better understanding on the acquired ratings in urban driving simulations, including their reliability and predictability, incongruences can be more accurately targeted and reduced.

4.1 Introduction

URBAN driving is an important use-case in driving simulation due to its high importance in vehicle development. Especially for the design of autonomous vehicles, driving in urban environments proves to be one of the most challenging use-cases. Interactions with the surroundings have a higher level of complexity [Zhan et al., 2016] and the likelihood of motion sickness due to the vehicle movements increases [Irmak et al., 2021; Salter et al., 2019; Turner and Griffin, 2000] compared to other scenarios. Driving simulators offer a unique ability to support the development of vehicle technologies by creating safe and repeatable test conditions [Bruschetta et al., 2018a]. Many driving simulators are equipped with a motion system to recreate the inertial motion of the simulated vehicle as closely as possible through the reproduction of its specific forces and rotational rates. This conversion is performed by the MCA. Especially for urban driving, with its characteristic sharp curves, roundabouts and lane changes (strong lateral motion) and frequent decelerations/accelerations (strong longitudinal motion) [Qazani et al., 2020], the workspace-constrained motion system can often not (fully) reproduce the reference motion [Ellensohn, 2020], such that mismatches occur. Not all mismatches are necessarily problematic, however, since some can go unnoticed by the driver [Berthoz et al., 2013]. Only when a driver notices a deviation between their expectation of the real vehicle motion and what they actually perceive can the simulator motion be considered *incongruent* [Cleij et al., 2018]. In an urban simulation, the presence of incongruences combined with the strong visual stimuli can induce relatively high simulator sickness levels [Himmels et al., 2022b]. Understanding *which* mismatches lead to incongruences is therefore paramount for improving these simulations.

Evaluating the (in)congruence of motion is most commonly based on subjective evaluations obtained from drivers. Such subjective ratings provide a direct measurement of the perceived quality of the presented motion cueing. Thus, they are crucial when design choices in motion cueing have to be made for (upcoming) driving simulator experiments, such as selecting a simulator, motion cueing algorithm, and/or MCA parameters. Several different subjective rating methods exist. For example, it is possible to extract an overall rating that summarizes a single maneuver [Cleij et al., 2018] or a whole drive [Biemelt et al., 2021; Rengifo et al., 2021]. A problem with these subjective rating methods is that they can only be obtained when the motion cueing is tested by human test drivers. In practice, it is not realistic to obtain statistically relevant rating data for all possible variations of motion cueings. Furthermore, some novel MCAs in development might not even be testable in a simulator yet. Only with an understanding of the relative importance of the various mismatch channels can attempts to improve the motion cueing be performed with a focus on the most critical mismatches. Thus, *predicting* subjective ratings would be a crucial development, which requires models on the expected subjective rating data. However, the subjective ratings methods that are generally applied in simulator driving experiments (e.g., maneuver-based and overall ratings) are often not of sufficiently high resolution that they can be used for extracting models. Cleij et al. [2018] therefore introduced a *continuous* rating method: while being driven around, drivers continuously give a rating that aims to reflect their impression at each point in time. The method has since been used in van der Ploeg et al. [2020] (same scenario as Cleij et al. [2018]), Ellensohn et al. [2020],Ellensohn et al. [2019],Ellensohn et al. [2019]

(rural scenarios), and Cleij et al. [2019] (rural-urban scenario). These continuous ratings, with their high temporal resolution, *do* allow for modeling how objective motion mismatches relate to perceived motion incongruences. Thus, they are the missing link in predicting motion cueing quality for driving simulator experiments.

Cleij et al. [2018] showed that for a scenario with three basic maneuvers (braking/acceleration, cornering and these combined), participants are generally able to successfully perform the continuous rating task and provide useful data. The latter was investigated by estimating the reliability of the data through the estimation of Cronbach's Alpha, although the relationship between the (in)congruence of the motion and the associated reliability has not yet been investigated. Furthermore, Cleij et al. [2018] showed that the *worst*-rated segment of the maneuver correlates most with the overall rating of that maneuver. This gives rise to the hypothesis that incongruent motion generally shapes the overall impression of drivers more than congruent motion. Whether this holds for longer drives (containing multiple maneuvers), where the worst-rated maneuver would also correlate most to the overall rating of this complete drive, is unknown, as the overall ratings could be biased through short-term memory effects, such as the serial position effect [Murdock Jr, 1962] or the peak-and-end-rule [Fredrickson and Kahneman, 1993]. Finally, Cleij et al. [2018] showed that their continuous rating data can be described by a moving average filter of weighted lateral and vertical specific force, as well as roll and yaw rotational rate mismatches terms. However, Ellensohn et al. [2019] showed that such moving average dynamics are not sufficient to predict the ratings in a more complex and longer rural scenario. It is thus unknown what model structure should be used for realistic urban scenarios, and what relative weightings best describe the data, as this could be different for each scenario. Due to the strong longitudinal motions in urban driving, it can be hypothesized that these motions strongly affect the ratings, in contrast to the findings of Cleij et al. [2018].

This chapter presents four contributions. First, it investigates whether the continuous rating method of Cleij et al. [2018] yields useful results for a realistic urban driving scenario. Second, it examines whether a general relation exists between the maximum of the continuous ratings in each maneuver and the overall ratings for a long and realistic urban scenario, in contrast to the short scenario described in Cleij et al. [2018]. Third, the relation between the ratings and their reliability is investigated through the estimation of Cronbach's Alpha and McDonald's Omega. The latter has shown to provide better estimates of reliability, as Cronbach's Alpha is known to underestimate reliability [Sijtsma, 2009]. Fourth, a rating model is developed, which was fit on the mismatch signals; these signals are selected based on their contribution to the fit.

To support these contributions, this chapter uses data from a driving simulator experiment in a realistic urban scenario with 60 participants, in which both continuous and overall ratings were recorded. Three MCA settings were tested: (i) a classical washout algorithm *without* tilt-coordination, with large mismatches in the longitudinal and lateral specific forces, expected to provide low motion cueing quality; (ii) the same algorithm *with* tilt-coordination, with smaller specific force mismatches (medium quality), and (iii) an optimization-based algorithm with perfect prediction, best able to reproduce the specific forces (highest quality). This range of motion cueing settings allows for a better understanding of the impact of (in)congruent motion on reliability and predictability.

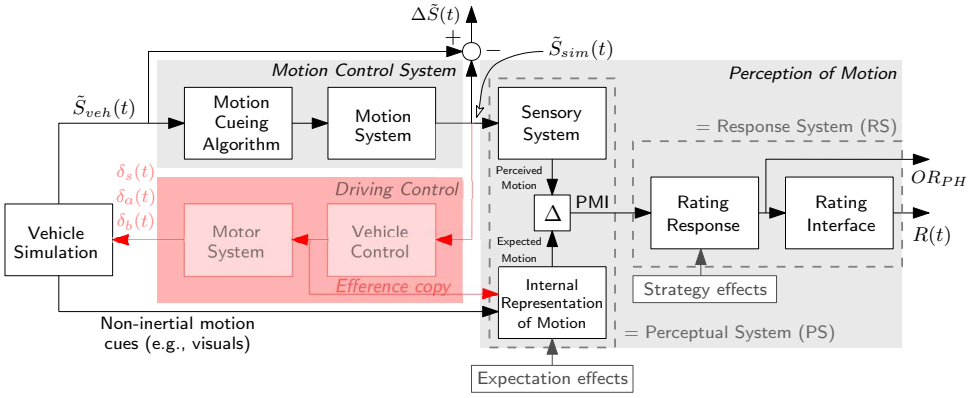


Figure 4.1: Block diagram of the rating process. PMI = Perceived Motion Incongruence, MIR = Motion Incongruence Rating. In the present experiment, the latter is extracted using a continuous rating $R(t)$ and an overall rating OR_{PH}. The red path represents the closed-loop driving control, not active in the present experiment.

The chapter is structured as follows. Section 4.2 introduces the rating task, reliability estimates, and the modeling method. The experiment set-up is explained in Section 4.3. Results are presented in Section 4.4, followed by a discussion in Section 4.5. Concluding remarks are given in Section 4.6.

4.2 Methods

4.2.1 Rating Task

In the experiment, participants were driven around passively (referred to as "open-loop"), rather than driving themselves. Their task was to evaluate how well the perceived inertial motion in the simulator matched to what they would expect to feel from the simulated vehicle, i.e., their PMI [Cleij, 2020]. A block diagram of the human rating process in such tasks is shown in Figure 4.1. As participants do not know exactly what the vehicle would feel like in a particular situation, they must form an internal representation [Stassen et al., 1990] of the expected motion based on non-motion cues (such as the visuals) of the simulation. This internal representation can be affected by "expectation effects", i.e., the participant's level of experience with the task (driving) and with the vehicle that is simulated. While the simulator motion is identical for all test drivers, the *expected* motion signal can thus be different for each participant. The simulator motion is perceived through the human vestibular and proprioceptive systems, indicated as "Sensory System". The internal representation and sensory system combined are indicated in Cleij et al. [2018] as the "Perceptual System (PS)". Furthermore, most driving simulations would drive closed-loop. In that case, an additional path representing the driving control would be present (Figure 4.1, red elements). Given the open-loop driving in the present experiment, this path is thus not active.

The PMI defines a participant's impression of what is (in)congruent, and would be the most useful quantity to measure. It is, however, internal to the human and not directly measurable. Instead, Cleij et al. [2018] proposed to measure a subjective Motion

Incongruence Rating (MIR) that represents the PMI. The Response System (RS) between the PMI and MIR can include the rating strategy, which can vary between drivers, as well as any dynamics of the rating interface. In the case of continuous ratings, the MIR is typically given through a rotary knob that can be adjusted at any time, resulting in a time signal $R(t)$. After each run, an overall rating representing the overall impression is given verbally, yielding a single rating measurement OR_{PH}. The subscript $_{PH}$ denotes post-hoc, as the rating is taken after the completion of the drive. For both methods, the MIR varies between 0 and 10, in steps of 1, where 0 indicates "fully congruent motion" and 10 indicates "highly incongruent motion" [Cleij et al., 2019; Ellensohn et al., 2019a; Ellensohn et al., 2020; Ellensohn et al., 2019c]. Based on earlier experience with participants, it is expected that, especially for the continuous ratings, the RS can be affected by a number of rating "strategy effects":

1. *Task motivation* describes the willingness to focus (on the motion) and actively perform the (rating) task [McRuer and Jex, 1967; Mulder et al., 2018].

2. *Cueing reference* refers to what values drivers apply for the given incongruences, which depends on which PMI-level they associate with the maximum (10) MIR score. In Cleij et al. [2018], participants were shown the full range of the incongruences before the experiment. In the present experiment, they were presented with a false cue in the training sessions to anchor to the highest MIR (10).

3. *Anticipation* can occur when incongruences of upcoming maneuvers are expected based on previous drives or from recognizing that a certain MCA setting is active.

4. *Task understanding* of the participant that only the PMI is to be evaluated, and no other motion-related phenomena (e.g., visual motion, sound or vibrations).

4.2.2 Reliability

Recordings of continuous ratings over various conditions yield a collection of rating time signals $R_{cjp}(t)$, with c the condition, j the condition repetition and p the participant. If along one of these elements the average is taken, this element is taken out of the subscript, such that, for example, $R(t)$ represents the rating of the average participant across all repetitions in a given condition.

In the experiment described in the current chapter, each run lasted 255 s, with continuous rating data being recorded at 100 Hz ($\Delta t = 10$ ms); each recording R_{cjp} contains $N = 25{,}500$ samples. In psychometric theory, the *total score* is the sum of the run items $X_{cjp} = \sum_t R_{cjp}(t)$, where $\sigma^2_{X_{cp}}$ is the variance of total scores over multiple repetitions. Theoretically, if an infinite number of identical and independent repetitions were performed by a participant, the average of all total scores would result in the *true score*, i.e., the expected value of the rating: $T_{cp} = E[X_{cjp}]$. Each separate test result is bound to end up with a random, stochastic measurement error $E_{cjp} = X_{cjp} - T_{cp}$. Reliability is defined by how much of the test score variance can be explained by the true score variance [Sijtsma, 2009]. As the true score cannot be determined, only estimations of a lower bound of reliability can be made. Here, the most common method (for continuous ratings [Cleij et al., 2019; Cleij et al., 2018; Ellensohn et al., 2019b]) is by determining Cronbach's Alpha,

which represents a reliability value for each participant p [Cronbach, 1951]:

$$\alpha_{cp} = \frac{J}{J-1}\frac{\sum_j \sigma^2_{cjp}}{\sigma^2_{X_{cp}}}. \tag{4.1}$$

Here, J is the total number of repetitions and σ^2_{cjp} is the variance of the individual samples. The coefficient α is unbounded on the lower side, i.e., $[-\infty < \alpha \leq 1]$, where the upper bound of 1 indicates full reliability. The main assumption in the derivation of Cronbach's Alpha is "tau-equivalence" [Sijtsma, 2009], meaning that all repetitions of a single condition share the same true score. Due to this constraining assumption, the use of Cronbach's Alpha has been criticized [Sijtsma, 2009] as it can lead to underestimations of reliability. As an alternative, McDonald's Omega [Sijtsma, 2009; Trizano-Hermosilla and Alvarado, 2016], as introduced in McDonald [2000], is calculated as:

$$\Omega_{cp} = \frac{(\sum_j \lambda_{cjp})^2}{(\sum_j \lambda_{cjp})^2 + \sum_j (1-\lambda^2_{cjp})}, \tag{4.2}$$

where λ_{cjp} are the factor loadings. McDonald's Omega is in the same range as Cronbach's Alpha. As a crucial difference, however, McDonald's Omega allows the variation of the true scores, i.e., does not require the assumption of tau-equivalence. This provides a more accurate estimation on reliability than Cronbach's Alpha. Due to the true score variation, McDonald's Omega is always equal to or higher than Cronbach's Alpha [Sijtsma, 2009]. The factor loadings λ_{cjp} were determined using `factoran` in MATLAB R2018b, yielding Ω_{cp} using (4.2).

4.2.3 Explanatory Model

4.2.3.1 Model Selection

To develop a response system model (Figure 4.2), a Multiple-Input-Single-Output (MISO) AutoRegressive eXogenous (ARX) model is fitted. Its polynomial relationships $\frac{B_m(z)}{A(z)}$, with the discrete-time complex variable z, represent the transfer functions $H_m(z)$ between the measured mismatches $\widetilde{P}_m(t)$ (inputs) and a modeled rating signal $\tilde{R}(t)$ (output):

$$\tilde{R}(t) = \frac{1}{A(z)}\epsilon(t) + \sum_m \frac{B_m(z)}{A(z)}\widetilde{P}_m(t), \tag{4.3}$$

with polynomials of the form:

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{n_a} z^{-n_a}, \tag{4.4}$$

$$B_m(z) = b_{m,1} z^{-1} + b_{m,2} z^{-2} + \ldots + b_{m,n_b} z^{-n_b} \tag{4.5}$$

Here, m represents the channel of the mismatch, e.g., $m \in [f_x, f_y, \ldots]$; n_a and n_b are the orders of the dynamics and $\epsilon(t)$ is the error term reflecting the noise to the system.

The signals $\tilde{P}_m(t)$ are formed by a model of the perceptual system ($\widetilde{PS}$), with the mismatches $\Delta\widetilde{S}_m(t)$ between the vehicle motion $\widetilde{S}_{veh,m}(t)$ and the simulator motion $\widetilde{S}_{sim,m}(t)$ as inputs. The absolute value represents that both positive and negative mismatches result in an increase of the rating. K_m represents the gains of the mismatches.

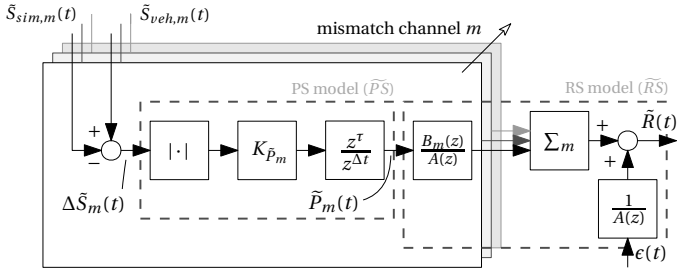


Figure 4.2: Proposed human rating model structure in open-loop driving. The rectangle layers represent the various mismatch channels present in the model.

To express the fit quality, the Variance-Accounted-For (VAF) is determined using $e(t) = R(t) - \tilde{R}(t)$:

$$\text{VAF} = \left[1 - \frac{\sigma^2_{e(t)}}{\sigma^2_{R(t)}}\right] \cdot 100\%. \tag{4.6}$$

The $\sigma^2_{(\cdot)}$-terms indicate the variances. The VAF indicates how much of the variance of the difference between the modelled and measured signals can be explained by the measured signal variance [van der El et al., 2018]. A value of 100% indicates a perfect fit, whereas it is unbounded on the lower side, i.e., $[-\infty < \text{VAF} \leq 100\%]$.

To only select and include the most influential mismatch signals ranked on their contribution to the model quality of $\widetilde{PS}$, a Forward Regression (FR) algorithm is used [Demir et al., 2021]. At the start of the selection, the mismatch signals in the translational acceleration and jerk, as well as the rotational velocity, acceleration, and jerk are considered as possible candidates. These signals only relate to the mismatches in the inertial motion. Any mismatches in the visual motion channels (i.e., the realism of the visuals) are not explicitly considered, as the prime research motivation lies in understanding incongruences as a function of inertial motion mismatches. However, note that the definition of an incongruence considers the difference between the perceived simulator motion and what the participants would *expect* to perceive in the real vehicle. In reality, this expected motion is primarily based on what the participant sees through the visuals of the simulation. Thus, this visual information is implicitly incorporated in the perception of motion incongruence.

Starting with an empty model, each mismatch signal is fit separately to the data. The signal that provides the highest VAF is selected. In the second iteration, all other remaining signals are tested in combination with the signal of the first iteration, selecting the second signal for the model. This process is repeated until no term provides at least an increase of 1% VAF. This method allows for testing all mismatch signals, such that only the most influential signals are included, and unnecessary model complexity is avoided.

The time delay in $\widetilde{PS}$ is modeled by a term $z^{-\tau/\Delta t}$, where τ is the time delay constant. As the ARX-structure cannot estimate a time delay, the FR method is repeated for delay constants ranging between 0 s and 2.5 s with steps of 0.05 s, considering the delay of 1.45 s found by Cleij et al. [2018]. The method is again repeated with orders $N = n_a = n_b$ ranging from 1 upwards until less than a 1% increase in VAF is observed.

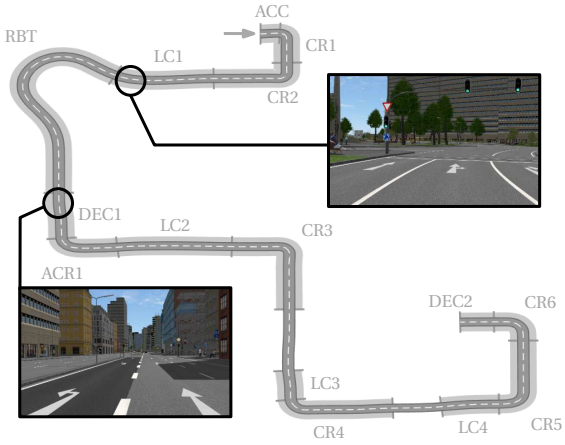


Figure 4.3: Top-down view of the driven route with urban maneuvers with acceleration (ACC), corners (CR), decelerations (DEC), lane changes (LC) and a roundabout (RBT).

4.2.3.2 Parametric Model

The FR ARX method delivers transfer functions $H_m(z)$ in the z-domain to the most influential mismatch channels. These are converted to transfer functions $H_m(s)$ using the bilinear transformation. To obtain more flexibility in the model structure and to acquire explicit parameter values, a parametric model is fitted using the same mismatch channels as estimated by the FR method. As in the FR ARX fit, a fixed $A(z)$ term for all mismatch channels is assumed, such that all mismatch channels pass through the same rating response filter. The model is fitted in the time-domain through the minimization of a cost function:

$$\arg\min_{\Theta} J = \sum_{t} [R(t) - \tilde{R}(t|\Theta)]^2, \tag{4.7}$$

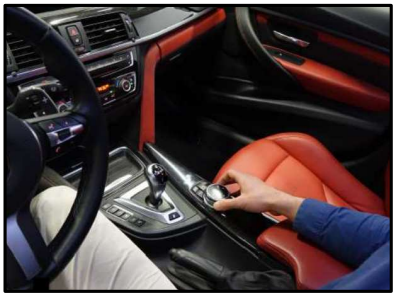
where Θ is the parameter set. In contrast to an ARX fit, this method does not guarantee to find the global optimum. Therefore, 50 iterations are performed with uniformly distributed random numbers between 0 and 3 as initial conditions. The parameter set leading to the lowest cost is then selected.

4.3 Experiment Set-up

4.3.1 Scenario

In the experiment all participants experienced the same recording of a drive through typical urban maneuvers (Figure 4.3), consisting of lateral/yaw maneuvers (corners (CR), lane changes (LC) and roundabout (RBT)) as well as longitudinal maneuvers (accelerations (AC) and decelerations (DEC)). As later runs might induce more anticipation effects, the driving direction (left/right) was shown, together with the vehicle velocity.

(a) The Ruby Space simulator while moving.



(b) A test driver using the rating knob.

Figure 4.4: The experiment set-up (photos adapted from Ellensohn et al. [2019]).

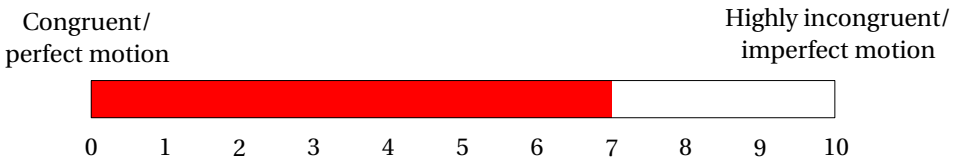


Figure 4.5: The rating scale, ranging from 0 (Congruent motion) to 10 (Highly incongruent motion). It currently displays a rating value of 7.

4.3.2 Apparatus

BMW Group's Ruby Space simulator (Figure 4.4a) was used, with nine DoFs. It consists of a hexapod on a tripod system, where the latter adds additional workspace in longitudinal, lateral and yaw directions. The kinematic relations are summarized in Table 3.2 in Chapter 3. The iDrive knob on the center console was used by participants to give the continuous rating (Figure 4.4b). The 240° projection screen showed the visuals and a "rating bar" [Cleij et al., 2018], displaying the current continuous rating value. The size and color of the rating bar changed from 0 (short, white) to 10 (long, red).

It was checked at the beginning of every experiment session (i.e., for each new participant) whether the participant could comfortably and fully rotate through the rating range (0-10) with one hand movement, which all participants were able to do without problems. During the experiment, participants could rest their right arm on the center console. Feedback obtained from the participants showed that they generally found the knob easy and intuitive to operate. No comments or complaints regarding discomfort and/or difficulty operating the rating knob were made. A typical (fast) transition time of the rating recorded in the experiment required 40 ms per rating step. The rating knob was connected to the CAN bus of the simulator, which is synced with the central simulation software, together with the MCA control data and the motion system. This ensured that the recorded ratings were always synchronized with the motion of the simulator.

4.3.3 Independent Variables

Three MCAs were tested, reflecting different levels of (expected) quality. Their resulting mismatch signals are shown in Figure 4.6, of which the specific forces were calculated using (3.5) and the rotational rates using (3.6). As the first condition, a CWA [Conrad et al., 1973; Reid and Nahon, 1985] was used, where the vehicle motion is distributed over the hexapod (high-frequency) and tripod (low-frequency) channels, and washed-out with second-order high-pass filters. There was no tilt-coordination ("NTC"), such that large mismatches in the f_x and f_y channels (Figure 4.6) are present. This is expected to result in high ratings (i.e., highly incongruent), as tilt-coordination can be used to improve the cueing of sustained longitudinal and lateral specific forces, as long as the tilting rates are not noticeable [Stratulat et al., 2011]. Therefore, it is expected that this condition provides a reference for 'low' quality.

Second, a variant of the same classical washout algorithm was used, with active tilt-coordination ('CWA'). Due to the tilt-coordination, the sustained specific forces in f_x and f_y cause the mismatches to be smaller. The yaw rate remains unaffected, as shown in Figure 4.6f. The tilt-coordination was tuned aggressively such that the roll rate could be noticeable (> 3 deg/s [Reymond and Kemeny, 2000], see Figure 4.6b) to obtain a better reproduction of lateral specific force. The aim of this condition is to represent a state-of-the-art algorithm that can potentially be used in real-time simulations. The condition "CWA" is expected to represent "medium" quality due to two inherent limitations: as it uses linear filters, a CWA must always be tuned to account for the worst-case maneuver, limiting the simulator motion in all other maneuvers. Second, as the algorithm uses causal filters, it cannot incorporate future states in the motion cueing.

As the third condition, an optimization-based algorithm was tested [Ellensohn et al., 2019c], where the simulator motion along the complete recorded drive was optimized offline, the *Oracle* ("ORC"). This algorithm can only be used in open-loop simulations, but allows for the investigation of how the available simulator workspace may be fully exploited. As a result, this condition has the smallest mismatches (Figure 4.6). Therefore, this condition is expected to represent 'high' quality. The rotational rates ω_x and ω_y were below the perceptual threshold (<3 deg/s).

4.3.4 Participants and Procedures

60 subjects participated (50 men, 10 women), all employees of the BMW Group with a European car driver's license B (M = 22.38 yrs, SD = 10.16 yrs) and an average yearly driven distance of M = 18,833 km (SD = 13,207 km). The average age was M = 40.1 yrs (SD = 10.1 yrs). 33 participants had previous experience in driving simulators. Participants provided informed consent. The experiment was approved following BMW's internal ethics review procedure.

The experiment started with two training runs, after which participants drove with either CWA, NTC or ORC. Each condition was repeated three times, yielding nine runs. After every third run a five minute break was taken. Ten participants were unable to finish the experiment due to various reasons.

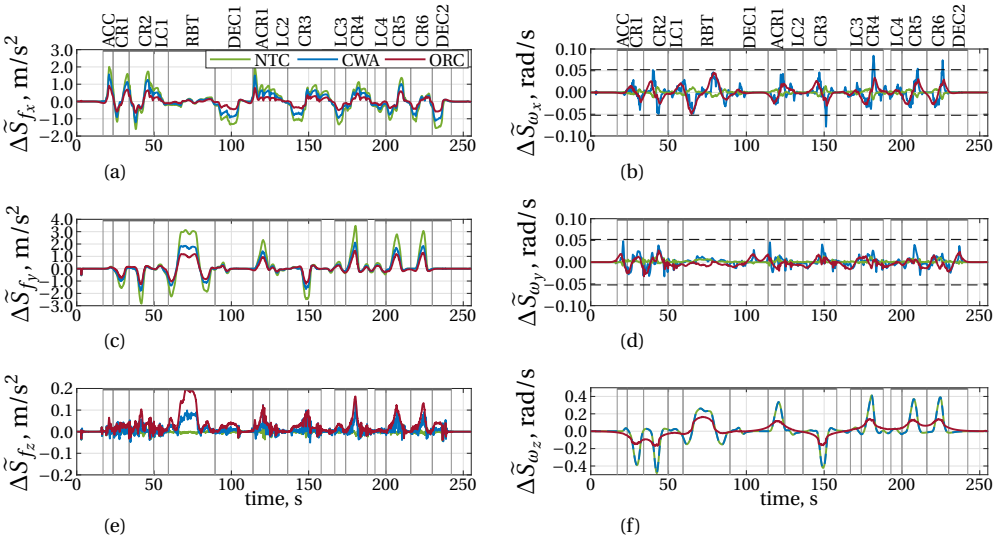


Figure 4.6: Mismatches of the three algorithms (NTC, CWA, ORC). Grey vertical lines indicate the maneuvers, with the grey text entries in (a) the urban maneuvers from Figure 4.3. The dashed horizontal lines in (b) and (d) indicate the rotational threshold of 3 deg/s [Reymond and Kemeny, 2000], relevant for the use of tilt-coordination.

4.4 Results

For nine participants, the experiment could not be finished (eight due to simulator sickness, one due to technical problems). The data of these participants were discarded. Figure 4.7 shows the continuous ratings (left) and overall ratings (right) of the remaining 51 participants. For the former, the lines indicate the mean ratings over the participants and repetitions, $R(t)$; the shaded areas represent the standard deviation. Lower ratings indicate better perceived motion cueing quality. The individual rating data are included in Appendix D.1.

4.4.1 Overall ratings

For the overall ratings, Lilliefors tests ($p \leq 0.01$) showed that the data were normally distributed. When analyzing the means of the ratings, ORC ($\mu = 2.44$, $\sigma = 1.04$) is rated better than CWA ($\mu = 3.89$, $\sigma = 1.79$). The classical washout condition without tilt-coordination, NTC, is rated the worst with $\mu = 5.18$ and $\sigma = 2.17$.

Furthermore, it was checked whether there were significant differences between the three MCA pairs (CWA-NTC, CWA-ORC and NTC-ORC). Within these pairs of independent variable settings, a repeated measures one-way ANOVA was performed. A post-hoc analysis showed very significant differences ($p \leq 0.01$) between the CWA-NTC pair and highly significant differences for the other two pairs ($p \leq 0.001$). A Bonferroni correction was used to correct for multiple comparisons, as each data set is used twice.

4.4.2 Continuous ratings

The continuous ratings are shown in Figure 4.7 (left). The lines indicate the continuous ratings, averaged over the participants. The shaded areas indicate the standard deviation. Here, it is clear that the NTC algorithm performs the worst overall (highest rating), whereas ORC is the lowest rated. The means and standard deviations are for ORC: $\mu = 0.44$, $\sigma = 0.44$, for CWA: $\mu = 0.72$, $\sigma = 0.67$, and for NTC: $\mu = 1.35$, $\sigma = 1.14$. Thus, the order of quality is in line with the overall ratings.

Figure 4.7 also shows the division of the maneuvers. Ratings within each maneuver were analyzed by averaging the rating in that maneuver. As these data were not normally distributed for any of the maneuvers except for 'Corner 4', the Friedman's test with Bonferroni correction was used for all maneuvers to check for significance. Similar to the overall ratings, the mean values are compared between the three MCA pairs to look for significance. These results are shown in Table 4.1.

These results show that when analyzing the pair 'CWA-NTC', the condition NTC is the worst rated of the two, as all other mean values are higher than those of the CWA conditions. Only for the initial acceleration the NTC has a slightly lower mean value, but this difference is not a significant. Furthermore, it was found that only the maneuvers at the start of the run (ACC, CR1, CR2 and LC1), which are relatively weak compared to the rest of the run, do not have significant differences between the two conditions, whereas all other maneuvers do. This indicates that participants generally do prefer the CWA condition with tilt-coordination. Especially the roundabout, where constant tilt-coordination is the only method to produce the sustained lateral acceleration, sees a strong beneficial effect of adding tilt-coordination.

Furthermore, when comparing CWA with ORC, the latter results in lower average ratings compared to the classical washout strategy. In all maneuvers that contained major longitudinal cues (ACC, DEC1, ACR1 and DEC2), no significant differences were found. However, all cornering maneuvers are rated significantly better for ORC than for CWA. This indicates that for the highly dynamic 90° corners, typical for urban environments, there is an advantage when using an optimization-based algorithm. Even though the latter is not possible in real-time, human-in-the-loop driving, optimization-based algorithms, also those that *are* capable of supporting human-in-the-loop driving, might improve cueing quality in urban simulations compared to washout algorithms.

It is clear from the average values that NTC and ORC are the worst- and best-rated conditions, respectively, meaning that their respective differences compared to the CWA algorithm are most interesting. Nevertheless, Table 4.1 also includes this pair for completeness. Significant differences for almost all maneuvers were found.

4.4.3 Rating Relationships

To better understand the relation between the continuous ($R(t)$) and overall ratings (OR_{PH}), the Pearson correlation ρ between OR_{PH} and the maximum of $R(t)$ within each maneuver is calculated (Figure 4.8a, note that the horizontal axis is sorted by the average correlation over the three conditions for increased readability), similar as in Cleij et al. [2018]. Some maneuvers correlate well (for CR6 and CR3 in NTC, $\rho = 0.88$) with the overall ratings, similar to values as found in Cleij et al. [2018]. There is a clear difference between the three conditions, where the lower rated (i.e., better) condition ORC also

Table 4.1: Average MIR per MCA and significance levels per maneuver between each MCA pair. A * is significant ($p < 0.05$), ** is very significant ($p < 0.01$) and *** is highly significant ($p < 0.001$).

		Average $R(t)$			Significance pairs		
Maneuver	Abbr.	CWA	NTC	ORC	CWA-NTC	CWA-ORC	NTC-ORC
Acceleration	ACC	0.31	0.24	0.25	-	-	-
Corner 1	CR1	0.95	1.22	0.56	-	***	***
Corner 2	CR2	1.02	1.48	0.63	-	*	***
Lane change 1	LC1	0.44	0.54	0.32	-	-	*
Roundabout	RBT	1.05	2.85	0.79	***	-	***
Deceleration 1	DEC1	0.40	0.89	0.22	***	*	***
Acceleration + Corner	ACR1	0.94	1.58	0.49	**	-	***
Lane change 2	LC2	0.57	1.04	0.26	**	***	***
Corner 3	CR3	1.08	1.76	0.62	***	*	***
Lane change 3	LC3	0.60	0.87	0.32	**	-	***
Corner 4	CR4	1.30	2.32	0.71	***	***	***
Lane change 4	LC4	0.48	0.91	0.34	*	-	***
Corner 5	CR5	1.09	1.86	0.60	**	**	***
Corner 6	CR6	1.01	1.85	0.60	**	**	***
Deceleration 2	DEC2	0.46	0.91	0.26	**	-	***

correlates the *least* to its own overall ratings and NTC correlates best. To further investigate the relation between the rating and the correlation, the same values of ρ are plotted as a function of the given rating in Figure 4.8b. A positive linear relationship exists between the continuous rating and its correlation with the overall rating. The maneuver with the highest correlation, CR4, predict the overall ratings through the relationship $OR_{PH} = 2.0 + 0.8 \cdot \max[R(t)]$.

4.4.4 Reliability Estimates

Figure 4.9 shows the estimated reliability coefficients for the continuous MIR data for all participants, split over the three conditions. The average reliabilities of NTC, CWA, and ORC are for Ω: 0.79, 0.68 and 0.65 and for α: 0.74, 0.62 and 0.55, respectively. The reliability values per participant are also shown as a function of the corresponding average rating in that condition (hence, the rating averaged over time and averaged over three runs). The overall trend shows that the higher the ratings, the more reliable the obtained data is. This again confirms our expectation that more incongruent motion results in more reliable data, and vice versa.

The figure contains both reliability metrics α and Ω, where Ω is by definition equal or higher compared to α (see Section 4.2.2). The vertical bars show the difference between both metrics. Differences are prominent (up to 0.3) for participants for whom α is low, in line with predictions by Savalei and Reise [2019]. The spread of the reliabilities between participants also becomes larger for smaller average ratings. It is thus at more congruent motion where the use of Ω is beneficial, as it provides a higher lower bound of reliability, avoiding the false conclusion that some participants' data are unreliable at this point.

A regression of the form $r = a - 1/(b\bar{R}_p + c)$ is fit to the data, with a, b, and c the fit coefficients and $\bar{R}_p$, the average rating (over time and repetitions) per participant p. This follows the range of both α and Ω, i.e., $[-\infty < r \leq 1]$ and describes the trend of the

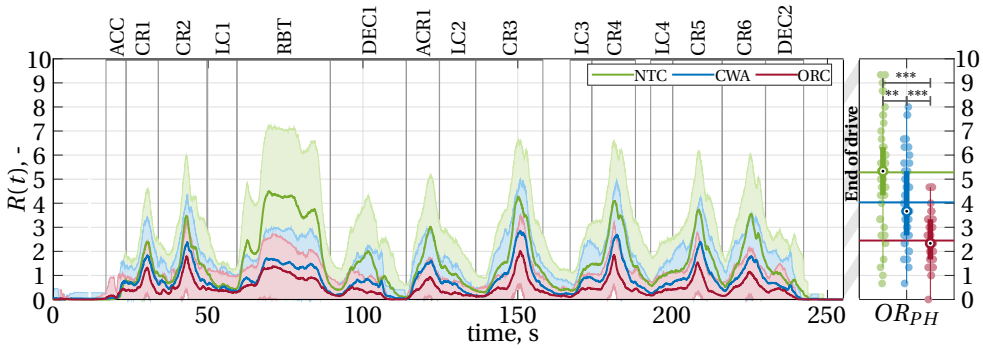


Figure 4.7: Left: The averaged MIRs per MCA (as a function of time in seconds) with the standard deviation displayed as shaded areas. Right: Box plots of the three distributions of the overall ratings; their means are indicated by horizontal lines.

reliability. This function allows for *predicting* reliability based on measured ratings.

Reliability is also calculated for the overall ratings. For the continuous ratings, the presented values represent within-subject reliability. This cannot be calculated for the overall ratings, as per subject and per condition only one data point exists. Therefore, the *between*-subject reliability is calculated, i.e., the reliability of the whole group. The values for Ω are 0.91, 0.89 and 0.73, for conditions NTC, CWA, and ORC, respectively. For the overall ratings, the values of Ω for the between-subject reliability are 0.92, 0.81 and 0.72. These values also indicate a decrease in reliability of the overall ratings, such that the decrease can be considered *inherent* to the difficulty of rating congruent motion, rather than a limitation in the continuous ratings.

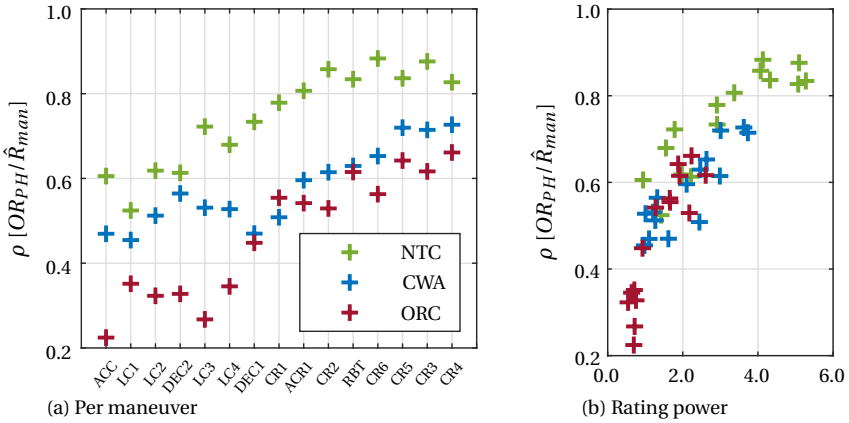
4.4.5 Model Predictions

4.4.5.1 ARX Forward Regression

Results of the ARX FR method are shown in Figure 4.10. Note that the method was applied for the ratings of the three conditions separately (referred to as models a-CWA, a-NTC, and a-ORC), as well as for all conditions grouped together in a single rating signal (a-ALL). The estimated time delay parameter τ was 0 s (a-NTC and a-ALL) and 0.05 s (a-CWA and a-ORC), independent of the model order. For $N = 1$, in all models except a-CWA, the mismatch signal $\widetilde{P}_{f_y}$ (lateral specific force) forms the most important contribution to the model followed by the longitudinal specific force mismatch $\widetilde{P}_{f_x}$. Model a-ORC contains an additional yaw rate term $\widetilde{P}_{\omega_z}$. The model fit on the CWA data (a-CWA) has a different structure: its most important term is the yaw rate mismatch $\widetilde{P}_{\omega_z}$, followed by the longitudinal specific force and yaw acceleration mismatches. Higher orders, as shown in Table 4.2, do not provide a meaningful contribution to the model fits.

To calculate the relative contributions of the most important terms to the (first-order) models, an influence factor is calculated as:

$$I_m = \frac{\sum_t \widetilde{P}_m(t)}{\sum_t \widetilde{P}(t)} \cdot 100\%, \tag{4.8}$$



(a) Per maneuver (b) Rating power

Figure 4.8: Pearson correlation coefficients between the overall ratings (OR_{PH}) and the maximum of the continuous ratings within each maneuver ($\hat{R}_{man}$).

4

with $\widetilde{P}_m(t) = K_{\widetilde{P}_m}|\Delta\widetilde{S}_m(t)|$, see Figure 4.2, and m the mismatch channel. This value represents the relative contributions of the mismatches of the channels, such that the sum of all channels in the model is always 100%. This metric was introduced by Cleij et al. [2018] and thus allows for a direct comparison to their reported values. The values are shown in Table 4.3 under "ARX FR", showing similar contributions of $\widetilde{P}_{f_y}$ and $\widetilde{P}_{f_x}$, except for a-CWA. In the latter, the $\widetilde{P}_{\omega_z}$ also provides a strong contribution at 72.0%. Note that although $\widetilde{P}_{\alpha_z}$ was included in the model a-CWA, its contribution relative to the other terms is negligible.

When repeating the process for higher orders (i.e., $N = 2$, $N = 3$), the same orders of contributions are obtained and negligible increases in VAF are observed, such that it is concluded that first-order dynamics are sufficient to explain the rating data and are thus used for further analysis. The Bode plots in Figures 4.11a-4.11h show the estimated first-order dynamics. The responses resemble those of low-pass filters, such that participants apply smoothing to form their ratings. Furthermore, the phase responses in each model are generally equal, due to the equal $A(z)$ terms in all mismatch channels (as shown in (4.3)). These phase shifts are within $0°$ and $-90°$, indicating that the low-pass filters have positive gains: An increase of the mismatches also leads to an increase in the rating. The phase responses further reveal that possibly an additional response exists at high frequencies, however, with negligible impact on the magnitude ($< 10^{-3}$).

4.4.5.2 Parametric Model

The parametric models (denoted "p-") are based on the estimated dynamics of the ARX FR method. The additional dynamics at high frequencies, as estimated by the ARX FR method, are not included, as it provided only negligible contributions to the magnitude of the estimated dynamics. In addition, as the lack of a time delay cannot be readily

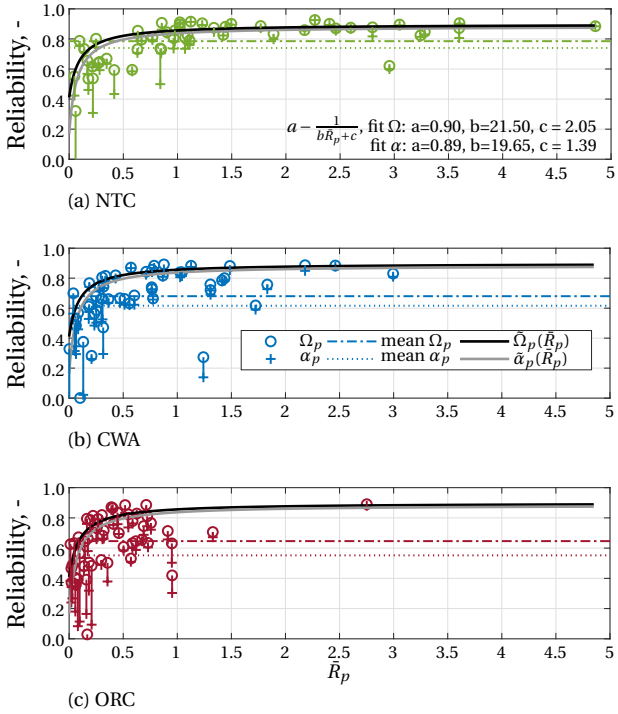


Figure 4.9: Reliability coefficients α_p and Ω_p of all subjects per condition, showing that reliability decreases with lower ratings. The legend in Subfigure 4.9b also applies to the same elements visible in the other subfigures.

explained, a delay is still included in the parametric model; the model is fit in the form:

$$H(j\omega) = \sum_m K_{\tilde{P}_m} \frac{\omega_c}{j\omega + \omega_c} e^{-\tau j\omega}. \tag{4.9}$$

Each mismatch channel has a gain $K_{\tilde{P}_m}$, whereas ω_c is the cut-off frequency and τ the time delay constant, assumed equal in all mismatch channels. The parameter sets that describe the dynamics are $\Theta = [\tau\ \omega_c\ K_{\tilde{P}_{f_y}}\ K_{\tilde{P}_{f_x}}]^T$ for p-NTC and p-ALL, $\Theta = [\tau\ \omega_c\ K_{\tilde{P}_{f_y}}\ K_{\tilde{P}_{f_x}}\ K_{\tilde{P}_{\omega_z}}]^T$ for model p-ORC, and $\Theta = [\tau\ \omega_c\ K_{\tilde{P}_{\omega_z}}\ K_{\tilde{P}_{f_x}}\ K_{\tilde{P}_{\alpha_z}}]^T$ for model p-CWA. The resulting parameters are shown in Table 4.4. Generally, similar values are obtained between the models, indicative of the same rating dynamics and similar weightings being applied by participants between the various conditions.

Notable is that the time delay, as in the ARX FR, is estimated as 0 s, although it is expected that humans would require a processing delay [van der El et al., 2018]. The work of Cleij et al. [2018] also found a non-zero delay of 1.45 s. Their applied model for the RS dynamics was a moving average of the form $(1 + z^{-1} + z^{-2} \ldots + z^{-N_{ma}+1})/N_{ma}$ with a window N_{ma} of 300 samples (= 3 s). If the same model structure is used on our data and the delay $\tau_{ma} = N_{ma}/100$ is estimated for maximization of the cross-correlation between the ratings and moving averaged mismatch signals, a similar value of τ_{ma} =

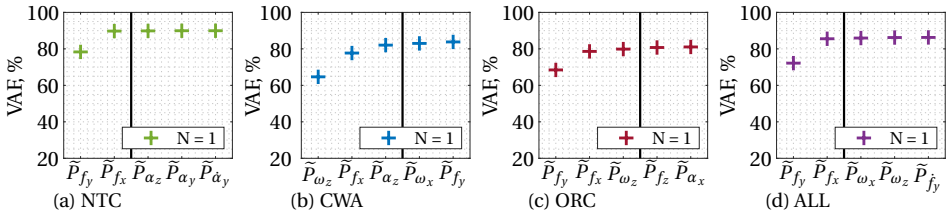


Figure 4.10: Variance-Accounted-For (VAF) values of the AutoRegressive eXogenous (ARX) Forward Regression (FR) method, showing the consecutive contribution of the mismatch signals from left to right for the delay providing the highest VAF. The vertical bars indicate a 1% cut-off rule, such that signals that provide a lower contribution (to the right of the bar) are not considered in further analyses.

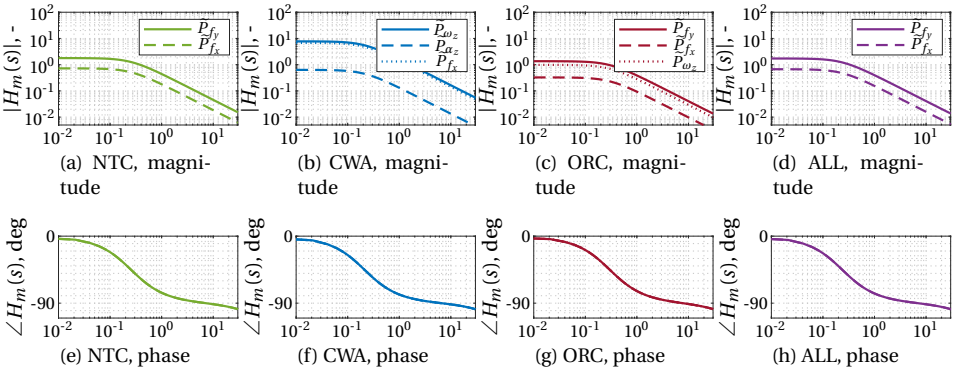


Figure 4.11: Bode diagrams (magnitude (top row) and phase (bottom) of $H_m(s)$ as a function of frequency in radians/sec) of the first-order AutoRegressive eXogenous (ARX) Forward Regression (FR) estimations, showing low-pass filter dynamics in all mismatch channels.

Table 4.2: Variance-Accounted-For (VAF) values of the AutoRegressive eXogenous (ARX) Forward Regression (FR) method, showing the consecutive contribution of the mismatch signal for the first, second, and third order systems. Bold face indicates the selected model components.

	NTC		CWA		ORC		ALL	
	$\widetilde{P}_{(\cdot)}$	VAF	$\widetilde{P}_{(\cdot)}$	VAF	$\widetilde{P}_{(\cdot)}$	VAF	$\widetilde{P}_{(\cdot)}$	VAF
$N=1$	$\boldsymbol{f_y}$	78.38%	$\boldsymbol{\omega_z}$	65.10%	$\boldsymbol{f_y}$	68.99%	$\boldsymbol{f_y}$	72.21%
	$\boldsymbol{f_x}$	89.72%	$\boldsymbol{f_x}$	78.70%	$\boldsymbol{f_x}$	78.49%	$\boldsymbol{f_x}$	87.27%
	α_z	89.92%	$\boldsymbol{\alpha_z}$	82.50%	$\boldsymbol{w_z}$	79.96%	ω_x	87.42%
	α_y	89.93%	w_x	83.42%	f_z	81.22%	w_z	87.42%
	$\dot{\alpha}_y$	89.93%	$\dot{\alpha}_x$	83.66%	α_x	81.53%	$\dot{f}_y$	87.45%
$N=2$	f_y	78.90%	w_z	65.75%	f_y	69.11%	f_y	72.46%
	f_x	89.90%	f_x	78.53%	f_x	78.52%	f_x	87.24%
	α_z	89.93%	f_y	82.56%	w_z	79.74%	f_z	87.25%
	$\dot{f}_x$	89.95%	w_x	83.46%	f_z	81.30%	w_z	87.34%
	α_y	89.96%	$\dot{\alpha}_x$	83.71%	w_x	82.03%	$\dot{f}_x$	87.35%
$N=3$	f_y	78.99%	ω_z	65.83%	f_y	69.26%	f_y	72.59%
	f_x	89.92%	f_x	78.60%	f_x	78.79%	f_x	87.24%
	$\dot{f}_x$	89.97%	f_y	82.67%	w_z	80.25%	f_z	87.28%
	α_z	90.00%	w_x	83.69%	f_z	81.75%	w_z	87.39%
	α_y	90.02%	$\dot{\alpha}_x$	83.88%	w_x	82.27%	$\dot{f}_x$	87.41%

1.88 s is obtained for all conditions grouped together. This shows that although a phase shift is present between the mismatches and ratings, the phase of the estimated low-pass filter response currently captures all of the phase present in the system.

4.4.5.3 Model Fits and Generalizability

The model fits are shown in Figure 4.12. Each figure shows the measured ratings of that condition, as well as how well the four parametric models predict the ratings in terms of VAF. Note that each conditions has two models that were fit on the ratings, the respective condition and the p-ALL model. The two other conditions were fit on the other two conditions and thus provide an insight in the generalizability between the conditions. From these results, it is clear that the model p-CWA generalizes the worst. However, for the CWA data, the p-NTC, p-ORC, and p-ALL models provide reasonable VAF values at 71.8%, 80.4%, and 76.9%, respectively. When considering all three conditions, using only two model terms, the model p-ALL explains most of the measured rating data well. Thus, a surprisingly simple model description can be used to predict the continuous rating data of all three conditions.

A notable exception is maneuver CR3, where all models underestimate the actual ratings as given by the participants. One explanation that followed from participant feedback is that this corner is specifically tight and was taken at a relatively high velocity, which might have resulted in measured ratings that are higher than the models predict.

Note that the right of the figures also includes the measured overall ratings ("∘"-

Table 4.3: Influence factors of the identified channels for the AutoRegressive eXogenous (ARX) Forward Regression (FR) and parametric models, as well as reported values by Cleij et al. [2018]. Dashes indicate channels were not present in the model.

		$I_{\tilde{P}_{f_x}}$ [%]	$I_{\tilde{P}_{f_y}}$ [%]	$I_{\tilde{P}_{f_z}}$ [%]	$I_{\tilde{P}_{\omega_x}}$ [%]	$I_{\tilde{P}_{\omega_y}}$ [%]	$I_{\tilde{P}_{\omega_z}}$ [%]	$I_{\tilde{P}_{\alpha_z}}$ [%]
ARX FR	a-NTC	25.4	74.6	-	-	-	-	-
	a-CWA	28.0	-	-	-	-	72.0	0.0
	a-ORC	13.0	76.2	-	-	-	10.9	-
	a-ALL	24.6	75.4	-	-	-	-	-
Parametric	p-NTC	31.0	69.1	-	-	-	-	-
	p-CWA	35.5	-	-	-	-	64.5	0.0
	p-ORC	24.9	63.0	-	-	-	12.2	-
	p-ALL	34.0	66.0	-	-	-	-	-
	Cleij et al. [2018]	*0*	*37*	*18*	*26*	*2*	*17*	-

Table 4.4: Estimated parameters for the four parametric models.

model	τ [s]	ω_c [rad/s]	K_{f_x} [-]	K_{f_y} [-]	K_{ω_z} [-]	K_{α_z} [-]
p-NTC	0.00	0.33	0.89	1.66	-	-
p-CWA	0.04	0.37	0.78	-	6.71	3.71
p-ORC	0.07	0.52	0.62	1.11	1.08	-
p-ALL	0.00	0.36	0.91	1.50	-	-

symbols), as well as the predicted ("+"-symbols) overall ratings, using the relation $OR_{PH} = 2.0 + 0.8 \cdot \max[\tilde{R}(t)]$. This again shows the generalizability of the p-ALL model, which can predict the overall ratings of all three conditions with reasonable accuracy.

4.4.5.4 Individual Predictions

The developed models deliver a prediction for the "average" participant. However, to form an indication on prediction power of individual ratings, Figure 4.13 shows the VAF values calculated between the "ALL" model and the three data sets together. On the individual level, individual scaling differences in the rating strategy become prominent, which lead to low VAF values. In three cases, the VAF is lower than 0. Therefore, the values are manually set to a value of 0. With an average VAF of 34.5%, these values are lower than the model fits of the average rating data.

4.5 Discussion

The presented experiment applied the continuous rating task of Cleij et al. [2018], who tested short drives each with a single maneuver, in a realistic setting: a long scenario combining a large number of maneuvers characteristic for urban driving. Overall, the 51 participants were well able to distinguish the differences in incongruences between the motion cueing conditions and rate these accordingly. Whether the rating task provides *useable* results is discussed below, in terms of how the continuous and overall ratings correspond, their reliability, and the ability to model and predict the acquired ratings.

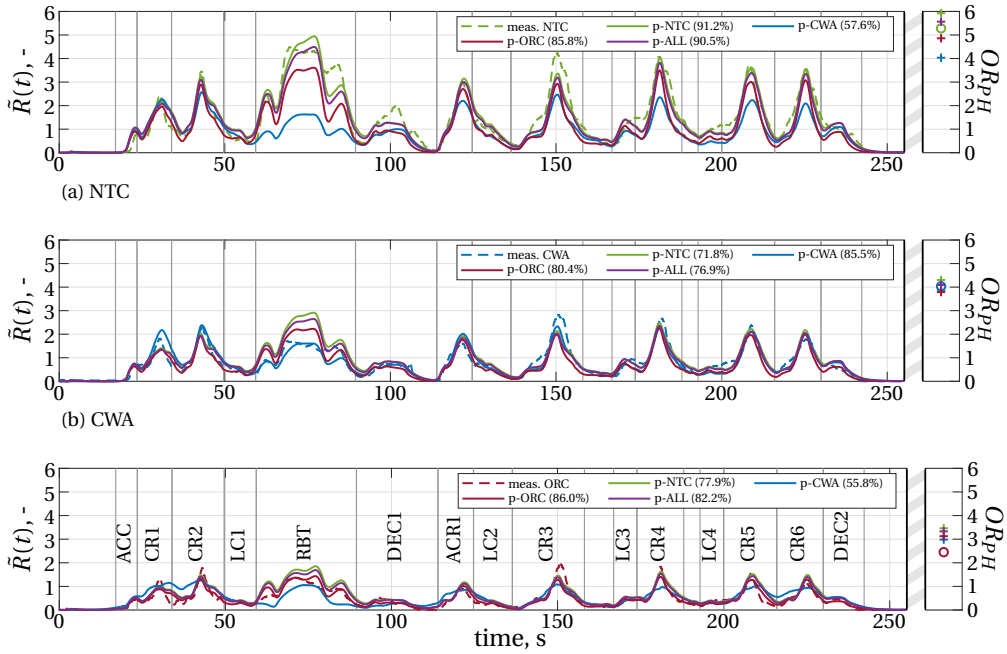


Figure 4.12: The measured continuous (left) and overall (right) ratings of three conditions, each with the four applied models. Percentages in the legend indicate the Variance-Accounted-For (VAF) values for the continuous rating models.

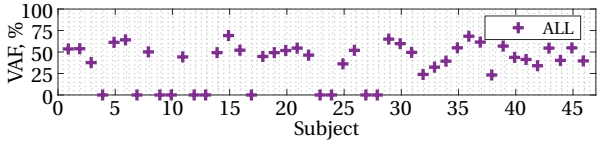


Figure 4.13: Prediction quality of the "ALL" model for each subject.

4.5.1 Continuous and Overall Rating Correlations

Analyzing the correlation between the maximum of the continuous ratings per maneuver and the overall ratings revealed that the most *in*congruent motion dominates a participant's overall impression of the provided simulator physical motion. Recency effects, in which maneuvers occurring later in the scenario have a stronger influence on the overall rating, were not observed. This confirms the findings of Cleij et al. [2018], but also extends this finding for longer-duration and realistic urban driving scenarios containing a large number of maneuvers.

4.5.2 Reliability

Reliability estimates, mainly based on the estimation of McDonald's Omega, show that the urban driving scenario is generally rated in a consistent manner with reliability levels of α $(0.6-0.8)$ similar as reported by Cleij et al. [2018],Cleij et al. [2019],Ellensohn et al. [2019] and Ellensohn et al. [2020]. The most striking result regarding reliability estimates of the continuous MIR data is that they were found to be *inversely related* to the rating power: the lower the ratings, i.e., the better the motion is rated, generally the less reliable the ratings are. A possible explanation is that it is easier for participants to point out that something is wrong, incongruent, rather than that something is right, congruent. This also explains why the worse-rated maneuvers correlate more to the overall ratings.

This leads to a paradoxical situation, as the more one improves the simulator motion cueing, the less reliable the subjective assessment methods to confirm so become. This conclusion is independent of the choice between Cronbach's Alpha and McDonald's Omega. However, in continuous rating studies where reliability estimates are used as a cut-off requirement (such as in Ellensohn et al. [2019]), i.e., by removing data that do not meet a certain value of reliability, Omega *can* be beneficial, as it is shown that for more congruent motion, the difference between Omega and Alpha becomes significant. Thus, it is at these points that Alpha often underestimates the reliability, which can lead to the wrongful conclusion that certain rating data are unreliable. Generally, if incongruences are to be further reduced, reliability can become an issue, such that increasing the number of repetitions or deliberately inducing incongruences in the motion are required to boost reliability.

4.5.3 Model Predictions

In the model predictions, no effect of the reliability is directly observed: The four models (NTC, CWA, ORC and ALL) provide reasonable fits and a decent level of cross-validation when predicting ratings of the other data sets. Overall, the system identification results show that the ratings can be modeled by a low-pass filter response and are dominated by the lateral specific force mismatch. In Cleij et al. [2018] a similar finding was reported, with 37% of the measured ratings attributed to this channel. In our case, this contribution ranges from 33.2% to 70.5%. The lagged response to the mismatches can likely be attributed to the rating dynamics. For example, operating the rating knob to change the rating from a 1 to a 7 requires rotating the knob through all in-between rating values.

Similar as in Cleij et al. [2018], a contribution of the yaw rate mismatch was found, but only in the CWA and ORC conditions. One explanation, strengthened by participants' comments, is that whereas the lateral specific force mismatches were more promi-

nent and easier to identify, yaw motion mismatches were not. That is, the yaw rate mismatches *can* be sensed, but are secondary to the lateral specific force mismatches. NTC and CWA had identical yaw rate mismatches, but NTC also had large lateral specific force mismatches, which therefore became dominant in the rating. For ORC, the yaw rate mismatch was smaller and might have been less noticeable, yielding a smaller contribution.

A notable difference is that an f_x term is identified between 24.9%-35.5%, which is 0% in Cleij et al. [2018]. The obtained values are indeed realistic for an urban scenario in which strong accelerations and decelerations are present. Significant is also that other channels, such as ω_x (which was tuned above the perceptual threshold of 3 deg/s) and ω_y, did not provide a meaningful contribution, such that these were not noticeable or too short to have a meaningful impact on the rating.

When only fitting on one condition, and validating on the other, some generalizability issues are revealed. Due to its different terms and associated weightings, the CWA condition performs less in cross-validating the other two conditions. However, the ALL model (with only contributions of $\widetilde{P}_{f_y}$ and $\widetilde{P}_{f_x}$), which is fit to *all* data together, provides a reasonably good quality of the fit on all conditions and could thus be used as a general model for predicting incongruences, independent of the motion cueing architecture.

4.5.4 Future Work

4.5.4.1 Experiment Differences

The present experiment investigated the applicability of measuring and modeling continuous and overall ratings in a realistic urban scenario. The main motivation for this investigation is to use the gained knowledge to make predictions on the motion cueing of future driving simulation experiments. This can, for example, be used to support decision making when selecting an appropriate simulator and motion cueing settings, and offline tuning of MCA parameters. Other, future experiments for which these evaluations are used might be performed on the exact same urban scenario, on a different urban scenario, or on a completely different scenario type (e.g., highway, rural). Therefore, it is suggested to investigate how the ratings are affected under each of these three steps. The presented model is thus an *explanatory* model, as it provides an explanation on what formed the ratings under the circumstances of the experiment. Its validity needs to be confirmed as a *predictive* model, such that it has the quality to predict the ratings under different circumstances as well.

First, if the scenario would be exactly the same, future work should investigate how ratings are affected by a different participant group and/or a different simulator or motion cueing settings. Cleij [2020] showed that when two experiments expose a different range of motions (for example by using a larger and a smaller simulator), the obtained ratings of these experiments need to be corrected for through a linear scaling factor. The next step, using a different urban scenario could explicitly investigate whether possibly the length or a different order in which maneuvers are presented affects the provided ratings. Finally, extending the results to completely different scenario types would be an important step. For example, a highway scenario might have more interaction with surrounding traffic, which could induce different types of motion (e.g., more lane changes), which might affect the balance between the mismatch channels. Furthermore maneuvers might be harder to rate, as their occurrence might be harder to anticipate than the

visually clear corner maneuvers in an urban scenario. As a result, such scenarios might inherently have a lower reliability.

4.5.4.2 Open-loop Driving Experiments

A main motivation for the presented work is to leverage continuous rating prediction models, which can only be extracted from data collected in open-loop driving experiments, for predicting the quality of closed-loop driving simulator experiments. However, a central assumption so far made in the existing continuous rating literature is that open-loop ratings are also representative for closed-loop driving. However, it is possible that differences between open-loop and closed-loop driving occur due to perceptual differences [Nesti et al., 2016; Valente Pais et al., 2012]. Thus, future work should explicitly investigate whether motion cueing in closed-loop and open-loop driving are in fact rated equivalently. Explicitly proving this would further increase the validity of the continuous rating method for closed-loop testing. An example for which the continuous rating method may be applied is studying the effect of masking of cues [Greig, 1987].

4.5.4.3 Error Types

In the present work, the mismatches in the motion were analyzed through objective difference functions between the vehicle reference and simulator motion. As a result, the rating models linearly depend on the overall magnitude of the mismatch, without making any distinction between what *type* of cueing error is present. However, humans may have different sensitivities to different error types. For example, Grant and Reid [1997] defined three different types of errors for flight simulation motion cueing: false cues, missing/scaling error cues, and phase-error cues. In their definition, false cue motion results in errors in the opposite direction of the true vehicle motion, or a motion cue whereas no motion is expected from the vehicle. A scaled cue is correct in its direction, but mismatched in magnitude compared to the vehicle reference motion, of which the missing cue is a special case (i.e., no simulator motion). Phase errors were also defined by Grant and Reid [1997], in which the simulator's motion is shifted in time (i.e., leading or lagging) with respect to the vehicle reference motion. Variations of these definitions exist, such as defined in Cleij [2020] and Kolff et al. [2022]. Grant and Reid [1997] noted that false cue motion was generally perceived as worse than scaled or missing motion, although without providing experimental proof. Following on preliminary investigations by Cleij [2020], future research should investigate explicitly how these error types compare and if predictive rating models may be improved when different error types are weighted independently in the rating model.

4.6 Conclusion

The difficult trade-off and selection of motion cueing settings would greatly benefit from accurate prediction methods of subjective ratings. This chapter describes the application of continuous and overall motion incongruence ratings in a realistic urban driving experiment through reliability and predictability. From analyzing the correlation between the continuous and overall ratings, it is concluded that incongruent motion strongly determines the overall impression of drivers. This is explained by the reliability of the acquired continuous ratings, which is generally high, but *inversely related* to

the incongruence ratings: the more congruent the presented motion is, the less the acquired ratings can be trusted. Reducing incongruent motion thus requires more effort in the subjective confirmation. This is done either through the deliberate presentation of incongruent motion cues or by increasing the number of repetitions. For the rating data presented in this chapter, the reliability of the data is sufficient, as the estimates are similar to values in literature and no effects on the predictability of the rating data are observed. Non-delayed, first-order linear low-pass filtered responses to the lateral (66.0%) and longitudinal (34.0%) specific force mismatches are sufficient to predict the measured motion incongruence ratings in an urban scenario. Through this model and the gained knowledge on its associated reliability, incongruences can be accurately targeted and reduced in the development, selection, and tuning of future motion cueing.

5

Incongruences in Highway Simulations

Key findings

- As in urban simulations, ratings of highway simulations become less reliable the better the motion cueing is rated, and vice versa. However, highway simulations induce a lower *overall* reliability of the ratings, showing that the scenario type is more difficult to rate.
- The linear rating model of Chapter 4 can predict the measured continuous rating in highway simulations, although an additional response to the lateral specific force mismatches in lane change maneuvers is present.
- As in urban simulations, continuous ratings of highway simulations are dominated by the lateral and longitudinal specific force mismatches. However, a stronger contribution (45%) to the longitudinal specific force mismatch is present compared to urban driving (55%).

This chapter is based on the following publication:

Title:	Subjective Motion Incongruence Ratings in Highway Driving Simulations
Authors:	M. Kolff, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder.
Journal:	To be submitted.

Abstract

*This chapter investigates the reliability and model predictions of subjective motion incongruence ratings in highway driving simulations. A simulator experiment was performed in which participants evaluated the motion incongruences using both continuous and overall rating methods. The experiment had the same set-up as an earlier urban experiment, to compare scenario-specific differences between urban and highway simulations. Three motion cueing algorithms were tested that represented different levels (low-medium-high) of motion cueing quality. Results show that the more congruent the motion is, the less reliable the ratings become, a result reproduced from the urban experiment. However, the reliability of the highway scenario data is overall lower, indicating highway driving is more difficult to rate than its urban counterpart. The linear rating model, previously fitted to the urban experiment's data, is nevertheless successfully used to predict the ratings based on the longitudinal (*45.0%*) and lateral (*55.0%*) specific force mismatches. The rating data can be better described when including a term describing the lateral motion in lane changes. Together, these findings further validate the continuous rating method and the previous proposed rating models, strengthening the ability to make assessments of motion cueing quality without the need for expensive on-site simulator testing.*

5.1 Introduction

HIGHWAY driving is a crucial use-case in driving simulation due to its ability to support the development of important novel vehicle technologies, such as ADAS [Rossi et al., 2020], high-speed driving [Weiss and Gerdes, 2023], and connected vehicle applications [Schrapel and Vinel, 2024]. Driving simulators provide a cheaper and more controllable alternative to real vehicle testing [de Winter et al., 2012]. Especially for highway driving, where the vehicle velocity is relatively high compared to other driving scenarios (e.g., urban or rural), the inherent safety provided by driving simulators is another important benefit. The goal of the motion system of the simulator is to recreate the sensation of inertial motion as closely to the real-life vehicle as possible. However, it is generally not possible to fully reproduce the motion due to the limited size of driving simulators' motion systems and limitations in the MCA, which controls the motion system. This results in mismatches between the vehicle reference and simulator motion. The generated mismatches depend on the simulator, the MCA, and its tuning parameters [Kolff et al., 2023]. In highway simulations, mismatches can negatively affect driving behavior [Wang et al., 2023], reduce realism, and induce simulator sickness [Himmels et al., 2022b]. Those motion mismatches that are perceived by the driver are known as *incongruences* [Cleij et al., 2018]. Experimental measurements of the (in)congruence of motion are most commonly based on subjective evaluations from drivers [Biemelt et al., 2021; Cleij et al., 2018; Kolff et al., 2024b; Rengifo et al., 2021]. Providing accurate motion for a highway simulation thus in practice requires the subjective verification of the motion by test drivers through dedicated, and costly, simulator tests.

Being able to identify and *predict* how motion cueing settings affect subjective evaluations would therefore be a crucial tool. This would allow for improving the inertial motion reproduction of the simulator and more rapid testing and selection of various motion cueing settings, *without* the need for expensive on-site simulator testing. Cleij et al. [2018] introduced a *continuous* subjective rating method suitable for model development. While being driven around ("open-loop" driving), drivers then continuously give a rating that reflects their perceived motion incongruence at each point in time. The method has since been successfully used in [van der Ploeg et al., 2020] (same scenario as [Cleij et al., 2018]), [Ellensohn et al., 2019a; Ellensohn et al., 2020; Ellensohn et al., 2019c] (rural scenarios) and [Cleij et al., 2019] (rural-urban scenario). In Chapter 4, this approach was furthermore shown to be valid in realistic urban simulations: the obtained average continuous ratings could be predicted using a first-order low-pass filter response to a weighted sum of the lateral and longitudinal specific force mismatches. The reliability of the (continuous) subjective ratings, however, was shown to decrease with higher motion cueing quality. While highly useful for modeling, overall ratings (a single rating value representing the whole drive) are more practical for a direct trade-off of motion cueing settings or parameter sensitivity analyses. Both Cleij et al. [2018] and Chapter 4 also showed that overall ratings strongly correlate with the most incongruent point in the continuous ratings.

In the current state-of-the-art, no similar work on measuring and predicting subjective ratings in highway simulations exists. The motion of highway simulations is typically easier to reproduce on driving simulators than the more dynamic urban scenarios [Ellensohn, 2020]. The resulting smaller motion mismatches may affect the reliability, rating

relationships, and predictability of the subjective ratings. The decreased reliability of the ratings in congruent motion, as described in Chapter 4, makes it then likely that *less* reliable rating data will be collected on highway simulations, complicating the assessment of motion cueing settings. Furthermore, as the relative weighting of the mismatch channels may be different in highway driving, a different parametrization of the rating model may also be necessary.

This chapter therefore describes an open-loop driving experiment in which 33 participants provided continuous and overall ratings for a realistic highway simulation. The findings on rating reliability, the correlation between continuous and overall ratings, and predictability are directly compared to those acquired in the urban simulations of Chapter 4, where a highly similar experiment set-up was used (simulator, experimenter, briefing, and training procedures). Three different MCAs were tested in a within-subjects design: a Classical Washout Algorithm (CWA), the same algorithm with increased tilt-coordination gains, and an optimization-based algorithm ("Oracle"), having full knowledge of the future states, only possible in open-loop driving. First, the rating model of Chapter 4 is applied to the current highway driving data to test its application without adaptations, which would be the ideal case. Second, the model is refitted to the highway data, allowing different mismatch channels and parameters than the model of Chapter 4. Finally, an extended model is fit that also includes a response to the simulator motion *itself*, rather than the corresponding mismatch, which will be shown to positively impact the match with the continuous rating data.

The chapter is structured as follows. Section 5.2 introduces the model structures. Section 5.3 shows the experiment set-up. The results are shown in Section 5.4, followed by the discussion in Section 5.5. The chapter is concluded in Section 5.6.

5.2 Methods

5.2.1 Driving Task

The vehicle simulation produces the reference motion signal $\widetilde{S}_{veh}(t)$ that is sent to the Motion Control System (MCS), as shown in Figure 5.1. During the experiment, the reference vehicle simulation was a pre-recorded drive, meaning that participants passively (i.e., as passengers) experienced the vehicle's motion ("open-loop" driving). The MCS block consists of two parts, the Motion Cueing Algorithm (MCA) and the Motion System (MS). The latter is the physical component of the simulator, together with its control software, that produces the actual inertial platform motion. The MS receives its setpoints from the MCA, which, using knowledge of the configuration and dimensions of the motion system, converts the vehicle motion to simulator motion that can be reproduced by the motion system. Therefore, using the input signal $\widetilde{S}_{veh}(t)$, the MCS gives the output $\widetilde{S}_{sim}(t)$. Differences between these signals (vehicle reference and simulator motion) are then the objective mismatches of the simulator motion, i.e., $\Delta\widetilde{S}(t) = \widetilde{S}_{veh}(t) - \widetilde{S}_{sim}(t)$. The mismatches highly depend on the simulator configuration used [Kolff et al., 2023].

5.2.2 Rating Task

During the open-loop driving scenario, the participants performed a rating task, comparing the perceived simulator motion to the motion expected of the real vehicle. This is

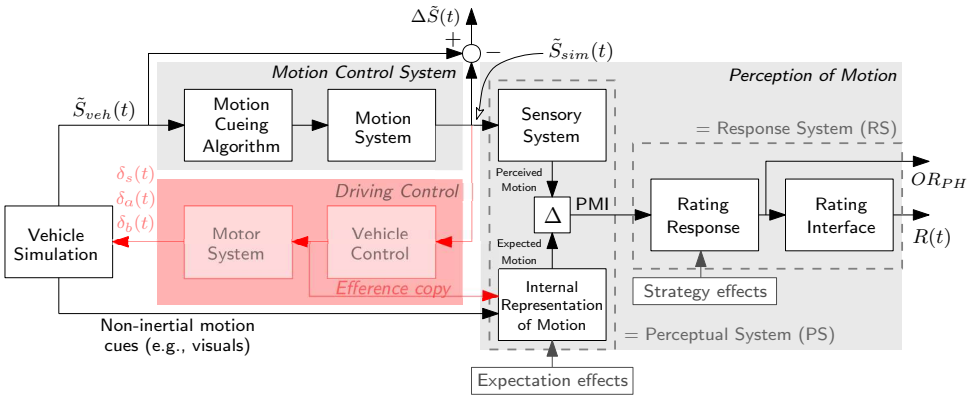


Figure 5.1: Block diagram of the driving and rating tasks from Chapter 4, highlighting the difference between the subjective PMI and objective mismatches $\Delta\tilde{S}(t)$. The part "Driving Control" (red, dashed lines) is only present in closed-loop driving, and therefore not active in the present experiment.

described in Figure 5.1 by the block *Perception of Motion*. The participant cannot exactly know what the real vehicle's motion would feel like in a particular situation (i.e., the real vehicle motion signal S_{veh} is unavailable to them). They must therefore use an *internal representation* [Stassen et al., 1990] of the vehicle motion, for example based on previous experience and the non-inertial motion cues (e.g., visuals) provided in the simulation. Individual variations can arise due to different familiarity or experience with the simulated vehicle (Chapter 4), as indicated by the "expectation effects" in Figure 5.1. Both the expected and the vehicle model motions can thus differ from the real vehicle motion.

In closed-loop driving (i.e., when driving the car oneself), the red path in Figure 5.1 would be active, denoted by *Driving Control*. The expected motion is then not only based on what drivers expect to feel from the non-inertial motion cues, but also what they expect to feel as a result of their own intended control actions of the steering wheel, accelerator pedal, and braking pedal deflections ($\delta_s(t)$, $\delta_a(t)$, and $\delta_b(t)$, respectively). In open-loop driving (as in the present experiment), the lack of driving control implies that no efference copy is available. This means that the internal representation of motion, and in turn the expected motion, is only based on the non-inertial motion cues.

If a participant notices a difference between the perceived and expected motion, an *incongruence* occurs. The Perceived Motion Incongruence (PMI) defines a participant's perception of what is (in)congruent over time [Cleij et al., 2018]. Hence, the PMI would be a highly useful quantity to measure; it is, however, internal to the human and thus not directly measurable. Instead, a Motion Incongruence Rating (MIR) can be asked from the drivers Cleij et al., 2018. This is a subjective rating that functions as a proxy for the PMI. The PMI and MIR can differ, as the latter can be affected by the *rating response* (Figure 5.1). This shows that the rating can be affected by several "strategy effects" (Chapter 4): 1) the participants' motivation to actively perform the rating task, 2) their reference of what counts as congruent or incongruent motion, 3) the anticipation of motion of upcoming maneuvers, and 4) their understanding of what is to be rated.

Especially the anticipation of upcoming maneuvers can be different between urban

and highway open-loop driving scenarios. In urban driving, the motion of the vehicle and simulator can, to a large extent, be predicted through the fixed geometry of the surroundings. In contrast, highway driving is less deterministic, as certain maneuvers (e.g., accelerations and lane changes) are difficult to predict without knowing the intention of the driver represented in the recording.

In the present experiment, rating data were acquired continuously (during the drive), as well as post-hoc (after each drive), to remain close to the work described in Chapter 4. For both rating methods, the MIR values were on a semantic differential scale, ranging between values of 0 (fully congruent) and 10 (highly incongruent), with steps of 1, to remain close to the work of Cleij et al. [2018] and Chapter 4. In the case of continuous ratings (denoted by $R(t)$ in Figure 5.1), the MIR is given through a rotary knob that can be adjusted at any time, yielding continuous recordings over time. These ratings therefore aim to represent the subjective opinion of the subject at any point in time.

For overall ratings (denoted OR_{PH}), only a single rating value is acquired after each drive. This rating aims to represent the quality of the whole drive. In principle, participants can decide for themselves what determines their overall rating. However, previous research [Cleij et al., 2018; Kolff et al., 2024b] has shown that this rating strongly correlates with the highest continuous rating during a drive, i.e., the most *in*congruent point during a drive.

5.2.3 Rating Model

The "p-ALL" rating model of Chapter 4, see Figure 5.2, is used 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 $\widetilde{S}_{veh,m}(t)$ and the simulator motion $\widetilde{S}_{sim,m}(t)$, i. e., $\Delta\widetilde{S}_m(t)$, with $\widetilde{P}_m(t) = K_{\widetilde{P}_m}|\Delta\widetilde{S}_m(t)|$. Here, m represents the mismatch direction. The model was developed on the combined (hence: "ALL") rating data of two CWA variants and the Oracle MCA, as in the present experiment.

The signals $\widetilde{P}_m(t)$ are formed by a model of the perceptual system ($\widetilde{PS}$), with the mismatches $\Delta\widetilde{S}_m(t)$ between the vehicle motion $\widetilde{S}_{veh,m}(t)$ and the simulator motion $\widetilde{S}_{sim,m}(t)$ as inputs. The absolute value block in Figure 5.2 indicates that both positive and negative mismatches result in an increase of the rating value. $K_{\widetilde{P}_m}$ represents the gains on the mismatches m, e.g., $m \in [f_x, f_y, \ldots]$.

The parametric "p-ALL" model of Chapter 4 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 $\widetilde{R}(t)$ (output):

$$\widehat{\widetilde{R}}(j\omega) = \sum_m K_{\widetilde{P}_m}\left(\frac{\omega_c}{j\omega+\omega_c}\right)\Delta\widehat{\widetilde{S}}_m(j\omega), \tag{5.1}$$

with the cut-off frequency ω_c and $K_{\widetilde{P}_m}$ being the weights on the different mismatch channels. Chapter 4 showed that the continuous ratings as measured in that study could be largely explained when considering the mismatch channels $\widetilde{P}_{f_y}$ and $\widetilde{P}_{f_x}$, with respective gains of 1.50 and 0.91, together with $\omega_c = 0.36$ rad/s.

5.2.4 Model Fitting Procedure

Several models of increasing complexity are tested for the current highway dataset:

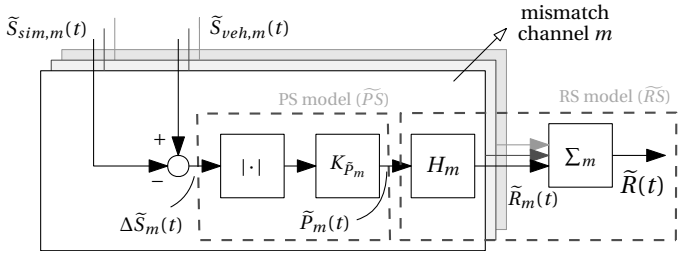


Figure 5.2: The linear model "p-ALL urban" from Chapter 4.

1. In the ideal case, the existing parametric model "p-ALL urban", extracted from an urban driving simulation experiment in Chapter 4, would apply one-to-one. In this case, the exact same model, with the same mismatch channels and parameter values, is applied.

2. Second, the same model structure as "p-ALL" is applied, but the fitting process is revisited to allow for different parameter values that may better match the current highway data. This can result in a better fit, if, for example, the participants weighted the mismatch channels differently than in the urban data. This model is referred to as "p-ALL highway".

3. Third, an extended model is considered, in which the fitted model includes a response to the simulator lateral motion magnitude during the lane change maneuvers, see Section 5.4, denoted as "p-ALL highway + LC".

For the second and third cases, the models are fitted in the time-domain using a Nelder-Mead simplex algorithm [van der El et al., 2018] through the minimization of a sum-of-squared-errors cost function:

$$\arg\min_{\Theta} J = \sum_{t} [R(t) - \widetilde{R}(t|\Theta)]^2, \tag{5.2}$$

where $\Theta = [\omega_c,\ K_{f_x},\ K_{f_y}]$ is the parameter set. This gradient-based method does not guarantee attaining the global optimum. Therefore, 30 iterations are performed with uniformly distributed random numbers between 0 and 3 as initial conditions. The parameter set leading to the overall lowest cost is then selected as the final solution. The fit was always performed on all the rating data of the three conditions grouped together, similar to the "ALL" strategy in Chapter 4.

5.2.5 Variance Accounted For

The Variance-Accounted-For (VAF) is used to quantify the quality of the fit, with $e(t|\Theta) = R(t) - \widetilde{R}(t|\Theta)$:

$$\mathrm{VAF}(\Theta) = [1 - \sigma^2_{e(t|\Theta)} / \sigma^2_{R(t)}] \cdot 100\%, \tag{5.3}$$

where $R(t)$ is the measured continuous rating signal (i.e., the reference) and $\bar{R}(t|\Theta)$ is the modeled signal. The VAF describes how much of the measured signal's variance can be

explained by the modeled signal: a VAF of 100% describes a perfect fit, whereas there is no lower bound that represents the "poorest of fits".

To calculate the relative contribution of the mismatch channels to the ratings, the *influence factor* [Cleij et al., 2018] is calculated as:

$$I_m = \sum_t \widetilde{P}_m(t) / \sum_t \widetilde{P}(t) \cdot 100\%, \tag{5.4}$$

with $\widetilde{P}_m(t) = K_{\widetilde{P}_m} |\Delta \widetilde{S}_m(t)|$, see Figure 5.2, and m the mismatch channel. $\widetilde{P}(t)$ is the total modeled PMI. The influence factor therefore represents the relative contribution of the different mismatch channels. The sum of all channels' influence in the model is therefore always 100%. This metric allows for a comparison on the relative importance of the different mismatch channels with those in Chapter 4.

5.2.6 Reliability

In Chapter 4, reliability of the continuous ratings was explicitly investigated. Although the continuous ratings are continuously recorded during a drive, the recorded ratings can be unreliable, for example if the participants do not consistently perform the rating task. The reliability then quantifies the extent to which the obtained continuous ratings can be trusted.

Recording continuous ratings over various conditions yields a collection of rating time signals $R_{cjp}(t)$, with c the condition, j the condition repetition and p the participant. If along one of these elements the average is taken, this element is taken out of the subscript. $R_c(t)$, for example, represents the rating of the average participant across all repetitions in a given condition. In the experiment used in the current chapter, each run lasted 430.83 s, with continuous rating data being recorded at 100 Hz ($\Delta t = 10$ ms); each recording R_{cjp} contains $N = 43,083$ samples. The total score is the sum of the run items $X_{cjp} = \sum_t R_{cjp}(t)$, where $\sigma^2_{X_{cp}}$ is the variance of total scores over multiple repetitions.

Theoretically, if an infinite number of identical and independent repetitions would be performed by a participant, the average of all total scores would result in the *true score*, i.e., the expected value of the rating: $T_{cp} = E[X_{cjp}]$. Each separate test result contains a random, stochastic measurement error $E_{cjp} = X_{cjp} - T_{cp}$. Reliability is defined by how much of the test score variance can be explained by the true score variance [Sijtsma, 2009]. As the true score cannot be determined, only an *estimation* of a lower bound of reliability can be made. Here, the reliability is calculated using Cronbach's Alpha:

$$\alpha_{cp} = \frac{J}{J-1} \frac{\sum_j \sigma^2_{cjp}}{\sigma^2_{X_{cp}}} \tag{5.5}$$

where J is the total number of repetitions of each test condition and σ^2_{cjp} is the variance of the individual samples. The coefficient α is unbounded on the lower side, i.e., $[-\infty < \alpha \leq 1]$, whereas a value of 1 indicates perfect reliability.

Chapter 4 showed that, for the urban rating data, more *congruent* motion leads to *lower* reliability. The following relationship was found to predict the average reliability of the participants:

$$\widetilde{\alpha} = a - \frac{1}{b\bar{R} + c}, \tag{5.6}$$

with the coefficients $a = 0.89$, $b = 19.65$, and $c = 1.39$, which were found to well describe the obtained reliabilities and can be used to predict reliability based on the continuous rating signal. $\bar{R}$ is the average rating across time.

In Chapter 4, the reliability estimates, and thus the resulting fits, were based on three repetitions of each condition, allowing the calculation of the more accurate reliability metric *McDonald's Omega* rather than Cronbach's Alpha. This, however, requires at least three repetitions of the same condition. In the present highway experiment, there were only two repetitions, due to constraints on the experiment length. Therefore, the calculation of McDonald's Omega was not possible.

Furthermore, a comparison of the reliability of the present highway data (with two repetitions) to the urban data fit (based on three repetitions) would be unfair, as increasing the number of repetitions also increases the reliability of that data. Therefore, the calculation of Cronbach's Alpha described in Chapter 4 was repeated three times, once for each repetition pair (runs 1 & 2, runs 2 & 3, and runs 1 & 3), of which the average value for α was taken. This allows for a direct comparison of the reliability between the highway and urban datasets. This yields the coefficients $a = 0.82$, $b = 15.89$, and $c = 1.42$ for (5.6).

5.3 Experiment Set-up

5.3.1 Scenario

All participants experienced the same recording of a drive on a fictional German highway. It consisted of various typical highway maneuvers, as shown in Figure 5.3. The drive accelerated ("ACC") from a parking place, including several lane changes, after which the highway was entered ("ENT"). Traffic was present and always exactly the same (i.e., part of the recording). Driving by a static traffic accident of two other road users (see Figure 5.3) and corresponding slow traffic were simulated to induce additional decelerations and accelerations ("D/A"). Furthermore, a sustained curve ("CUR") (see Figure 5.3) to change highway and several lane changes ("LC") to overtake traffic were simulated, either to the left ("[L]") or right ("[R]"). After a section of constant braking ("CB") and two double lane changes ("DLC"), the vehicle entered the same urban environment as was simulated in the urban experiment of Chapter 4. Drivers experienced the same two last corners ("CR") as in that work to allow for a comparison of identical maneuvers between the experiments.

5.3.2 Apparatus

For maximal comparability, the experiment had the exact same experiment-setup as in Chapter 4: BMW Group's Ruby Space simulator (Figure 5.4a) was used, with nine DoFs. It consists of a hexapod on top of a tripod system, where the latter adds additional workspace in the longitudinal (f_x), lateral (f_y) and yaw (ω_z) degrees-of-freedom. BMW's iDrive navigation knob on the center console was used by participants to give their continuous rating (see Figure 5.4b), recorded at 100 Hz. The 240° projection screen showed the visuals, and the current rating value in the form of a "rating bar" [Cleij et al., 2018]. The size and color of the rating bar changed from 0 (small, white) to 10 (long,

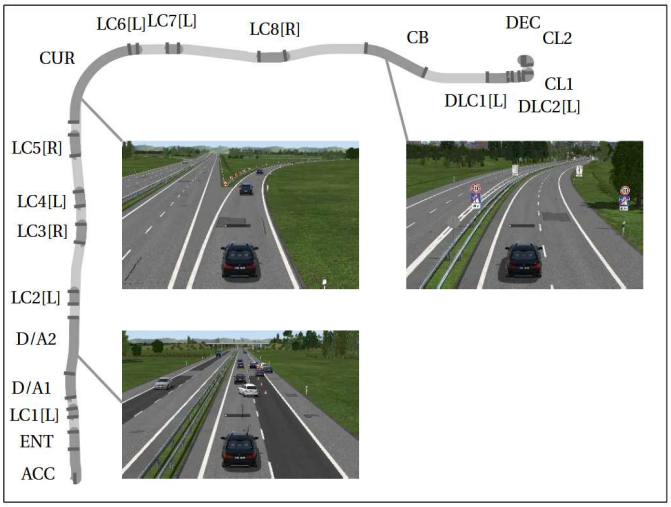


Figure 5.3: Overview of the driven route with the maneuvers indicated. Three representative screenshots are shown.

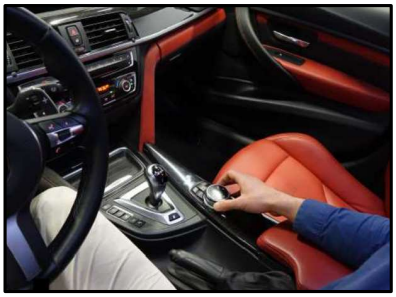
red), which makes the rating system more intuitive for participants to use. The driving direction (left/right arrows) was also shown, together with the vehicle velocity.

5

5.3.3 Independent Variables

Three MCA were tested. First, a classical washout algorithm ("CWA") [Conrad et al., 1973; Reid and Nahon, 1985] was used, where the vehicle motion is distributed over the hexapod (high-frequency), tripod (medium-frequency) channels and washed-out with second-order high-pass filters. The low-frequency forward and lateral accelerations were reproduced by tilt-coordination, in which rotations over roll and pitch are used to create the sensation of a sustained acceleration. This *somatogravic* illusion is only com-

(a) BMW Group's Ruby Space simulator.



(b) A test driver using the iDrive rating knob.

Figure 5.4: The experiment set-up (adapted from Ellensohn et al. [2019]).

plete as long as the accompanied tilting rates are not noticeable [Stratulat et al., 2011]. The aim of this condition is to provide a reference of a state-of-the-art algorithm that can be used in real-time. Due to the worst-case tuning and the "blindness" of the CWA to future states, it cannot utilize the simulator's full potential at all moments during the scenario. The CWA tuning set was exactly the same as that of Chapter 4, to maximize comparability.

Second, a variant of the same CWA was used, which used increased tilt-coordination ("HTC"). Here, the gains of the tilt-coordination channels were increased (31% in longitudinal direction, 80% in lateral direction), such that the roll and pitch rates were above the perceptual threshold (> 3 deg/s [Reymond and Kemeny, 2000], see Figure 5.5d), potentially increasing motion incongruence due to spurious rotations. The associated benefit is that a smaller mismatch is present in the longitudinal (Figure 5.5a) and lateral (Figure 5.5b) specific forces, potentially *decreasing* ratings. The purpose of this condition is to investigate the trade-off between specific force and rotational rates, often an important focus point in an MCA tuning process.

Third, an optimization-based algorithm was tested [Ellensohn et al., 2019c] as in Chapter 4, where the simulator motion along the complete recorded drive was optimized offline: the Oracle ("ORC") MCA. This algorithm can only be used in open-loop simulations, but allows for the investigation of how the available simulator workspace may be fully exploited. It serves as the upper bound of motion cueing quality possible on the simulator. The rotational rates ω_x and ω_y are constrained to remain below the perceptual threshold (< 3 deg/s). Figure 5.5 illustrates that even ORC is not able to perfectly reproduce the vehicle motion of the six channels, although it comes the closest in reproducing the longitudinal and lateral specific forces, as well as the yaw rate. It is therefore hypothesized that this condition will receive the best (i.e., lowest) ratings.

5.3.4 Participants and Procedures

Thirty-three participants (25 males, 8 females) performed the experiment. All of them were employees of BMW Group with a European driver's license B (M = 25.3 yrs SD = 10.10 yrs) and an average yearly driven distance of M = 22,230 km (SD = 13,599 km). The average age was M = 41.1 yrs (SD = 9.9 yrs). Ten participants had previous experience in driving simulators. All participants provided informed consent and the experiment was approved following BMW's internal ethics review procedure.

The experiment started with a training run to familiarize participants with the scenario and rating method. After that, the experiment phase started, in which participants started with either CWA, HTC or ORC. Each condition was repeated two times, yielding six runs. After every second run, a five minute break was taken. 30 complete data sets were obtained, as three participants were unable to finish the experiment due to technical reasons or simulator sickness. For the latter, the Motion Illness Symptoms Classification (MISC) [Bos et al., 2005; Reuten et al., 2021] questionnaire was asked after every drive. The experiment was stopped if participants reached a value of 6 two drives in a row, or if a single 7 or higher was reached, as in Hogerbrug et al. [2020].

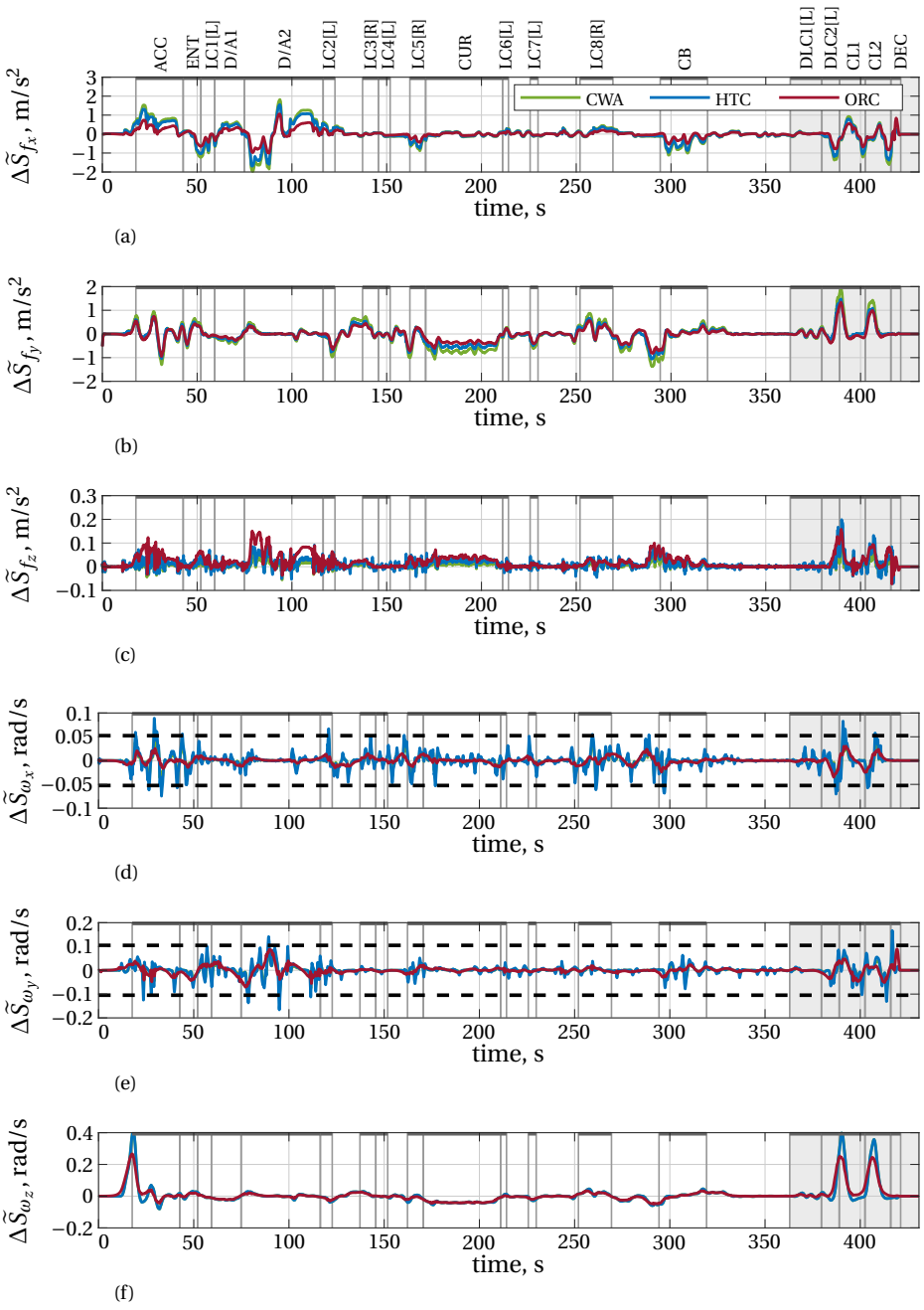


Figure 5.5: Mismatches induced by the three algorithms. Grey vertical lines indicate the separate maneuvers. The grey areas indicate the urban maneuvers, which were excluded from the model fitting step. The dashed lines indicate the rotational threshold of 3,deg/s [Reymond and Kemeny, 2000], relevant for the use of tilt-coordination.

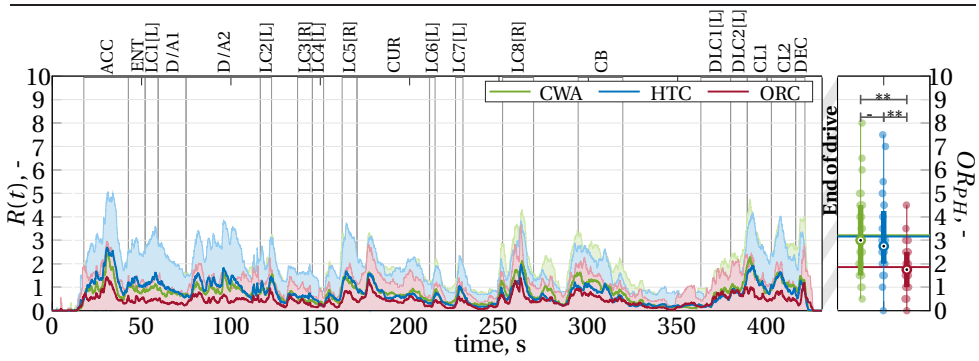


Figure 5.6: Left: measured continuous ratings, with the lines the mean and the shaded areas the standard deviation. The maneuvers from Figure 5.3 are also indicated. Right: overall ratings, displayed as box plots, showing the medians and the inter-quartile ranges. The horizontal lines show the means.

5.4 Results

5.4.1 Rating Measurements

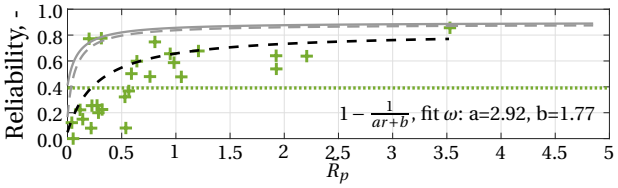
Figure 5.6 shows the continuous ratings (left), with the maneuvers from Figure 5.3 annotated with vertical lines and their acronyms. The individual rating data are included in Appendix D.2. Generally, the ratings are low, especially compared to the rating values found in Chapter 4, where the highest peaks in the average ratings were around 4.0. In the present data, the highest peak value is 2.8 and occurs for the HTC and CWA conditions, for the initial acceleration maneuver (ACC). Overall, the ORC motion receives the lowest (i.e., the best) ratings on average for all maneuvers. CWA and HTC are highly similar. Furthermore, the right subplot in Figure 5.6 shows the distributions of the collected overall rating data. The overall ratings for CWA and HTC are highly similar, and that ORC was better rated. HTC is rated at $\mu = 3.32$ ($\sigma = 1.94$), while the CWA condition is rated only marginally better at $\mu = 3.28$ ($\sigma = 2.09$). The overall ratings for ORC are substantially lower on average: $\mu = 2.44$ ($\sigma = 0.96$).

5.4.2 Reliability

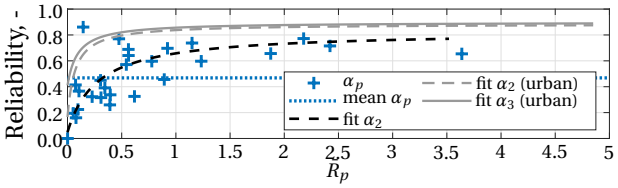
The Cronbach's Alpha reliability estimates for the continuous rating data are shown in Figure 5.7a-5.7c. The observation from Chapter 4, i.e., that reliability goes down with decreasing rating values (i.e., better motion quality), is confirmed here: the better the motion cueing is rated, the less we can trust the acquired ratings. The average reliabilities (dot-dashed lines) are 0.47 for HTC, 0.39 for CWA, and 0.25 for ORC. These are substantially lower than found for an urban scenario in Chapter 4, where the Alpha averages of its CWA, NTC, and ORC conditions were 0.74, 0.62, and 0.55, respectively, based on three repetitions. When the average of each two-pair Cronbach's Alpha, as described in Section 5.2, is taken, the averages are 0.65, 0.53, 0.47, for the urban's CWA, NTC, and ORC conditions, respectively.

In Figure 5.7, the black line is the fit of the reliability model of (5.6), thus best describing the reliability estimates of the recorded data. The grey dashed line is the fit of Chapter 4 estimated on three repetitions. For the fairer comparison, the solid grey line

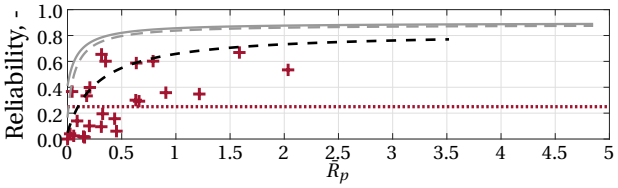
shows the fit performed on *two* repetitions ("fit α_2 urban"), which allows for comparing to the present highway data. The latter is indicated by the black line ("fit α_2"). The α_2 urban fit again shows considerably higher reliability compared to the α_2 highway fit of the present data.



(a) CWA



(b) HTC

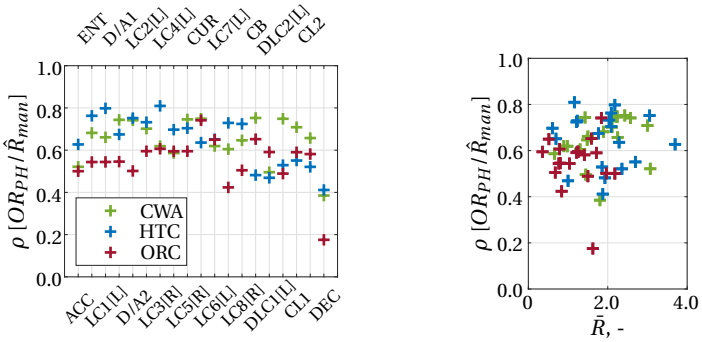


(c) ORC

Figure 5.7: Reliability coefficients α_p of all subjects per condition, as function of the average of their continuous ratings $\bar{R}_p$, showing that reliability decreases with lower ratings.

5.4.3 Rating Relationships

Previous research in Chapter 4 showed a strong correlation between the most incongruent point in a simulated scenario (highest continuous rating value) and the overall rating, yielding the regression relation: $OR_{PH} = 2.0 + 0.8 \cdot \max[R(t)]$. This analysis is repeated here. Figure 5.8a shows that the highest correlation coefficient occurs for the maneuver "LC3[R]", at 0.81. This gives the regression relation $OR_{PH} = 1.2 + 0.6 \cdot \max[R(t)]$. Furthermore, Figure 5.8b shows the correlation coefficients as a function of the average continuous rating. Here, no clear trends are visible, in contrast to the urban data. This shows that the most incongruent point in the continuous rating did not correlate most with the overall ratings for the current highway driving data. In fact, the "LC3[R]" maneuver received relatively *low* continuous ratings (see Figure 5.6).



(a) Per maneuver

(b) As a function of the rating

Figure 5.8: Pearson correlation coefficients between the overall ratings (OR_{PH}) and the maximum of the continuous ratings within each maneuver ($\hat{R}_{man}$).

5.4.4 Rating Models

The continuous ratings predicted by the model "p-ALL urban", as well as the refitted model ("p-ALL highway", allowing for different parameter values), are shown in Figure 5.9, together with the measured ratings. As visible in the figure, and confirmed by the VAF values of the fits in Table 5.1, both models capture the general trends of the measured ratings reasonably well. For the model "p-ALL highway", the VAF is slightly higher than of the model "p-ALL urban", which indicates a better model fit. However, the VAF values are clearly lower compared to the 80 – 90% described in Chapter 4.

Some maneuvers are not predicted well. First, the variations in maneuver D/A2, which was the simulation of the traffic accident and subsequent traffic jam, are not modeled accurately. Here, it is likely that the close interaction with the traffic at low speeds affected the ratings, which is not captured in the rating model, which only predicts changes in ratings due to changes in physical motion mismatches.

Second, specifically the variations that occur in the lane change maneuvers ("LC#", as well as "ACC"), are not described well. From informal comments of the participants, it was also noted that specifically the lane change maneuvers were difficult for them to anticipate, considering that a pre-recorded drive was used. These maneuvers may thus have lead to a sense of surprise from the sudden lane change movement. To represent this response, the mismatches in the lane changes were added to the model fitting as a separate response.

This signal is shown in Figure 5.10 and is only defined in the lane change maneuvers. Therefore, additional high-pass filter dynamics $H_{y,lc}(j\omega)$ to the lane changes were included:

$$H_{y,lc}(j\omega) = \frac{j\omega}{j\omega + \omega_{c,lc}}, \tag{5.7}$$

with the cut-off frequency $\omega_{c,lc}$. The adapted model structure is shown in Figure 5.11. This furthermore includes the gain $K_{y,lc}$, which is zero outside of the lane change maneuvers. The model fits are thus repeated with these additional dynamics, leading to the model "p-ALL highway + LC'. As shown in Figure 5.9 and Table 5.1, this indeed leads to a better fit for all three conditions, as confirmed by the increase in VAF.

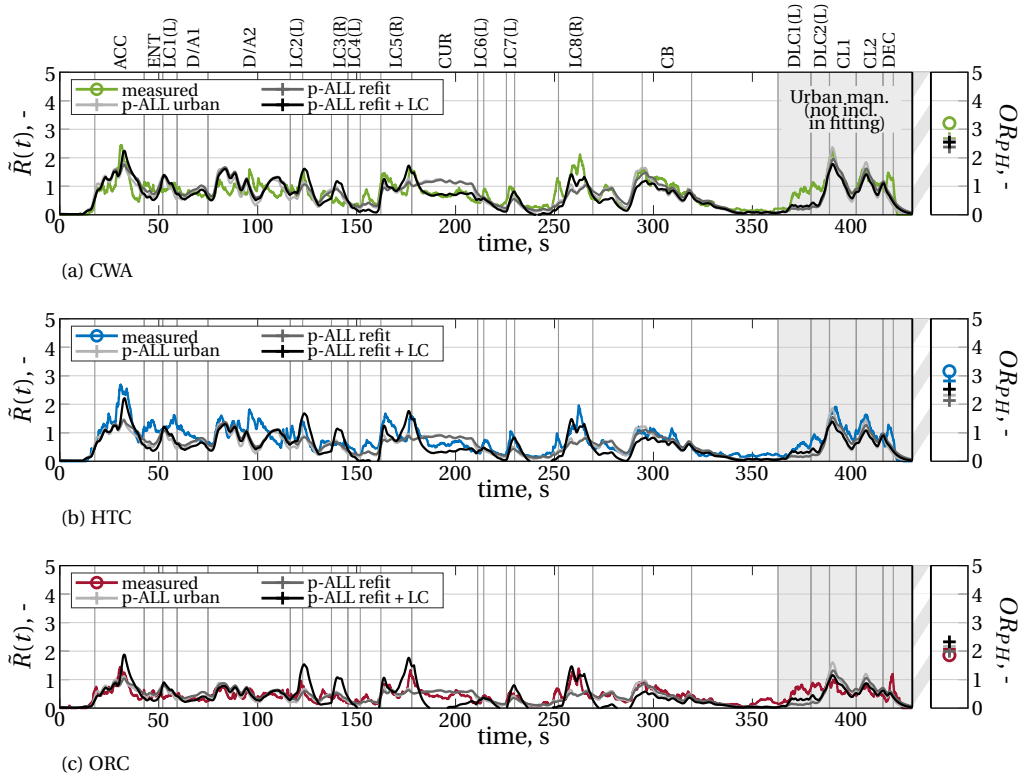


Figure 5.9: The measured average continuous (left) and overall (right) ratings of three conditions, each with the three applied models. The grey areas represent the urban maneuvers, which were not included in the fitting.

Table 5.1: VAF values for the three considered continuous rating models.

	p-ALL urban	p-ALL highway	p-ALL highway + LC
CWA	48.9%	49.2%	55.0%
HTC	58.3%	61.7%	65.1%
ORC	31.9%	30.0%	33.4%

Table 5.2: Estimated parameters for the parametric models.

model	ω_c [rad/s]	K_{f_x} [-]	K_{f_y} [-]	$K_{y,lc}$ [-]	$\omega_{c,lc}$ [rad/s]
p-ALL urban	0.36	0.91	1.50	-	-
p-ALL highway	0.21	1.09	1.46	-	-
p-ALL highway + LC	0.24	1.10	1.29	2.1	0.052

The parameter values of the three models are shown in Table 5.2. The values between the models "p-ALL urban" and "p-ALL highway" are highly similar, although the gain K_{f_x} to the longitudinal mismatches is higher. This shows that participants rated the longitudinal specific force mismatches more strongly than in the urban driving experiment. The lower cut-off frequency of the low-pass filter indicates the participants' stronger filtering to high-frequency motion. Finally, Table 5.2 also shows the parameters of the longitudinal, lateral, and lane change specific force mismatch channels, as well as the cut-off frequency of the low-pass filter, remain largely unaffected when considering the "p-ALL highway + LC" model, although the gain K_{f_y} slightly decreases, from 1.46 in the "p-ALL highway" model to 1.29 in the "p-ALL highway + LC" model.

The models' influence factors (always summing up to 100%) are shown in Table 5.3. The bottom row shows the influence factors from the urban data sets from Chapter 4 for reference. In all three highway models the influence of the longitudinal specific force is larger than observed for the urban data. This can be explained to the stronger longitudinal motion in highway driving compared to urban driving [Bosetti et al., 2014], which was also the case between the present highway and urban experiments, when consider-

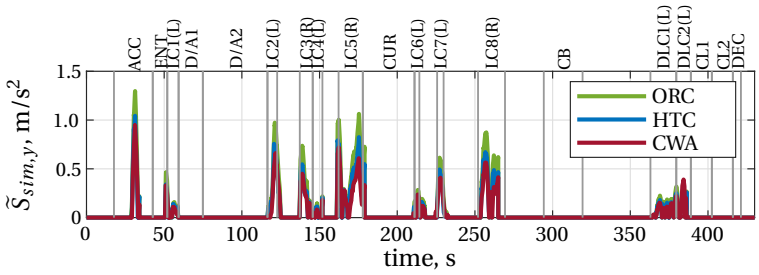


Figure 5.10: Lateral specific force contributions during the lane changes in the three conditions.

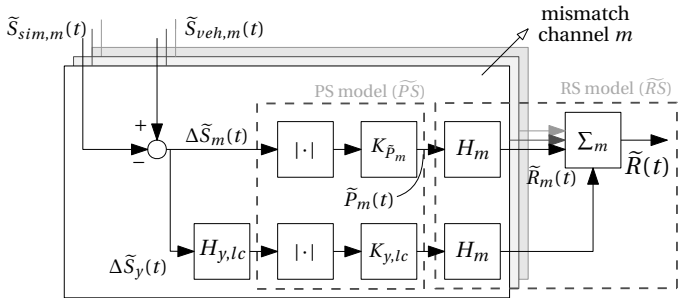


Figure 5.11: The linear model from Chapter 4, describing the additional response to the lateral simulator motion in lane change maneuvers ("p-ALL highway + LC").

Table 5.3: Influence factors of the identified channels in the parametric models, as well as the values for the urban data set of Chapter 4. Dashes indicate channels were not present in the model.

data	model	$I_{\widetilde{P}_{fx}}$	$I_{\widetilde{P}_{fy}}$	$I_{\widetilde{P}_{y,lc}}$
Highway	p-ALL urban	40.0%	60.0%	-
	p-ALL highway	45.0%	55.0%	-
	p-ALL highway + LC	44.0%	47.1%	8.9%
Urban	p-ALL	34.0%	66.0%	-

ing the ratio of the average vehicle motion strength in both scenarios:

$$r_f = \frac{\text{mean}[|\widetilde{S}_{veh,f_x}(t)|]}{\text{mean}[|\widetilde{S}_{veh,f_y}(t)|]}, \tag{5.8}$$

which yields $r_f = 1.01$ for the highway experiment and $r_f = 0.83$ for the urban experiment. Furthermore, although having a clear impact on the model fits, the lane change component only provides a contribution of 8.9% to the rating model, as it is only active for short periods of time.

The considered recorded driving scenario ended in the exact same urban environment tested in the experiment of Chapter 4, which allows for a direct comparison of the rating model's prediction quality *between* both experiments. The grey areas in Figure 5.9 indicate these urban maneuvers, which were not included in the model fitting described above, and are thus only analyzed descriptively. All models are found to describe the ratings in the urban part of the scenario surprisingly well, further showing the *transferability* of the rating model *between* experiments and thus predictive validity.

Finally, Figure 5.9 (right) also shows the measured ("o"-symbols) and predicted ("+"-symbols) overall ratings, based on the earlier determined mapping from continuous ratings to overall rating. Regardless of the relatively low VAF values for modeling the continuous ratings, the models are able to predict the overall ratings within one unit on the

rating scale. This falls within its measurable resolution, as participants can only change the rating value by discrete steps of 1.

5.5 Discussion

5.5.1 Rating Measurements

The experiment described in this chapter applied the continuous rating method of Cleij et al. [2018] – who tested short drives each with a single curve driving maneuver – in a realistic setting: a long highway scenario containing maneuvers characteristic for highway driving. Overall, the 30 participants were well able to distinguish the differences in incongruences between the different tested motion cueing conditions. The optimization-based Oracle (ORC) condition, with the objectively lowest motion mismatches, was indeed rated as having the most congruent motion. The CWA and HTC conditions were rated equivalently, indicating that no clear improvement due to the increased rotational rate gains in the HTC condition was found.

Generally, the experiment's motion cueing was rated positively, with the highest occurring average rating being just below 3. This is lower than the ratings of the urban experiment, as well as the rural rating data of Ellensohn [2020], where the highest ratings were around 5. This can be explained by the comparatively small mismatches present in the highway experiment, where $\max[\Delta\widetilde{S}_{f_x}] = 1.8\,\mathrm{m/s^2}$ and $\max[\Delta\widetilde{S}_{f_y}] = 2.0\,\mathrm{m/s^2}$. For the urban experiment, the largest mismatches were $\max[\Delta\widetilde{S}_{f_x}] = 2.0\,\mathrm{m/s^2}$ and $\max[\Delta\widetilde{S}_{f_y}] = 3.2\,\mathrm{m/s^2}$. On the other hand, considering that participants could have used the whole scale for rating the full range of incongruences that is *specific* to the current experiment, it can be considered surprising that participants only gave such low ratings. In previous research, it was such differences in interpretation of the continuous rating scale that required a *Model Transfer Parameter (MTP)* to predict between different sets of experiment data [Cleij et al., 2019]. However, considering that the training, before the measurement phase started, contained a strong false cue motion in the "CB" maneuver, inducing a mismatch of similar strength ($\sim 6\,\mathrm{m/s^2}$) as in the urban experiment, this may have served as an "anchoring", resulting in rating measurements that occur on the same relative rating range. Such anchoring with mismatches of equal strength between experiments is thus beneficial, as it increases the comparability of the acquired results.

5.5.2 Reliability

Consistent with the findings in Chapter 4, the reliability decreased with decreasing rating values (i.e., less incongruence), and vice versa. This implies that the better the motion cueing is rated, the less these ratings can be trusted. The reliability estimates are, however, lower than those found for the urban reliability model of Chapter 4. This shows that the lower reliabilities are not only caused by a lower number of repetitions (2 instead of 3), but also by the data generally being less reliable in the highway scenario. There are various causes for this, such as participants understanding the rating task better in the urban experiment than in the highway experiment. However, considering that the experiment set-up (simulator, briefing, experimenter, training procedure) were identical in both experiments, there is no specific reason to assume this was the case. Instead, a

more likely explanation is that the highway scenario is simply more difficult to rate for participants.

This could by due to the fact that maneuvers are more difficult to anticipate. In the urban scenario of Chapter 4, it was likely more clear where certain maneuvers took place (e.g., by visually recognizing upcoming corners and traffic lights). In the current highway scenario, the maneuvering is mainly based on the traffic surroundings, which may be more difficult to anticipate or remember from earlier runs. Thus, the more unpredictable maneuvers may have had a negative effect on the participants' ability to reliably evaluate the motion in every repetition. To further improve the reliability the anticipation of the more unexpected maneuvers, such as lane changes, the indicator sound may be simulated, or the future path of the vehicle in the simulation itself may be explicitly shown. This reduced ability to recognize the maneuvers between runs may also be related to the length of each drive: in urban the urban experiment of Chapter 4 each run lasted 255 s, whereas in the current highway experiment, each run lasted 430 s. Another, arguably the easiest method of boosting reliability is to increase the number of repetitions of each tested condition to three (as in Chapter 4) or more.

Another explanation for the lower reliability may come from the fact that the highway simulation contained relatively short (e.g., lane change) maneuvers for 430 s of driving, in combination with large parts of inactivity, compared to the longer maneuvers (e.g., curves, roundabout) in the shorter 255 s urban simulation of Chapter 4. As the reliability is calculated over the whole drive, and the ratings at all points in time contribute equally to the reliability, the reliability estimates of the present highway simulation may be dominated by the parts in the drive where not so much motion, and therefore not so many mismatches occur. In accordance with the findings of Chapter 4, the "local" reliability of the ratings in these inactive parts is likely to be low, even if the reliability of the ratings in the more active maneuvers is high. These relatively long parts with a low reliability would then bias the reliability of the whole drive, explaining the observed low values. This hypothesis will need explicit testing in future work. It would, however, also make it difficult, especially for continuous ratings, to define a cut-off value for reliability under which the ratings should be discarded, as is often done in literature (e.g., Ellensohn et al. [2019]). Nevertheless, as mainly the most incongruent moments in the simulation determine the overall impression, the presented approach of predicting continuous and overall ratings still upholds. Thus, whereas obtaining reliable data is important, and the reliability estimates are lower than literature suggests for considering the data reliable, the level of reliability obtained in the present ratings was adequate to predict the ratings with sufficient quality.

5.5.3 Rating Relationships

Chapter 4 showed that the most incongruent motion dominates participants' overall impression of the provided simulator physical motion in urban simulations. From the analysis of the correlation between the continuous and overall ratings for the current highway experiment, this finding could not be directly reproduced. It should be noted, however, that the analysis on these rating relationships may have suffered due to the choice of including the urban maneuvers at the end of the drive. While this had benefits for comparing the continuous rating model between both experiments, the overall ratings

may have been confounded due to the mixture of highway and urban contents. Whereas this approach showed to work well for the present simulation, it is thus recommended to develop explicit highway driving rating relationships for simulations in which only highway driving occurs. Nevertheless, the accuracy in predicting the overall ratings through the rating model and the rating relationship show that using the most incongruent point in the motion is still a well-functioning approach.

5.5.4 Model Predictions

Whereas the urban model of Chapter 4 worked well in predicting the general trends of the measured continuous ratings *of all three conditions*, the fit quality improved using the highway-specific model parameters. Considering the reasonable fit quality, the conditions were also not analyzed separately, e.g., by fitting condition-specific model parameters. The generalizability, by using an identical model structure and a simple change of model parameters, is a major strength for evaluating the motion cueing across a large selection of available MCAs.

Including an additional lateral specific force mismatch term, representing a response to the lane changes, resulted in the best prediction of the ratings. Through a lack of possible anticipation, as discussed in Subsection 5.5.2, these lane changes may have lead to a "surprise" response to participants, in which their response to the perceived mismatches is stronger than for the other maneuvers. Whereas this lane change model structure itself is still linear, the identification of lane changes to predict the ratings introduces additional difficulties, especially in closed-loop simulations. For example, whereas the urban model may be directly used as a cost function in future MPC algorithms, an additional prediction method to identify upcoming lane changes must be part of the optimization. In fact, this additional response may be the a result of the open-loop driving, as these maneuvers are difficult for the participant to anticipate. As this lack of anticipation is not present in closed-loop driving, as the driver knows its own intentions, this additional response may thus be less pronounced or even absent in closed-loop driving. Future work should therefore explicitly investigate the equivalence of closed- and open-loop driving in highway simulations.

5.6 Conclusion

This chapter described measuring and modeling continuous motion incongruence and overall ratings for a realistic highway driving scenario. The same experiment setup and procedures also considered for the urban experiment of Chapter 4, were used, to enable a direct comparison between the urban and highway scenario types. For the current highway experiment data, a lower reliability of the continuous rating data was found, which indicates that that the highway scenario was more difficult to rate than the previously tested urban scenario. However, the previous finding that rating reliability decreases for lower (i.e., more positive) ratings is reproduced, confirming that the more congruent the presented motion is, the less the measured ratings can be trusted. A simple model accounting for a first-order low-pass filter response to a weighted sum of lateral and longitudinal specific force mismatches, as introduced in Chapter 4, was found to still accurately predict the measured ratings. However, especially for the included

lane-change maneuvers the measured ratings were consistently underpredicted, likely due to the unpredictable nature of these maneuvers in open-loop driving settings, requiring an additional response to the lane change mismatches. As this improved the prediction of the most incongruent point in the continuous ratings, an improvement in the prediction quality of the overall ratings was subsequently obtained. Hier mist nog een statement over de overall ratings. Overall, these findings strengthen the validity of the proposed method for modeling continuous and overall ratings and, as a result, occurring incongruences can be more accurately analyzed across scenarios for evaluating and improving future motion cueing.

6

Incongruences in Rural Simulations

Key findings

- Like in urban and highway simulations, the reliability of ratings in rural simulations decreases with increasing congruence of the motion.
- The linear rating model of Chapter 4, which was already validated for urban and highway simulations, is also applicable to rural driving simulations.
- Ratings of the motion cueing are explained as a weighted response to the longitudinal (46%) and lateral (54%) specific force mismatches.
- A linear correction factor, scaled to the relative mismatches, is necessary to correct for experiments with different rating mismatch magnitudes.

This chapter is based on the following publication:

Title:	Models of Motion Incongruence Ratings in Rural Driving Simulations.
Authors:	M. Kolff, J. Venrooij, D.M. Pool, and M. Mulder.
Journal:	To be submitted.

Abstract

*This chapter investigates the reliability and model predictions of subjective motion incongruence ratings in rural driving simulations. The continuous rating data of three experiments from literature were used, in which participants evaluated the motion incongruences through continuous and overall rating methods. The experiments were highly similar to the set-up of earlier urban and highway simulation experiments, to allow for a comparison of scenario differences between the three scenario types. Several motion cueing algorithms were tested that represent different levels (low-medium-high) of motion cueing quality. The results show that – as in the urban and highway data – the more congruent the motion is, the less reliable the ratings become. The linear rating model is valid for predicting the rural continuous ratings based on the perceived longitudinal (*46.0*%) and lateral (*54.0*%) specific force mismatches. However, as earlier research has shown, a linear correction factor, describing the ratio of the relative mismatches, is necessary to generalize between experiments. Combined, these findings further show the validity of the continuous rating method, and complete the validation of the linear rating model across main driving simulation scenario types. This greatly improves assessment and prediction of motion cueing quality, without the need for expensive on-site simulator testing.*

6.1 Introduction

RURAL driving is one of the most important driving scenario types, together with urban and highway driving, for automotive vehicle development. To support the development of novel automotive technologies in such scenarios, simulators are highly useful tools. When equipped with a motion system, their limited workspace often results in an imperfect inertial motion reproduction. This causes mismatches between the vehicle reference and the simulator motion. When a mismatch is sensed by the test driver, the motion is *incongruent* [Cleij et al., 2018]. The presence of incongruences can affect driving behavior [Wang et al., 2023], reduce realism, and induce simulator sickness [Himmels et al., 2022b], which all negatively affect driving simulator experiment validity. For this reason, the design of the Motion Cueing Algorithm (MCA), which is the algorithm that drives the simulator motion, typically aims to reduce incongruences as much as possible. Being able to predict these incongruences beforehand would be a crucial tool, as that would allow for selecting the best-possible MCA and simulator combination without expensive on-site simulator testing.

To this end, Cleij et al. [2018] introduced a *continuous* rating method: while being driven around ("open-loop driving"), test drivers continuously give a subjective rating that aims to reflect their impression of the simulation quality at each point in time. The continuous rating data recorded in the relatively straight-forward driving scenario of Cleij et al. [2018] could be successfully predicted using a moving average model structure with the objective mismatches as inputs, i.e., the absolute difference between the vehicle reference and simulator motion. In Chapter 4, however, an improved model was proposed for a realistic urban simulation, which was shown to be also valid for highway simulations in Chapter 5. A first-order low-pass filtered response to a weighted sum of longitudinal and lateral specific force mismatches showed to adequately predict the measured continuous ratings. In both the urban and highway experiments, however, the reliability of the measured continuous ratings was shown to decrease with more congruent motion. Furthermore, the overall rating, a single rating that represents the overall quality of a drive and *is* feasible in closed-loop simulations (i.e., the test driver actively controlling the car), was shown to strongly correlate with the most incongruent point *during* that drive. Together with a continuous rating model, this thus provides a complete model for predicting subjective overall post-hoc ratings from objective simulator motion mismatches; a crucial tool in the analysis and trade-off of MCAs, simulators, and parameter settings.

The final step in the rating model development, and a current gap in literature, is the extension of the model validity to rural driving simulations. Three recent works have gathered valuable continuous and overall rating data in realistic rural simulations [Ellensohn et al., 2019a; Ellensohn et al., 2020; Ellensohn et al., 2019c]. These experiments were all performed on the same rural road and evaluated the quality of various types of MCAs. In these more realistic and complex rural simulations, the moving average model proposed by Cleij et al. [2018] was shown to be of limited predictive power [Ellensohn et al., 2019c]. Thus, the natural question arising is whether the findings regarding urban (Chapter 4) and highway (Chapter 5) simulations are valid in rural simulations as well, and whether the low-pass filter model of Chapter 4 *can* successfully predict the measured continuous ratings of the readily available rural data sets. This is especially

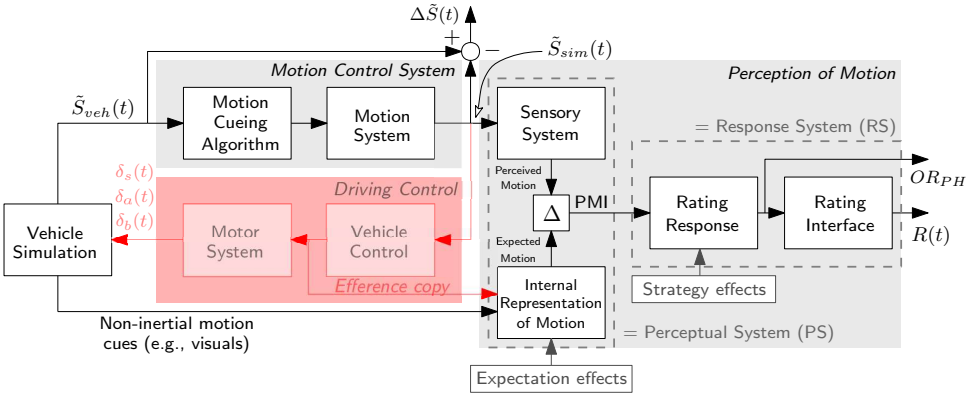


Figure 6.1: Block diagram of the driving and rating tasks, highlighting the difference between the subjective PMI and objective mismatches $\Delta\tilde{S}(t)$. The part "Driving Control" (red, transparent) is only present in closed-loop driving, and therefore not active in the present experiment.

challenging as it has been shown that participants use the available rating scale relative to the magnitude of the incongruences present in an experiment [Cleij, 2020], requiring a Model Transfer Parameter (MTP) to linearly correct for the scale of the mismatches *between* experiments of different mismatch magnitudes.

The contribution of this chapter is to analyze the reliability and predictability of the available rural data sets in an explicit comparison with the urban and highway rating data, respectively. To this extent, the modeling methodology of Chapter 4 is applied to the rating data sets presented in Ellensohn et al. [2019], Ellensohn et al. [2019], and Ellensohn et al. [2020]. First, the extent to which the low-pass filter model of Chapter 4 can be used without adaptations is verified. As a second approach, the model structure of Chapter 4 is refitted on the rural data sets, allowing variations in the relative weighting of the mismatch channels. Furthermore, the similarity of the three rural experiments is used to compared differences in the relative rating scale, which require a correction of an MTP in the model.

The chapter is structured as follows. Section 6.2 introduces the measuring and modeling methods. The experiment set-up and results of the considered experiments are described in Section 6.3. The results are shown in Section 6.4, followed by the discussion in Section 6.5. The chapter is concluded in Section 6.6.

6.2 Methods

6.2.1 Rating Task

In all three considered experiments, participants performed a rating task while being driven through a rural scenario, i.e., participants were driven around passively, rating than driving themselves. Open-loop driving allows the participant's full focus to be on evaluating the motion. The combined driving and rating task is shown in Figure 6.1. As both have been described extensively in Chapter 4 and Chapter 5, only the details of the rating task are repeated here.

For the rating task, drivers compared the *perceived motion* from the simulator to the *expected motion* from the real vehicle's motion, where the difference defines the Perceived Motion Incongruence (PMI). The simulator motion is perceived through the human vestibular and proprioceptive systems, indicated as "Sensory System". The internal representation and sensory system combined are indicated in Cleij et al. [2018] as the "PS". Participants must use an internal representation [Stassen et al., 1990] to form their expected motion based on the non-motion cues (such as the visuals) of the simulation. The internal representation can be affected by the participant's level of experience with the task (driving) and with the vehicle that is simulated, and can thus be different between participants, see Chapter 4.

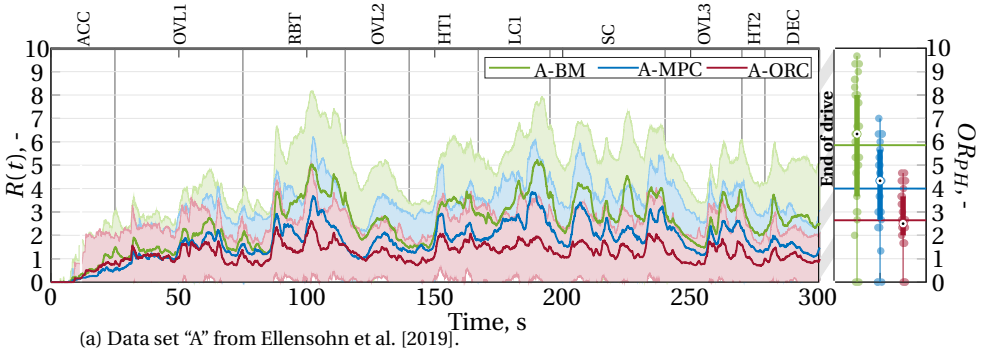
Whereas the PMI defines what is (in)congruent according to the participant, it is internal to the participant and cannot be measured directly. Therefore, a subjective Motion Incongruence Rating (MIR), which represents the PMI, can be measured [Cleij et al., 2018; Kolff et al., 2024b]. Differences between the PMI and MIR can occur due to differences in the rating strategy, i.e., how a participant converts the PMI to a rating, as well as any dynamics of the rating interface. Combined, these are described by the RS. As in Chapters 4 and 5, the MIR is given through a rotary knob [Cleij et al., 2018], which can be adjusted at any time, resulting in a time signal $R(t)$. The rating scale of the present experiments varied between 0 to 10, with steps of 1, where 0 indicates "fully congruent motion" and 10 indicates "highly incongruent motion".

In the experiments, participants rated the motion continuously, yielding a rating signal $R(t)$. Figure 6.2 shows the continuous (left) ratings as described in Ellensohn et al. [2019] (Experiment "A", Figure 6.2a), Ellensohn et al. [2019] (Experiment "B", Figure 6.2b), and Ellensohn et al. [2020] (Experiment "C"), Figure 6.2c. The various conditions will be explained in Section 6.3.3. The differences between the rating measurements have been described in the respective publications and are therefore not further analyzed in-depth. As in Chapter 4, the present experiments also recorded an overall post-hoc rating from the participants, as shown in Figure 6.2 on the right side. Overall post-hoc ratings aim to represent the overall quality of a drive. This rating were given verbally, yielding a single rating measurement OR_{PH}.

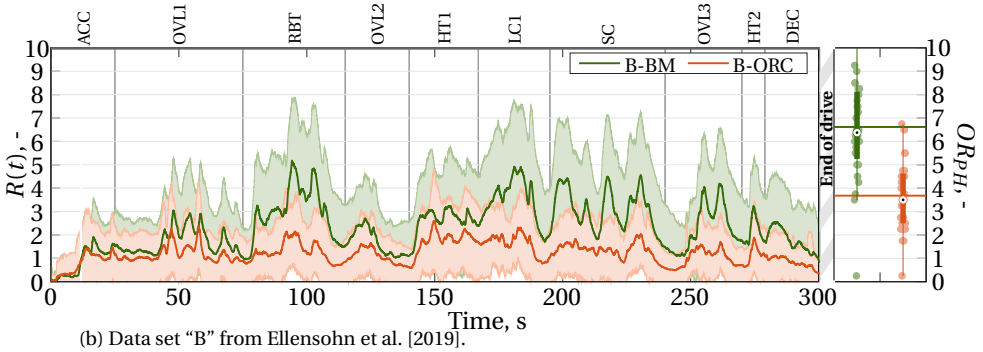
Although lacking temporal detail, overall post-hoc ratings have the benefit of being easy to extract and are viable for closed-loop driving simulator experiments, making them especially useful for validating rating predictions. Chapter 4 showed that the most incongruent point in an urban simulation correlates most with the overall post-hoc ratings, giving the linear relation $OR_{PH} = 2.0 + 0.8 \cdot \max[R(t)]$. Such a relation is especially useful to predict the overall ratings, suitable for closed-loop simulations, based on continuous rating models.

6.2.2 Reliability

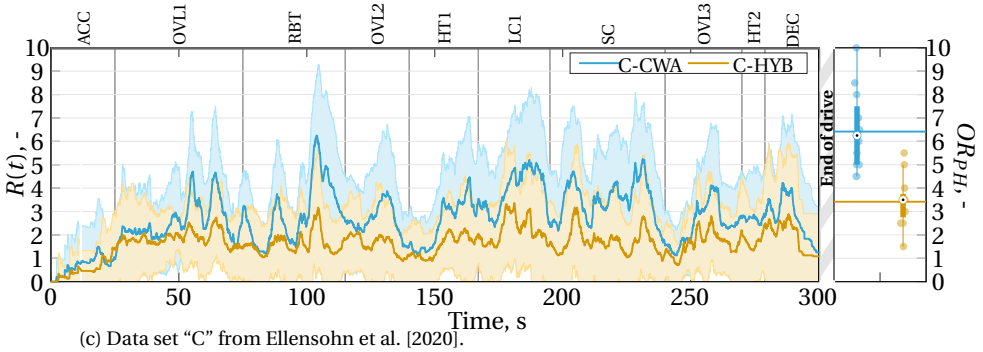
Although the continuous ratings are continuously recorded during a drive, the recorded ratings can be unreliable, for example if the participants do not properly understand the task. The reliability then describes how well the obtained data sets (in this case, the continuous ratings) can be trusted. Chapter 5 explicitly compared the reliability estimates of the continuous ratings between highway and urban simulations. To allow for an additional comparison to the present rural data, the reliability of the continuous ratings is



(a) Data set "A" from Ellensohn et al. [2019].



(b) Data set "B" from Ellensohn et al. [2019].



(c) Data set "C" from Ellensohn et al. [2020].

Figure 6.2: Continuous rating signals of the three experiments.

calculated using Cronbach's Alpha:

$$\alpha_{cp} = \frac{J}{J-1}\frac{\sum_j \sigma^2_{cjp}}{\sigma^2_{X_{cp}}}. \tag{6.1}$$

Here, J is the total number of repetitions and σ^2_{cjp} is the variance of the individual samples. The coefficient α is unbounded on the lower side, i.e., $[-\infty < \alpha \leq 1]$, where the upper bound of 1 indicates full reliability.

In Chapter 4, it was described that McDonald's Omega [McDonald, 2000] provides a more accurate estimation of reliability than the commonly used Cronbach's Alpha as the latter has been criticized [Sijtsma, 2009] as it can lead to underestimations of reliability. McDonald's Omega [Sijtsma, 2009; Trizano-Hermosilla and Alvarado, 2016], as introduced in McDonald [2000], is calculated as:

$$\Omega_{cp} = \frac{(\sum_j \lambda_{cjp})^2}{(\sum_j \lambda_{cjp})^2 + \sum_j (1-\lambda^2_{cjp})}, \tag{6.2}$$

where λ_{cjp} are the factor loadings. McDonald's Omega is always equal to or higher than Cronbach's Alpha [Sijtsma, 2009]. The factor loadings λ_{cjp} were determined using `factoran` in MATLAB R2018b, yielding Ω_{cp}.

6.2.3 Rating Model

The linear rating model of Chapter 4 is employed to predict the measured continuous ratings $R(t)$ as function of mismatch signals. Its structure is shown in Figure 6.3. The mismatches represent the difference in inertial motion (specific forces and rotational rates) between the vehicle motion $\widetilde{S}_{veh,m}(t)$ and the simulator motion $\widetilde{S}_{sim,m}(t)$, i. e., $\Delta\widetilde{S}_m(t)$, with $\widetilde{P}_m(t) = K_{\widetilde{P}_m}|\Delta\widetilde{S}_m(t)|$. m represents the mismatch channel, e.g., $m \in [f_x,\ f_y,\ \ldots]$. The model of Chapter 4 predicts the continuous rating of the *average* participant. Its structure consists of a first-order low-pass filter transfer function $H_m(j\omega)$ between the absolute mismatch signal $\widetilde{P}_m(t)$ and a modeled rating signal $\widetilde{R}(t)$:

$$\widehat{\widetilde{R}}(j\omega) = \sum_m K_{\widetilde{P}_m}\left(\frac{\omega_c}{j\omega+\omega_c}\right)\Delta\widehat{\widetilde{S}}_m(j\omega), \tag{6.3}$$

with the low-pass filter's cut-off frequency ω_c and the gains of the mismatch channels $K_{\widetilde{P}_m}$. The $\widehat{(\cdot)}$-terms indicate the Fourier transforms. The low-pass filter represents the participants' lagged response (representing the RS dynamics as shown in Figure 6.1) to the mismatches $\widetilde{P}_m(t)$. Chapter 4 showed that the continuous ratings of the conditions measured in that study could be largely explained when considering the longitudinal specific force mismatches $\widetilde{P}_{f_x}$, as well as the lateral specific force mismatch $\widetilde{P}_{f_y}$ (i.e., $m \in [f_x, f_y]$), with the parameters: $\omega_c = 0.36\,\text{rad/s}$, $K_{f_x} = 0.91$, and $K_{f_y} = 1.50$.

6.2.4 Model Fitting Procedure

The model is fitted in the time-domain using a Nelder-Mead simplex algorithm [van der El et al., 2018], as in Chapters 4 and 5. This method minimizes a sum-of-squared-errors cost function:

$$\arg\min_{\Theta} J = \sum_t [R(t) - \widetilde{R}(t|\Theta)]^2, \tag{6.4}$$

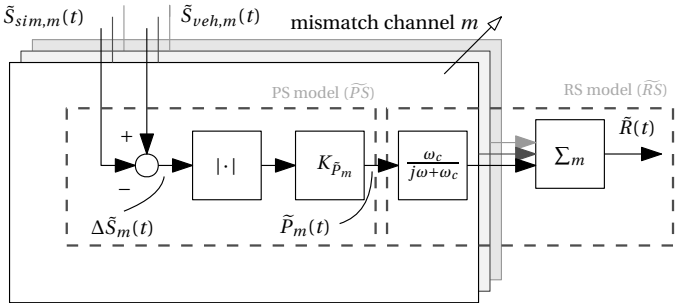


Figure 6.3: The linear rating model from Chapter 4.

where $\Theta = [\omega_c,\ K_{f_x},\ K_{f_y}]$ is the parameter set. As the gradient-based method does not guarantee attaining the global optimum, 30 iterations are performed with uniformly distributed random numbers between 0 and 3 as initial conditions. The parameters that lead to the lowest cost function value J are then selected as the final solution. The fit was always performed on all the rating data of the three conditions grouped together, similar to the "ALL" strategy in Chapter 4, as well as the models in Chapter 5.

6.2.5 Root Mean Square Error

The Root Mean Square Error (RMSE) is used to quantify the quality of the fit:

$$RMSE = \sqrt{\frac{1}{N}J} = \sqrt{\frac{1}{N}\sum_t [R(t) - \tilde{R}(t|\Theta)]^2}, \tag{6.5}$$

where $R(t)$ is the measured continuous rating signal (i.e., the reference) and $\tilde{R}(t|\Theta)$ is the modeled signal. The RMSE represents the quadratic mean of the difference between the measured and modeled signals. An RMSE of 0 thus describes a perfect fit, whereas there is no lower bound that represents the "poorest of fits".

Furthermore, the relative contribution of the mismatch channels to the ratings is calculated through the *influence factor* [Cleij et al., 2018]:

$$I_m = \sum_t \tilde{P}_m(t) / \sum_t \tilde{P}(t) \cdot 100\%, \tag{6.6}$$

where $\tilde{P}_m(t) = K_{\tilde{P}_m}|\Delta\tilde{S}_m(t)|$, see Figure 6.3, m is the mismatch channel, and $\tilde{P}(t)$ is the total modeled PMI, i.e., $\tilde{P}(t) = \sum_m \tilde{P}_m(t)$. The influence factor represents the relative contribution of each mismatch channel. The sum of all channels' influence in the model is therefore always 100%. The calculation of the influence factor allows for comparing the contribution of each mismatch channel balance between the urban (Chapter 4), highway (Chapter 5), and rural ratings.

6.3 Experiment Set-up

6.3.1 Scenario

The scenario was the same in all three experiments. It consisted of typical rural maneuvers, as shown in Figure 6.4:

- ACC: Acceleration
- OVL1: Overland 1
- RBT: Roundabout
- OVL2: Overland 2
- HT1: Hill top 1
- LC1: Left corner 1
- SC: S-curve
- OVL3: Overland 3
- HT2: Hill top 2
- DEC: Deceleration

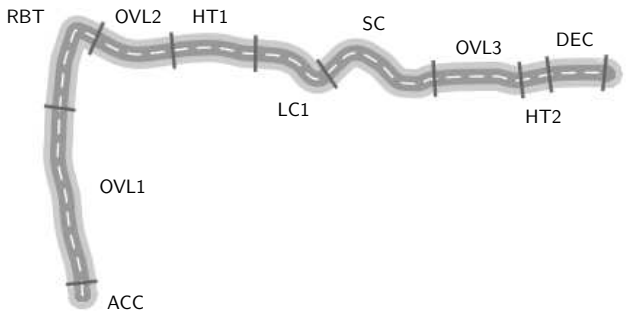


Figure 6.4: Overview of the rural route with the maneuvers indicated.

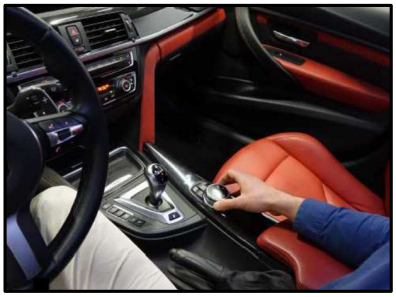
6.3.2 Apparatus

In all three experiments, the experiment-setup was highly similar as in Chapter 4 and Chapter 5: BMW Group's Ruby Space simulator (Figure 6.5a) was used. Its motion system consists of a hexapod on top of a tripod system, resulting in nine DoFs. The tripod adds additional workspace in longitudinal, lateral and yaw directions. Participants could give their rating using the iDrive knob on the center console (Figure 6.5b), which was sampled at 100 Hz. The 240° projection screen of the simulator showed the visuals, the current velocity, and the current rating value in the form of a "rating bar" [Cleij et al., 2018]. The size and color of the rating bar changed from 0 (small, white) to 10 (long, red), representing the current rating given by the participant. These settings were all identical to the experiments in Chapters 4 and 5.

6.3.3 Independent Variables

In the experiments, various experimental conditions were tested, as shown in Figure 6.2. These conditions were based on different MCA architectures and thus have different mismatches between the vehicle reference and simulator motion, see Figure 6.6. Experiment "Rural A" [Ellensohn et al., 2019a] evaluated three different MCAs. First, a "benchmark algorithm" (BM) was used [Bruschetta et al., 2017], which was not specifically tuned for the simulation. The algorithm filters the motion into low, medium, and high frequency components. An optimization-based approach is then used to solve an

(a) The simulator while moving.



(b) A test driver using the rating knob.

Figure 6.5: The experiment set-up, as in the experiments of Chapters 4 and 5. Figures adapted from Ellensohn et al. [2019]).

optimization problem using a MPC approach. Second, a full MPC was used, also capable of real-time simulation. Third, an optimization-based algorithm was used, which optimizes the simulator motion along the complete recorded drive offline [Ellensohn et al., 2019c], denoted as the Oracle ("ORC"). This algorithm can only be used in open-loop simulations, but allows for the investigation of how the available simulator workspace may be fully exploited [Kolff et al., 2022]. This MCA provides the lowest possible mismatches on the simulator, mainly visible in the longitudinal (Figure 6.6a) and lateral (Figure 6.6b) specific force mismatches, as well as the yaw rate mismatch (Figure 6.6f). It thus serves as the upper bound of motion cueing quality possible on the simulator. In experiment "B" [Ellensohn et al., 2019c], the same BM and ORC algorithms from experiment "A" were again tested.

In experiment "C" [Ellensohn et al., 2020], two MCAs were tested. First, a CWA was used [Conrad et al., 1973; Reid and Nahon, 1985], where the vehicle motion is distributed over the hexapod (high-frequency), tripod (medium-frequency), and tilt-coordination (low-frequency) channels, and washed-out using second-order high-pass filters. In the second, "hybrid algorithm", the optimal simulator motion is calculated beforehand using an Oracle strategy. Then, the algorithm can be applied in real-time by "washing out" any deviations from the optimal control output. Therefore, this MCA is also real-time capable and may potentially approximate the Oracle motion cueing in closed-loop simulations. In all conditions, the rotational rates ω_x and ω_y were tuned to be close to the perceptual threshold (< 3 deg/s) [Reymond and Kemeny, 2000], see Figures 6.6d and 6.6e.

In each of the three experiments, a pre-recorded drive was used, such that the same drive was used for all tested conditions. The recording of the drive, however, differed between the three experiments, such that slight differences in the driving style are present *between* the three experiments. Most notably, the initial acceleration of the vehicle in experiment "C" was later than that of experiments "A" and "B", leading a slight shift of ca. 5 s in when the maneuvers occurred (see Figure 6.6a). This has no effect on any of the results. Furthermore, participants were instructed to not actively rate the acceleration ("ACC") and deceleration ("DEC") maneuvers at the beginning and the end of the drive, respectively.

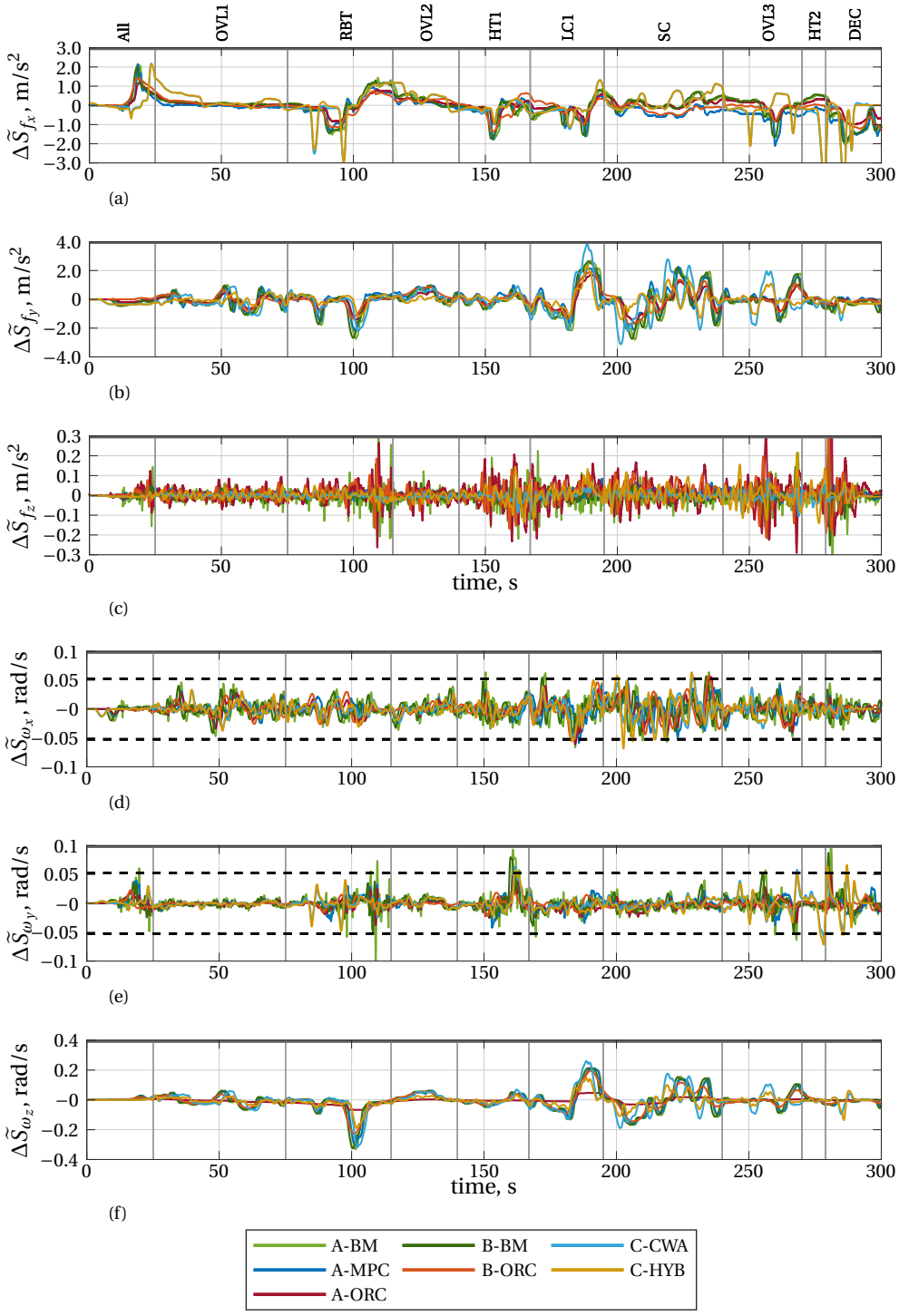


Figure 6.6: Mismatches of the experimental conditions tested in the three experiments. Grey vertical lines indicate the separate maneuvers. The dashed lines indicate the rotational threshold of 3 deg/s [Reymond and Kemeny, 2000], relevant for the use of tilt-coordination.

6.3.4 Participants and Procedures

The participant groups and experiment procedures have been discussed in detail in the respective source publications. Considering the importance for calculating reliability, the number of participants are nevertheless restated here: 35 participants in experiment "A", 30 in experiment "B", and 12 in experiment "C". Furthermore, experiment "A" had 3 repetitions of each condition (9 runs in total per participant), experiment "B" had 4 repetitions (8 runs in total), and experiment "C" had 2 repetitions (4 runs in total). All experiments contained a training segment before the actual experiment to let participants get accustomed to the simulation environment and the rating task. Although the rating task was the same, the present experiments had a different experiment leader than in the urban and highway experiments, which may cause differences in the instruction and strategy effects (see Chapter 4) compared to the urban and highway experiments.

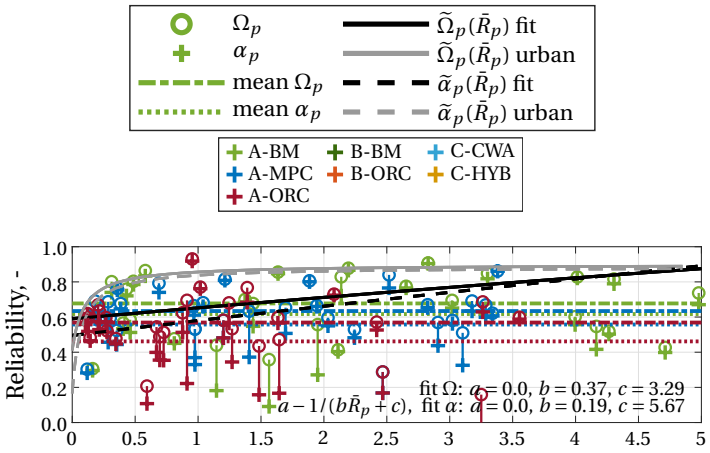
6.4 Results

6.4.1 Reliability

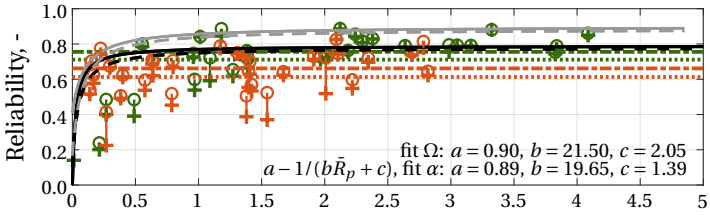
The reliability estimates of the three experiments' continuous rating data are shown in Figures 6.7a-6.7c. The average reliabilities over the participants are indicated by the dotted lines, their values are given in Table 6.1. The average reliabilities are highest for experiment "B", followed by experiment "A" and then "C". This follows the number of repetitions of each condition of the experiments (four, three, and two, respectively). Because only two repetitions of each condition were tested in experiment "C", Ω could not be calculated here. Therefore, only values of α are displayed in Figure 6.7c and Table 6.1 for the conditions C-CWA and C-HYB. For experiment "A" and "B", both reliability metrics α and Ω were calculated.

The individual data points in Figures 6.7a-6.7c represent the reliability estimates per participant, displayed for each condition separately, as a function of the corresponding average rating in that condition $\bar{R}_p$ (i.e., the rating averaged over time and averaged over the repetitions for a given participant). The vertical bars show the difference between both metrics (experiments "A" and "B" only). As described in Savalei and Reise [2019] and Chapter 4, values of Ω should be always equal to or higher than those of α, which is indeed the case here. Differences are prominent (up to 0.4), mainly for participants for whom α is low, in line with predictions by Savalei and Reise [2019]. It is thus at more congruent motion where the use of Ω is beneficial, as it provides a significantly higher estimate of the true reliability, avoiding the wrongful conclusion that some data sets are unreliable.

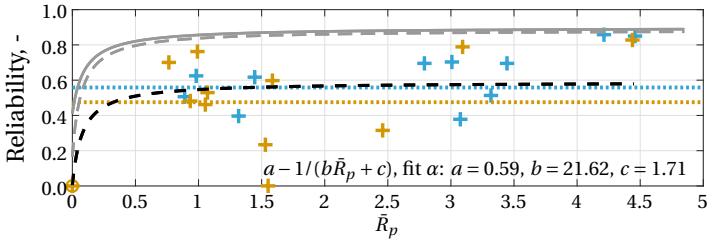
A key finding of Chapter 4 was that the lower the ratings are (i.e., more congruent), the more unreliable they are. The regression fit of Chapter 4 of the form $r = a - 1/(b\bar{R}_p + c)$ is used to describe this, with a, b, and c the fit coefficients, and r either $\tilde{\alpha}$ or $\tilde{\Omega}$. In Figures 6.7a-6.7c, these fits ($a = 0.89$, $b = 19.65$, $c = 1.39$ for $\tilde{\alpha}$, $a = 0.90$, $b = 21.50$, $c = 2.05$ for $\tilde{\Omega}$) are represented by the grey dashed ($\tilde{\alpha}$) and solid ($\tilde{\Omega}$) lines. Similarly, the regression fit is performed on the present rural data sets, represented by the black dashed ($\tilde{\alpha}$) and solid ($\tilde{\Omega}$) lines. For experiment "B" (Figure 6.7b), the data indeed follow the trend of decreasing reliability for lower ratings, although the reliabilities are lower than what the regression fit of Chapter 4 predicts. Despite the additional repetition (four) compared to



(a) Experiment "A".



(b) Experiment "B".



(c) Experiment "C".

Figure 6.7: Reliability coefficients α_p and Ω_p of all subjects per condition, showing that reliability decreases with lower ratings.

Table 6.1: Reliability values for the seven conditions of the experiments "A", "B", and "C". For the continuous ratings ("Cont."), the values represent the average of the reliabilities per subject. For the overall ("Ovrl."), the reliabilities are calculated over all subject, yielding one value per subject.

		A-BM	A-MPC	A-ORC	B-BM	B-ORC	C-CWA	C-HYB
Cont.	α	0.62	0.56	0.46	0.71	0.61	0.56	0.47
	Ω	0.68	0.63	0.57	0.76	0.66	-	-
Ovrl.	α	0.95	0.88	0.86	0.85	0.81	0.66	0.80
	ω	0.96	0.88	0.87	0.85	0.82	-	-

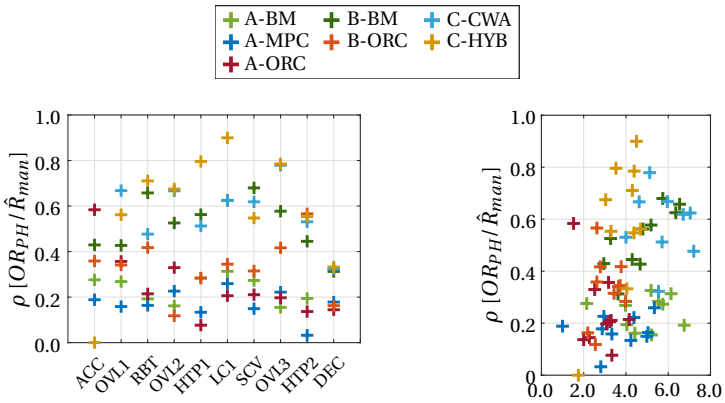
the urban data of Chapter 4 (three repetitions), the reliability is lower. This can indicate either a lower rating task understanding of the participants or the rural simulation being inherently more difficult to rate.

For experiment "A", this effect is only partially reproduced: one group of the data follows the regression fit of Chapter 4 (top/top left), with reliability decreasing for lower ratings. The remaining data show a more scattered distribution of reliability, without any apparent dependency on the rating magnitude. The presence of this grouping suggests that the lower reliability of some participants is caused by a lower rating task understanding, rather than rural simulations being inherently more difficult to rate (as the motion was the same for all participants). As a result of these two groups, the best-fit regression fit is in this case a linear model, which still describes decreased reliability for lower ratings (i.e., more congruent motion). Finally, for experiment "C", no clear patterns are visible due to the low number of participants (12). Reliability of the continuous ratings here is generally low, which is at least partly caused by only having two repetitions of each condition.

The reliability values of the overall post-hoc ratings are also shown in Table 6.1. These values summarize the reliability of the whole data set and are thus a function of the amount of subjects. The values are therefore also inherently higher than the estimated reliability estimates of the continuous ratings. Similar trends are visible, however, where, with the exception of C-HYB, the best rated conditions (i.e., A-ORC and B-ORC) have the lowest reliability.

6.4.2 Rating Relationships

Chapter 4 showed that a strong correlation between the most incongruent point and the overall rating existed in urban driving simulation. This analysis is repeated here for the rural experiment data. Figure 6.8a shows that the highest Pearson correlation (ρ = 0.90) between the continuous and overall post-hoc ratings occurred for the maneuver "LC1" (left corner 1), in the "HYB" condition of experiment "C". Figure 6.8b shows the correlation coefficients as a function of the average continuous rating. Here, no clear trends are visible, in contrast to the urban data. This shows that the most incongruent point in the continuous rating did not necessarily correlate most with the overall ratings. For further model predictions, the linear regression relation $OR_{PH} = 2 + 0.8 \cdot \max[R(t)]$, found in the urban analysis (Chapter 4), is still used.



(a) Per maneuver (b) Rating power

Figure 6.8: Pearson correlation coefficients between the overall ratings (OR_{PH}) and the maximum of the continuous ratings within each maneuver ($\hat{R}_{man}$).

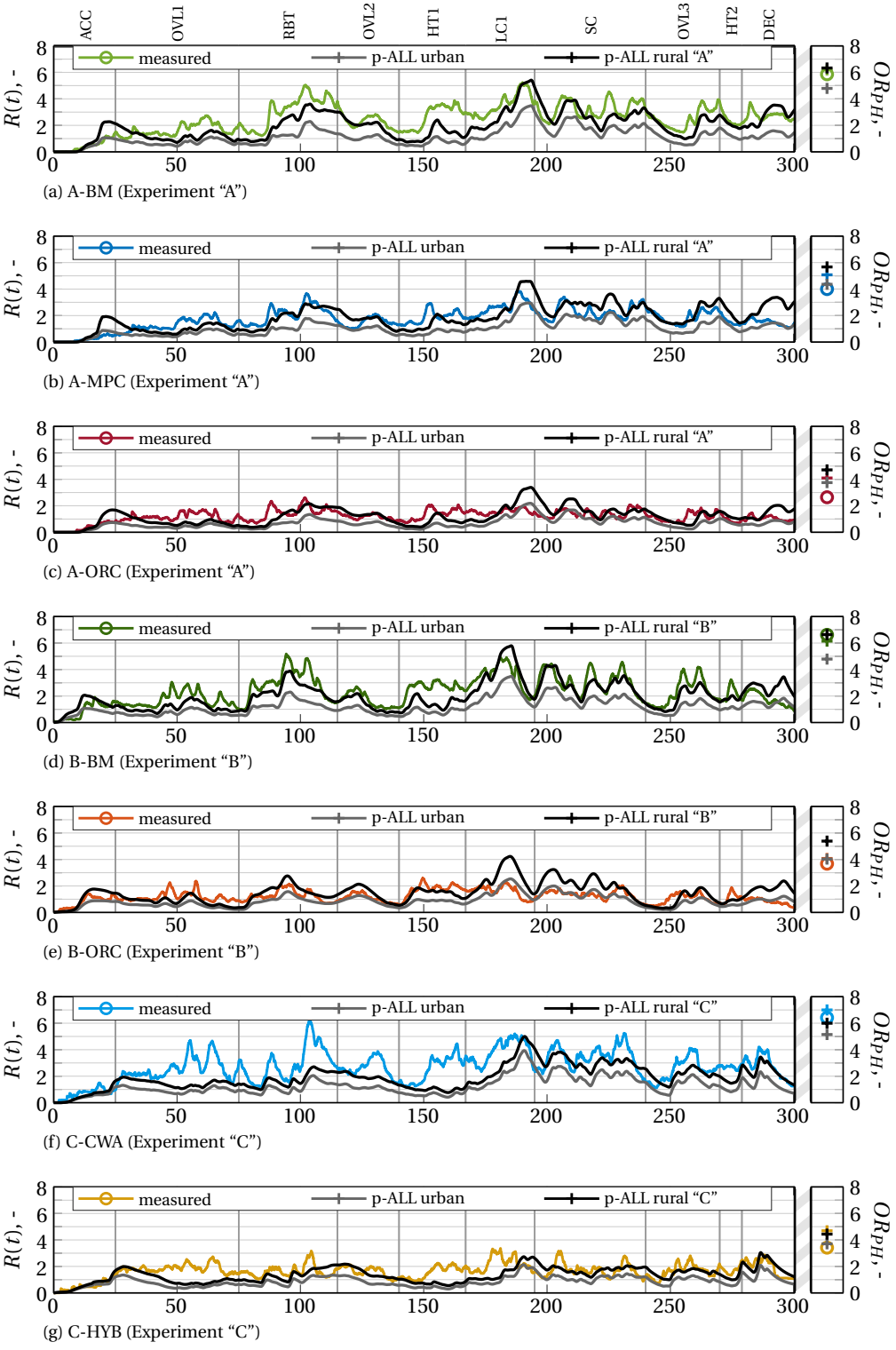
6.4.3 Rating Models

The model predictions, together with the measured continuous ratings, are shown in Figure 6.9. The "p-ALL urban" model from Chapter 4 generally underestimates the measured ratings, although it describes the variations in the ratings reasonably well. The three models "p-ALL rural A/B/C", fitted on each of the respective data sets, provide a better fit. As both the urban and rural models generally thus describe the variations in the ratings reasonably well, the VAF (as used in Chapters 4 and 5) was considered less suitable to compare between the models. This was the prime motivation for calculating the RMSE instead, whose values are shown in Table 6.2. These values confirm that the rural-specific models indeed performed better than the "p-ALL urban" model. Note that the acceleration ("ACC") and deceleration ("DEC") maneuvers were not actively rated by the participants, and therefore also not included in the analysis.

Table 6.2: RMSE values for the three considered continuous rating models.

		p-ALL urban	p-ALL rural A	p-ALL rural B	p-ALL rural C
Exp. "A"	BM	1.52	0.85	-	-
	MPC	0.85	0.70	-	-
	ORC	0.67	0.57	-	-
Exp. "B"	BM	1.32	-	0.76	-
	ORC	0.56	-	0.70	-
Exp. "C"	CWA	1.67	-	-	1.18
	HYB	0.96	-	-	0.69

There is a clear difference visible in the prediction quality, however, between different test conditions. For the conditions with low ratings (e.g., "ORC"), the ratings are not well predicted. This may be due to the low reliabilities of these data sets (see Table 6.1). Specifically for experiment "C" (Figures 6.9f and 6.9g), the low prediction quality may be



(a) A-BM (Experiment "A")

(b) A-MPC (Experiment "A")

(c) A-ORC (Experiment "A")

(d) B-BM (Experiment "B")

(e) B-ORC (Experiment "B")

(f) C-CWA (Experiment "C")

(g) C-HYB (Experiment "C")

Figure 6.9: The measured continuous (left) and overall (right) ratings of three conditions, each with the four applied models.

Table 6.3: Estimated parameters of the parametric models, including the parameters of the urban and highway models. For the rural models, the corrected parameters are calculated by multiplying with the MTP.

model	ω_c [rad/s]	K_{fx} [-]	K_{fy} [-]	MTP [-]	K_{fx}·MTP [-]	K_{fy}·MTP [-]
p-ALL rural A	0.35	2.36	2.07	0.67	1.57	1.38
p-ALL rural B	0.42	1.75	2.34	0.75	1.31	1.76
p-ALL rural C	0.25	1.34	2.50	1.11	1.49	2.78
p-ALL urban	0.36	0.91	1.50			
p-ALL highway	0.21	1.09	1.46			

explained by both the low number of participants and the number of repetitions (2) per condition, leading to unreliable rating data, as shown in Table 6.1 and Figure 6.7c.

The parameter values of the three fitted models are shown in Table 6.3, together with the estimated parameters found for the urban (Chapter 4) and highway (Chapter 5) models. The gains in the models of the three rural experiments are clearly higher than in the urban and highway simulations. This shows that participants gave higher ratings for incongruences of similar magnitude in the rural simulations compared to the urban and highway simulations. Using the approach of Cleij [2020], by introducing a Model Transfer Parameter (MTP) to correct for the difference in relative PMI *between* the experiments, this can be corrected for. The MTP is calculated as the ratio of the largest PMI occurring over time t and in the mismatch channels m in two experiments a and b:

$$\mathrm{MTP}_{a\to b} = \frac{\left(\max\limits_{m,t}[K_{\widetilde{P}_m}\Delta\widetilde{S}_m(t)]\right)_a}{\left(\max\limits_{m,t}[K_{\widetilde{P}_m}\Delta\widetilde{S}_m(t)]\right)_b}. \tag{6.7}$$

The MTP is calculated using, for rural "A" and "B", $\max\limits_{m,t}[\Delta\widetilde{S}_m(t)] = 2.9\,\mathrm{m/s^2}$, for rural "C": $\max\limits_{m,t}[\Delta\widetilde{S}_m(t)] = 4\,\mathrm{m/s^2}$, which in all three cases comes from the lateral specific force. For the urban and highway data sets, the largest mismatches were induced in the training drives: $\max\limits_{m,t}[\Delta\widetilde{S}_m(t)] = 6\,\mathrm{m/s^2}$, also in the lateral specific force. This yields values of MTP = 0.48 for experiments rural "A" and "B", and MTP = 0.67 for experiment rural "C", when comparing to both the urban and highway experiments. In words, this means that the mismatches in the rural experiments were smaller compared to the urban and highway experiments, such that participants likely adapted their use of the rating scale accordingly Using these MTP values, the "corrected" parameter values are also shown in Table 6.3, which indeed results in more comparable parameter values between all five experiments. Considering that the model "p-ALL rural A" of experiment "A" performed best (Table 6.2), and is based on three data sets, rather than two, this model is considered the most representative for rural driving.

To investigate the balance of the longitudinal and lateral specific force contributions, the corresponding influence factors are calculated. These are shown in Table 6.4, including the values found in the urban (Chapter 4) and highway (Chapter 5) experiments, for reference. A difference in balance between the contributions of $\widetilde{P}_{f_x}$ and $\widetilde{P}_{f_y}$ is visible

Table 6.4: Influence factors of the identified channels in the parametric models.

data	model	$I_{\tilde{P}_{fx}}$	$I_{\tilde{P}_{fy}}$
Rural "A"	p-ALL rural A	45.8%	54.2%
Rural "B"	p-ALL rural B	37.8%	62.2%
Rural "C"	p-ALL rural C	44.0%	56.0%
urban	p-ALL urban	34.0%	66.0%
highway	p-ALL highway	45.0%	55.0%

between the three rural experiments. For the model "p-ALL rural A", a highly similar balance of influence factors to the highway data exists. This contains a higher influence of the longitudinal specific force mismatch than the urban rating data, and a lower influence of the lateral specific force mismatch. Whereas rural simulations may indeed contain more longitudinal accelerations and higher speeds, this may also be caused by the hills that characterize the rural scenario, which also induce longitudinal specific force mismatches.

The final analysis step concerns the prediction of the overall ratings based on the modeled continuous ratings. These predictions are shown on the right in Figure 6.9, together with the measured overall ratings. As done for the continuous ratings, the overall ratings are predicted using the 'p-ALL urban" model, as well as using the three refitted rural models. The overall rating predictions are reasonably accurate (0-2 rating points deviation), but in some cases less accurate than in the urban and highway simulations; there the difference was always within a single rating point. This may have two reasons. First, the most incongruent point according to the rating models was always the "LC1" (left corner 1) maneuver, for which the models generally overestimated the participants' rating response. This then automatically results in an overestimation in the prediction of the overall ratings as well. Second, as no precise relation between the continuous and overall ratings could be established, the urban relationship of Chapter 4 is used here ($OR_{PH} = 2 + 0.8 \cdot \max[R(t)]$), which may not be fully representative of the rural scenario. Nevertheless, for the conditions with the most accurate continuous rating predictions in the "LC1" maneuver (A-BM, B-BM, C-CWA), these predictions fall within half a rating point.

6.5 Discussion

6.5.1 Reliability

As also done for the urban (Chapter 4) and highway (Chapter 5) simulations, the reliability of the acquired ratings was explicitly investigated. The phenomenon in which reliability decreases for more congruent motion, as consistently observed in the urban and highway simulations, was only partly observed. Especially in experiment "A", there were many participants that provided continuous rating data that were less reliable than the reliability model predicted. Considering this was the case for a *selection* of participants, this may be a sign of insufficient training or rating task understanding. Future

work should investigate the effect of task understanding on the data reliability in more detail.

Another important aspect in the analysis, however, is the magnitude of the continuous ratings. The fact that an MTP was required to predict the ratings shows that participants used a different mapping of perceived PMI to the rating scale than in the urban and highway simulations. Linearly correcting *measured* ratings for this should be done with great care, however. Instead, it is recommended to acquire rural rating data with mismatches equal in magnitude to those in the urban and highway experiments, such that this effect can be verified and explicitly corrected for.

6.5.2 Rating Relationships

A relationship between the incongruence of the motion, and the correlation between the overall and continuous ratings, could not be established with strong confidence for the rural experiment datasets analyzed in this chapter. An explanation for this is that maneuvers were more long-lasting compared to the urban simulation of Chapter 4, such that the maneuvers could not be captured by a single overall rating. The fact that the urban regression fit provided accurate predictions of the overall ratings, however, still shows that the most incongruent point provides a good predictor for the overall ratings, also in rural simulations. Future work should thus provide a more fundamental perspective on the formation of the overall ratings and, specifically, scenario differences. This is ideally done by comparing overall ratings of different scenarios within a single experiment. This could reveal whether inherent scenario differences exist, or if the lack of a strong relationship was a result of the experiments themselves, such as a lack of task understanding on the overall ratings.

6.5.3 Rating Models

The analysis of the rating models further confirmed the validity of the urban rating model structure, with the response to the longitudinal and lateral specific force mismatches, and comparable similar parameters. This shows that, like in urban and highway simulations, these two mismatch channels are the primary cause of incongruent motion in rural driving. The larger dependency on the longitudinal specific force mismatches, which matched that of highway simulations, may be explained through the higher speeds and additional hills in the simulations, which are typical for rural simulations.

Whereas the variations could be described using the urban model, the rating signal nevertheless needed to be corrected for using an MTP, as proposed in Cleij [2020]. Considering the similarity of the largest motion mismatches in the experiments of Chapters 4 and 5, this also explains why no MTP was necessary in the comparison of those data sets. Note that in future model predictions no MTP is required to predict between experiments, provided that *equivalent* parameters are used. Only when the predicted ratings would also be validated subjectively, would an MTP-based correction be necessary to correct between the predicted and measured ratings.

Whereas the continuous rating data of the urban and highway experiments contained mostly short “peaks” of mismatches in distinct maneuvers, the mismatches in the present rural data were more long lasting, causing overall trends in the ratings over

time. These trends were of added difficulty in predicting the ratings, as especially the magnitude of the continuous ratings showed to be difficult to predict using a single, linear rating model. As the magnitude is of high importance for determining the most incongruent motion, future work may thus investigate the suitability of a more advanced, non-linear rating model that can weigh mismatches differently over time, for example depending on the type of maneuver. This may be the necessary step to, for example, the C-CWA condition described in this chapter, in which the variations, but not the magnitude could be predicted accurately. For these approaches, model evaluations using the RMSE, as used in the present analysis, rather than the VAF (used in Chapters 4 and 5), are recommended.

Nevertheless, the successful model predictions, especially based on the rating data of conditions with large mismatches (i.e., A-BM, B-BM), can be directly applied in the analysis and trade-off of future MCA settings. With the validity of the linear rating model tested in the rural scenario as well, the trinity of rating predictions on the three most important scenario types (urban, highway, and rural) is now complete. Whereas the parameters differ slightly between the scenario types, given the fixed model structure, the model can be universally applied. In the evaluation and trade-off of various MCAs and simulators across the three scenario types, which is the primary use-case in this dissertation, the model can predict the quality of the motion without expensive on-site simulator testing. Another use-case is that the model can be used in MPC applications, where its linear model structure can be directly applied in the cost function to include a subjective component as part of the optimization. The subjective experience between different scenarios can then be directly implemented using a different parameter setting.

6.6 Conclusion

This chapter described measuring and modeling continuous motion incongruence ratings in realistic simulated rural driving scenarios. By analyzing the continuous and overall rating data of three existing rural driving experiments, a direct comparison to the other two main scenario types, urban (Chapter 4) and highway (Chapter 5), was made. Like in urban and highway simulations, reliability decreases for lower (i.e., more positive) ratings, confirming that the more congruent the presented motion is, the less the acquired ratings can be trusted. The generally lower reliability of the rural simulation shows that it is more difficult to rate than urban simulations. The low-pass filter response to the lateral and longitudinal specific force mismatches are again sufficient for predicting the measured ratings. However, an additional linear correction factor in the ratings was required to successfully predict between experiments. These findings complete the validation of the rating measurement and modeling methods in the three main scenario types (urban, highway, and rural), providing unmatched insights *across scenarios* in the evaluation, trade-off, and development of future motion cueing.

7

Incongruences in Closed-loop Driving

Key findings

- In urban simulations, participants being driven rate the motion cueing equivalently to when driving themselves.
- Similar to the findings of Chapter 4 on the overall impression of a whole drive, the summarizing impression of a maneuver's motion is dominated by the most incongruent point during that maneuver.
- The linear continuous rating model of Chapter 4, developed on data of a different participant group and simulator, can predict continuous ratings across experiments.

This chapter is based on the following publication:

Title:	Predicting Motion Incongruence Ratings in Closed- and Open-Loop Urban Driving Simulation.
Authors:	M. Kolff, J. Venrooij, E. Arcidiacono, D.M. Pool, and M. Mulder.
Journal:	IEEE Transactions on Intelligent Transportation Systems, 26(1), pp. 517-528, 2025.

Abstract

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

7.1 Introduction

DRIVING simulators are essential tools in the development of future driving technologies due to their ability to create safe and repeatable test conditions. When equipped with a motion system, their limited workspace often induces mismatches between vehicle and simulator inertial motion [Qazani et al., 2022]. While some mismatches are not perceived by the driver, the motion is *incongruent* if the driver *does* notice a deviation between their expectation of the real vehicle motion and the simulator motion they actually perceive [Cleij et al., 2018; Kolff et al., 2024b]. Incongruent motion can lead to an impaired perceptual fidelity of the simulation and induce simulator sickness [Irmak et al., 2021]. Therefore, the development, evaluation, and trade-off of MCAs typically aim at selecting the option with potentially the least incongruences. Acquiring and validating this information currently requires performing subjective evaluations in a driving simulator. Being able to *predict* such ratings through objective measures would be a crucial advancement [Casas-Yrurzum et al., 2020]. They would allow for rapid, systematic, and cost-efficient assessment of MCAs and guide developments, e.g., of MPC algorithms [Lamprecht et al., 2021]. However, making such predictions is notoriously difficult. For example, it is known that drivers generally consider scaled motion as more realistic than fully congruent one-to-one simulator motion [Berthoz et al., 2013], of which the cause is not yet understood.

In most driving simulations, drivers control the simulated vehicle themselves ("closed-loop"). Due to differences in driver behavior and driving style, each drive is different, resulting in different experiences of motion. Existing models to objectively predict subjective ratings [Cleij et al., 2018; Ellensohn et al., 2019c; Kolff et al., 2024b] are based on ratings of "open-loop" driving. Here, human drivers are driven around as passengers. The fact that they do not need to provide any steering control inputs has two crucial advantages. First, open-loop driving allows for performing multiple identical repetitions of exactly the same drive, e.g., to obtain more reliable subjective rating data [Kolff et al., 2024b]. Second, the absence of a driving task allows for a more invasive rating task, such as letting drivers *continuously* rate the motion cueing through a rating knob [Cleij et al., 2018; Kolff et al., 2021], providing unmatched insights into *when* and *where* in the simulation (in)congruent motion occurs. Due to the high temporal resolution of continuous ratings, their relation to objectively calculated mismatches between vehicle and simulated motion can be captured in mathematical models, which in turn allow for *predicting* continuous ratings [Cleij et al., 2018; Ellensohn et al., 2019c; Kolff et al., 2024b]. However, as drivers are expected to continuously assess their perceived motion and operate a rating knob with one hand, the continuous rating method cannot be used in closed-loop scenarios, i.e., when drivers need to operate the steering wheel with both hands. Rating methods that *are* suitable for closed-loop driving, such as providing a single rating after each drive or maneuver, are of such lower resolution that they are much less suitable to be used in a modelling approach.

Thus, it would be extremely useful if the high-resolution open-loop prediction models of continuous ratings can be used in the design, evaluation, and testing of motion cueing for closed-loop driving simulation. However, the central assumption of the continuous rating method, i.e., that it is representative of closed-loop simulations, has never been tested. Differences between the two driving methods might occur due to percep-

tual differences [Nesti et al., 2016; Valente Pais et al., 2012] or due to changes to the internal representation of motion [Kolff et al., 2024b]. With both the strengths and limitations of the continuous rating method in mind, three gaps are identified that would need to be answered to investigate whether continuous ratings of open-loop driving, and their predictions models, can be used to predict ratings of closed-loop driving. First, a rating model must be used to predict measured continuous ratings. This is challenging because existing rating models [Cleij et al., 2018; Ellensohn et al., 2019c; Kolff et al., 2024b] have not yet been confirmed to hold predictive power *between* experiments. Second, explicit rating relationships must be developed, that can link the continuous rating method to rating methods that are possible in closed-loop driving, such as after each maneuver or after the whole drive. Finally, no work so far has investigated the equivalence of open-loop and closed-loop driving. Their equivalence would be a requirement to be able to make predictions of closed-loop drives based on the open-loop rating models.

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

This chapter's main contribution is a complete, three-step approach that allows for predicting overall and maneuver-based subjective ratings of closed-loop driving as a function of objective motion cueing mismatch signals. First, a model for predicting continuous motion incongruence ratings from previous work [Kolff et al., 2024b] (Chapter 4) is employed to test whether the recorded continuous ratings (third phase) can be predicted. Although the same urban scenario of Kolff et al. [2024] is simulated, a different simulator, MCA parameters, and participant group were used. Second, it is investigated whether predictive relations exist from the continuous rating (third phase) to the overall and maneuver-based ratings which can be obtained in closed-loop driving (second phase). Kolff et al. [2024] (Chapter 4) showed that the maximum of the continuous rating highly correlates to the overall rating. These methods are extended by also considering the mean and median, as well as providing a similar analysis for the maneuver-based ratings. The rating model is then used to make predictions of both rating methods. Third, Bayes' theorem [Jeffreys, 1961] is used to verify whether maneuver-based and overall motion incongruence ratings provided in closed-loop and open-loop driving (first and second phases) are equivalent.

The chapter is structured as follows. The driving and rating tasks are discussed in Section 7.2. The experiment set-up is explained in Section 7.3. Results are presented in Section 7.4, and discussed in Section 7.5. Conclusions are stated in Section 7.6.

7.2 Methods

7.2.1 Driving Task

When driving closed-loop, illustrated in Figure 7.1 including the red elements, the driver controls the steering wheel $\delta_s(t)$, the accelerator $\delta_a(t)$ and brake $\delta_b(t)$ pedals. In a simulation, the vehicle simulation then calculates the corresponding vehicle motion states $\tilde{S}_{veh}(t)$, i. e., the specific forces $f(t)$ and rotational rates $\omega(t)$. As $\tilde{S}_{veh}(t)$ comes from a vehicle model, it is an approximation of the real vehicle motion $S_{veh}(t)$, hence the notation $\tilde{(\cdot)}$. The motion states are sent to the *Motion Control System*, consisting of the MCA and the Motion System (MS). The MCA converts the vehicle motion states to commanded platform motion. These are sent to the MS, i.e., the physical simulator, which determines the actual platform motion $\tilde{S}_{sim}(t)$ [Kolff et al., 2023]. These can differ from the commanded platform motion due to a variety of factors, such as the motion system latency. Differences between the vehicle reference and simulator motion are then the objective mismatches, i.e., $\Delta\tilde{S}(t) = \tilde{S}_{veh}(t) - \tilde{S}_{sim}(t)$.

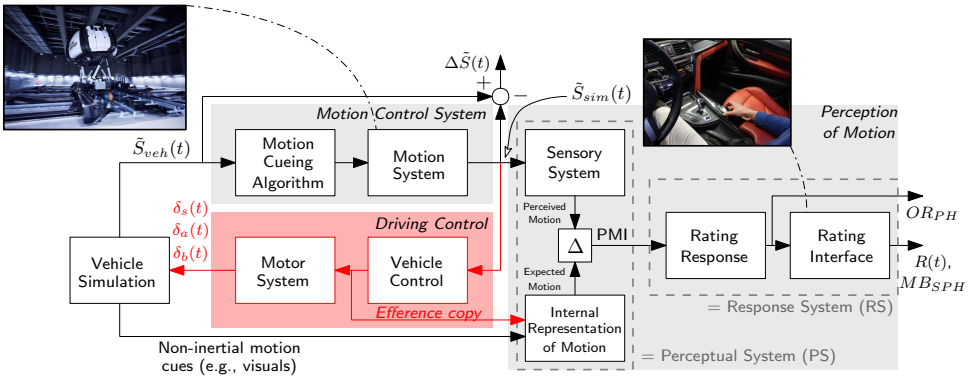


Figure 7.1: Block diagram of the driving and rating tasks. The part "Driving Control" (red) is only present in closed-loop driving. The top left image shows the Sapphire Space simulator (Image: BMW Group), the top right image shows the continuous rating interface, adapted from Ellensohn et al. [2019].

The platform motion is sensed by the driver through their sensory system. Based on the perceived inertial motion and all other non-inertial motion cues in the simulation, such as the visuals [Sivak, 1996], the driver chooses their intended control actions based on a desired state. The motor system of the body produces the actual control actions [$\delta_s(t), \delta_a(t)$ and $\delta_b(t)$], which are sent to the vehicle simulation, closing the driving control loop. In an open-loop driving task (Figure 7.1, excluding the red elements), the driver does not actively control the vehicle and the vehicle simulation is represented by a playback.

7.2.2 Rating Task

Next to the driving task, the participants also performed a rating task. They were tasked with evaluating how well the inertial motion they perceive in the simulator matched to what they would expect to feel from the simulated vehicle. This difference is defined as their PMI [Cleij et al., 2018], see Figure 7.1. As the driver does not exactly know what

the vehicle motion would feel like in a particular situation, they must use an *internal representation* [Stassen et al., 1990] of the vehicle motion based on non-motion cues (e.g., visuals) of the simulation. Individual differences can therefore arise due to familiarity or experience with the simulated vehicle [Kolff et al., 2024b]. Note that thus both the expected and the vehicle model motion $\tilde{S}_{veh}(t)$ can differ from the real vehicle motion $S_{veh}(t)$.

Additionally, in closed-loop control, an *efference copy* [Mulder et al., 2022] of the intended control actions is used to form the expected motion. The expected motion is then not only based on what drivers expect to feel from the non-inertial motion cues, but also what they expect to feel as a result of their own intended control actions. In open-loop driving, the lack of driving control implies that no efference copy can be present. This means that the internal representation can only be generated from the non-inertial motion cues. It is therefore possible that the PMI of open-loop driving is different than in closed-loop driving, for example, because drivers might know less well what motion to expect.

As the PMI is not measurable, a MIR was asked from the drivers. A typical choice is a numeric rating scale [Cleij et al., 2018], e.g., ranging between values of 0 (fully congruent) and 10 (highly incongruent), with steps of 1 [Kolff et al., 2024b], see Chapter 4. Another choice relates to the rating method, i.e., when and how these ratings are acquired. In the experiment, three types of rating methods were used:

7.2.2.1 Overall Ratings (OR_{PH})

After each simulation drive, a single rating value between 0 and 10, representing the overall impression of the drive, was asked from the drivers. As they do not interfere with the driving task(s), overall ratings can be used in closed-loop and open-loop driving. They can be extracted through various methods, such as verbally or through a rating interface. While beneficial due to their non-intrusive nature, the single rating values provide no direct information on which parts of the drive the overall rating is mostly based on. It has been shown that overall rating of PMI correlate with the most incongruent moment in the simulated drive [Cleij et al., 2018; Kolff et al., 2024b].

7.2.2.2 Maneuver-based Ratings (MB_{PH})

Here, a scenario is divided into different maneuvers and drivers give a single rating after those maneuvers [Ellensohn et al., 2020]. Like overall ratings, maneuver-based ratings can be acquired from closed-loop and open-loop driving, and can be given verbally or through a rating interface. A benefit compared to overall ratings is that maneuver-based ratings provided detailed information for each maneuver separately. A downside is that they require that drivers give their rating while driving, which may cause a slight distraction from the driving task.

7.2.2.3 Continuous Ratings ($R(t)$)

Here, drivers rate continuously throughout the drive using a rating interface [Cleij et al., 2018], reflecting their current PMI at each point in time. The main benefit of this method is its high temporal resolution, which allows for modelling approaches [Cleij et al., 2018; Ellensohn et al., 2019c; Kolff et al., 2022; Kolff et al., 2024b]. As it requires operating the rating interface with one hand (see Figure 7.1, top right), it is not possible to drive closed-

loop at the same time, which requires both hands on the steering wheel. Therefore, this task can only be performed in open-loop driving experiments.

7.2.3 Validation of Rating Predictions

As explained in Section 7.1, a main motivation is to predict how a certain motion cueing setting will be rated subjectively in closed-loop driving, based on objective signals of platform motion. For this, a three-step approach is used (Figure 7.2).

7.2.3.1 Rating Model Validation

First, a rating model is used to predict continuous ratings based on objective mismatch signals [path (1) in Figure 7.2]. The latter are defined as the differences in inertial motion (specific forces and rotational rates) between the vehicle motion $\widetilde{S}_{veh,m}(t)$ and the simulator motion $\widetilde{S}_{sim,m}(t)$, i. e., $\Delta\widetilde{S}_m(t)$, with $\widetilde{P}_m(t) = K_{\widetilde{P}_m}|\Delta\widetilde{S}_m(t)|$. Here, m represents the mismatch direction, e.g., $m \in [f_x,\ f_y,\ \ldots,\ \omega_z]$.

In Kolff et al. [2024] (Chapter 4) a linear model was proposed that predicts the continuous rating of the *average* participant. Its structure consists of a first-order low-pass filter transfer function $H_m(j\omega)$ between the absolute mismatch signal $\widetilde{P}_m(t)$ and a modeled rating signal $\tilde{R}(t)$:

$$\hat{\tilde{R}}(j\omega) = \sum_m K_{\widetilde{P}_m}\left(\frac{\omega_c}{j\omega+\omega_c}\right)\hat{\tilde{P}}_m(j\omega), \tag{7.1}$$

with the low-pass filter's cut-off frequency ω_c and the gains of the several mismatch channels $K_{\widetilde{P}_m}$. The $\widehat{(\cdot)}$-terms indicate the Fourier transforms. The low-pass filter represents the participants' lagged response (Response System (RS) in Figure 7.1) to the mismatches $P(t)$. In Kolff et al. [2024] it was shown that the continuous ratings of a CWA MCA condition as measured in that study could be largely explained when considering the longitudinal specific force mismatches $\widetilde{P}_{f_x}$, as well as the yaw rate mismatch $\widetilde{P}_{\omega_z}$ (i.e., $m \in [f_x, \omega_z]$), with the parameters: $\omega_c = 0.37$ rad/s, $K_{f_x} = 0.78$ and $K_{\omega_z} = 6.71$. This model is denoted as "p-CWA".

To express how well the model is able to predict the measured ratings, the VAF is used:

$$\mathrm{VAF} = \left(1 - \frac{\mathrm{var}\left[R(t) - \tilde{R}(t)\right]}{\mathrm{var}\left[R(t)\right]}\right)\cdot 100\%, \tag{7.2}$$

with $R(t)$ and $\tilde{R}(t)$ the measured and modeled rating signal, respectively. The VAF is a measure of how much of the measured signal's variance is explained by the modeled signal [Kolff et al., 2024b]. A value of 100% indicates a perfect fit, whereas it is unbounded on the lower side, i.e., $[-\infty < \mathrm{VAF} \leq 100\%]$.

7.2.3.2 Rating Relationships

In Kolff et al. [2024] it was shown that the maximum of the continuous ratings strongly correlate with the overall ratings [path (2a) in Figure 7.2], such that a linear relationship of the form $OR_{PH} = f_{OR_{PH}}[R(t)] = \alpha_{OR_{PH}} \cdot \max[R(t)] + \beta_{OR_{PH}}$ exists. A similar relationship, between maneuver-based and continuous ratings [path (2b) in Figure 7.2], does currently not exist. In the present work, the mean and median of the continuous ratings will also be considered as possible predictor for the overall ratings *and* the maneuver-based ratings.

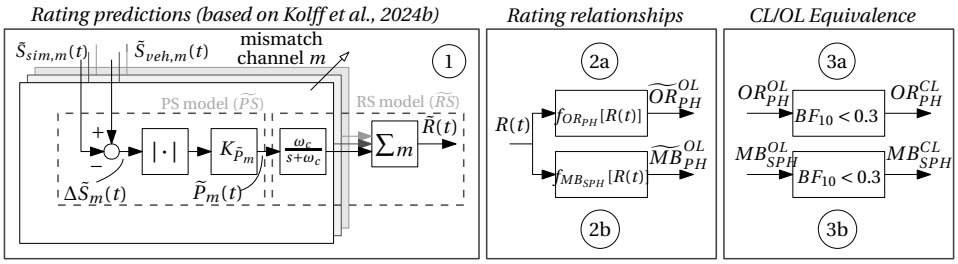


Figure 7.2: Contribution steps, representing (from left to right): Open-loop continuous ratings are predicted from objective mismatches using a rating model (1). Second, rating relationships between the continuous ratings to both the open-loop (2a) overall (OR_{PH}) and (2b) maneuver-based (MB_{PH}) ratings are determined. Third, equivalence testing relates (3a) overall and (3b) maneuver-based ratings of open-loop and closed-loop driving.

7.2.3.3 Equivalence Testing

Finally, to investigate whether OR_{PH} and MB_{PH} ratings of open-loop driving can be used for closed-loop driving, their equivalence is investigated [paths (3a) and (3b) in Figure 7.2, respectively]. In frequentist statistics, data are typically tested for significant differences, i.e., tested for a 95% probability that H_0 (null hypothesis; the data are equivalent) can be rejected in favour of H_1 (alternate hypothesis; the data are different). In the present case, the interest lies not in differences, but in *equivalence*, requiring proof of H_0. This cannot be tested through the same frequentist statistics procedure, as the lack of significant differences does not necessarily imply equivalence. Instead, it only shows that an effect cannot be proven [Lakens, 2017], which can also occur in the case of a lack of statistical power. Thus, using frequentist statistics, the H_0 cannot be accepted. This implies that the frequentist approach is not a suitable method for investigating the equivalence of the open-loop and closed-loop ratings. Specially developed alternative frequentist methods, such as the Two One-Sided Tests (TOST) [Lakens et al., 2018], require normally distributed data. Furthermore, the TOST method is considered to be less reliable for testing equivalence when the sample size is relatively small [Linde et al., 2023].

As an alternative, it is possible to use Bayesian statistics [Jeffreys, 1961], which does allow for explicit testing for equivalence of data. In Bayesian statistics, a degree of belief in a hypothesis is expressed as a form of conditional probability. An estimation of the distribution function is made about the data before even analyzing the data, resulting in a *prior belief*, which holds the ratio of the probability estimates of the hypotheses, i.e, $\frac{P(H_1)}{P(H_0)}$. The prior belief can stem from existing knowledge on the process under investigation, e.g., from previous experiments or from knowledge of underlying physical processes. No explicit assumptions on the distributions of the data, such as normality, are necessary [Jeffreys, 1961]. After the data are observed, the degree of belief is updated [Bolstad and Curran, 2007] to a *posterior* belief. This is expressed as $\frac{P(H_1|D)}{P(H_0|D)}$, with D the observed data (in this case, the maneuver-based ratings of open-loop and closed-loop

driving). The Bayes Factor can then be expressed through:

$$BF_{10} = \frac{P(D|H_1)}{P(D|H_0)} = \underbrace{\left(\frac{P(H_1)}{P(H_0)}\right)^{-1}}_{\substack{\text{Prior}\\ \text{Belief}}} \times \underbrace{\frac{P(H_1|D)}{P(H_0|D)}}_{\substack{\text{Posterior}\\ \text{Belief}}} \tag{7.3}$$

The Bayes Factor, denoted BF_{10}, represents the ratio in proof of H_1 over H_0. Therefore, the factor $BF_{01} = BF_{10}^{-1}$ equals the ratio of proof of H_0 over H_1. A value of $BF_{10} > 1$ indicates that H_1 is more probable [Jeffreys, 1961], but only $BF_{10} > 3$ is considered evidence for H_1. In contrast, $BF_{10} < 1$ means that H_0 is more probable, whereas only $BF_{10} < 0.3$ is considered evidence for H_0 (equivalence). Thus, to prove that the open-loop and closed-loop ratings are equivalent, BF_{10} must be calculated and be shown to be below 0.3. For this analysis, the Bayes factors are calculated using the JASP software [JASP Team, 2023], which calculates values of BF_{incl}. This Bayes factor indicates the change from prior to posterior inclusion odds [Kelter, 2020]. The same range of degrees of belief holds as for BF_{10} [Jeffreys, 1961].

7.3 Experiment Set-up

7.3.1 Experimental Conditions

Using the driving- and rating tasks presented in Section 7.2, the experiment was performed with the following three conditions: i) Closed-loop driving, maneuver-based rating ("CLMB"), ii) Open-loop driving, maneuver-based rating ("OLMB"), and iii) Open-loop driving, continuous rating ("OLCT"). To guarantee that drivers experienced exactly the same motion in the open and closed-loop tasks, the CLMB condition was performed first, such that in the open-loop conditions drivers could be presented with played-back recordings of their own drives. The overall rating was the only rating that was recorded in all three conditions. An overview of the conditions with the applied rating methods is shown in Table 7.1.

Table 7.1: Overview of the experimental conditions.

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

7.3.2 Scenario and Data Acquisition

For increased comparability, the driven route is exactly the same as in Kolff et al. [2024] (Chapter 4), see Figure 7.3. The maneuvers to be rated in the maneuver-based conditions were indicated on the road using green bars and consist of several typical urban driving maneuvers: corners ('CR'), lane changes ('LC'), as well as decelerations ('DEC'). A traffic light was present after 'DEC1', before which drivers had to stop, wait, and accelerate

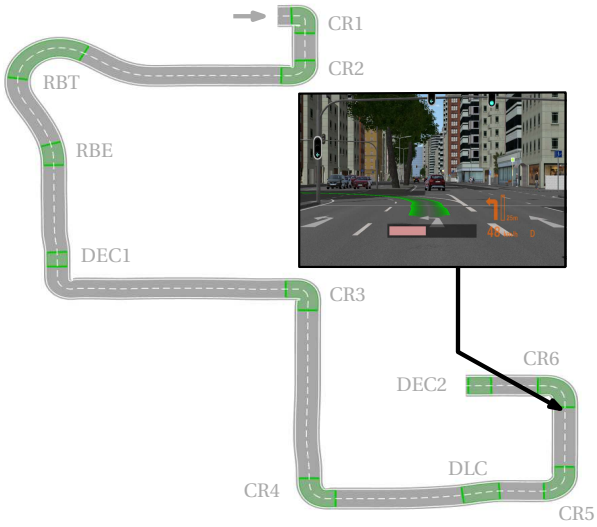


Figure 7.3: Top-down view of the driven route, as in Kolff et al. [2024]. The green areas were visible in the CLMB and OLMB simulations (see screenshot) and represent the maneuvers to be rated: corners (CR), decelerations (DEC), a double lane change (DLC) and a roundabout turn (RBT) and exit (RBE).

again. Compared to Kolff et al. [2024], there are two changes: First, the roundabout is split-up into the roundabout turn ('RBT') and exit ('RBE') to obtain separate maneuver-based ratings for both, increasing the amount of rating information. Second, three lane change maneuvers in Kolff et al. [2024], namely after 'CR2', after 'DEC1', and after 'CR3' were not used in the current experiment, as they were not found to result in informative ratings in Kolff et al. [2024]. Furthermore, this allowed for more time between the various 'CR' maneuvers for drivers to rate. Note that in Kolff et al. [2024] (Chapter 4) the division of the maneuvers was not visible to the participants at all, as in that experiment they only rated the motion continuously. In Kolff et al. [2024] (Chapter 4), the maneuvers were only introduced and shown for clarity to the reader. Therefore, the changes of the maneuvers compared to Kolff et al. [2024] is expected to only minimally impact the results.

7.3.3 Drive Matching Approach

Due to differences in driving style, all recorded closed-loop drives are inherently unique in terms of velocity and lane position. To visualize the differences in the motion that was presented in each drive, a "drive matching approach" was developed. Here, all recorded time signals are related to a common 'reference drive' (see Figure 7.4). Here, the data points of each drive of interest (black points) are linearly interpolated (black lines). For each data point i (red points) in the reference run, a line is constructed perpendicular to the closest linear line piece of the drive of interest, representing the shortest distance between point i and the line piece. The point where these lines intersect (red cross) is used to calculate the ratio $r = n_{k+1}/n_k$. The continuous rating signals are evaluated at these two points and the weighted average based on the ratio r is calculated.

This leads to a vector of indices of equal length for all analyzed drives at which the

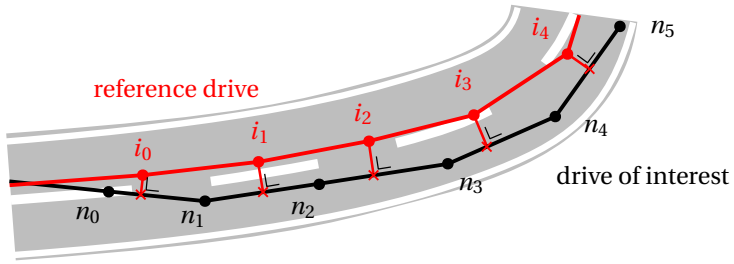


Figure 7.4: Drive matching method, in which for each of the points i of the reference drive (red), the ratio of the linear line segment that connects two points n yielding the shortest distance is calculated.

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

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

7.3.4 Apparatus

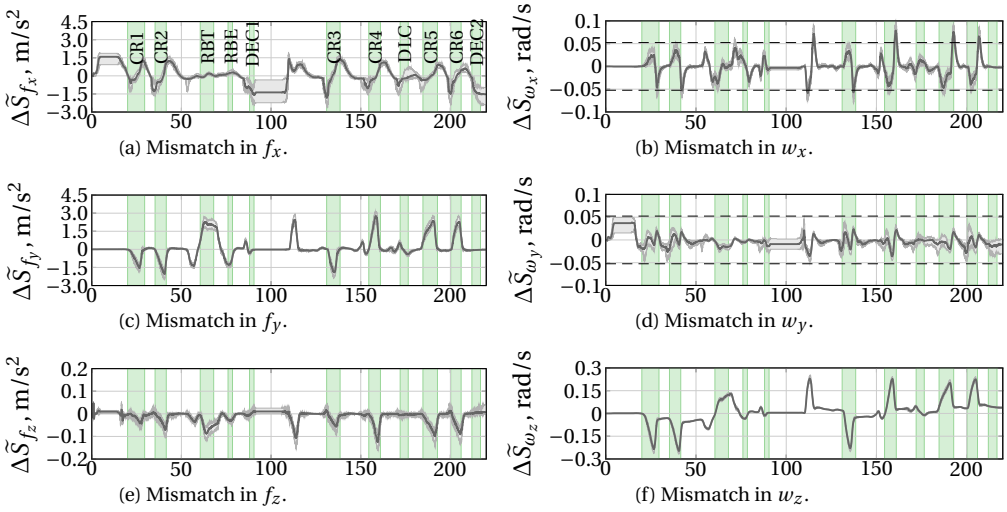
The experiment was performed on the "Sapphire Space" simulator at BMW Group in Munich (see Figure 7.1, top left), a custom designed simulator constructed by Van Halteren Technologies in 2021. Its kinematic structure consists of three motion subsystems: the base is formed by a 19.14 m×15.70 m xy-drive that allows for large excursions in the x and y directions. On top of the xy-drive stands a large 1.15 m stroke hexapod that can move in all six DoFs. Finally, on top of the hexapod, a 360° yaw-drive is installed, allowing for additional yaw rotations of ±180°. The total motion system thus has nine DoFs. The rating model of Kolff et al. [2024] was derived from data collected on the similar, but smaller, "Ruby Space" simulator at BMW Group (xy-drive: 1.6 × 1.5 m, yaw-drive: ±25°, hexapod stroke: 0.34 m).

A one-to-one mock-up of a BMW 3 series (G20) was used, which was fully enclosed by the simulator dome. Visuals were rendered on the inner dome wall using 12 Norxe P1 projectors, resulting in a full 360° projection around the mock-up. During the open-loop drives, the steering wheel remained stationary. The iDrive navigation knob on the center console was used as the rating interface by the drivers to give the continuous ($R(t)$) and maneuver-based ratings (MB_{PH}), see Figure 7.1. The overall rating was extracted verbally for consistency with Kolff et al. [2024]. The 360° projection screen showed the visuals and the current rating value in the form of a "rating bar" [Cleij et al., 2018]. The size

and color of the rating bar changed (See screenshot in Figure 7.3) from rating 0 (short, white) to rating 10 (long, red), to make the rating method more intuitive for drivers to use. The velocity of the vehicle was visible on the tachometer on the dashboard and in the out-of-the-window visuals, together with the driving direction (arrows). The rating knob was connected to the central simulation computer using a CAN bus. This allowed for the accurate and consistent synchronization between recordings of the simulator motion and the rating signals of the participants.

7.3.5 Motion Cueing Algorithm

A CWA was used as the MCA, as its linear filter-based structure ensures a deterministic output. As the motion cueing is calculated in real-time (see Figure 7.1), this is required to ensure that *identical* simulator motion is generated between the closed and open-loop driving conditions. The median mismatch signals of the MCA are shown in Figure 7.5, with the grey areas the interquartile ranges, and with the green areas representing the maneuvers. The CWA tuning did not fully utilize the motion system capabilities, to ensure that the limits were never reached. Tilt-coordination was used and tuned to keep the roll and pitch rate mismatches (Figures 7.5b and 7.5d) below the perceptual threshold of 3 deg/s [Reymond and Kemeny, 2000] (dashed lines). In longitudinal direction, drivers drove more aggressive than expected, resulting in the median pitch rate slightly exceeding its perceptual threshold mismatch (Figure 7.5b).



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

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

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

Figure 7.5: Median mismatches (grey lines) and interquartile ranges (grey shaded areas) over all drives between the reference and simulator as a function of equivalent time in seconds (participant 1, run 1). Green areas are the maneuvers; dashed lines in 7.5b and 7.5d are the perceptual thresholds (±3 deg/s, Reymond and Kemeny, 2000).

The scaling factors used in the MCA were set to 0.5 for the specific forces and 0.6 for the rotational rates. These values lie well within the range of scaling factors considered to be the most realistic, i.e., 0.4 – 0.8 found by Berthoz et al. [2013]. First-order filters were used to distribute the low- and high-frequency motion across the motion subsys-

tems (i.e., the xy-drive, the hexapod, and the yaw-drive). The break frequencies were set to 30 rad/s for the translational axes, such that motion below that frequency was reproduced by the xy-drive, whereas high-frequency accelerations were reproduced by the hexapod. A higher value of 50 rad/s was used for the yaw motion, such that the majority of yaw motion was reproduced by the yaw-drive, giving the hexapod more workspace to reproduce the roll and pitch motion. Finally, the lowest-frequency specific force motions in x and y directions were reproduced by the hexapod tilt-coordination, implemented using a low-pass filter break frequency of 0.5 rad/s.

Because all drivers drove themselves, the MCA output of each closed-loop drive is different. It is the longitudinal specific force mismatch ($\Delta\tilde{S}_{f_x}$, Figure 7.5a) that shows the largest spread, larger than $\Delta\tilde{S}_{f_y}$ (lateral specific force mismatch, Figure 7.5c) and $\Delta\tilde{S}_{\omega_z}$ (yaw rate mismatch, Figure 7.5f). This can be explained by the more varying nature of the driving behaviour in the longitudinal direction (i. e., braking and accelerating at different points in time) [Eppink et al., 2023], whereas the lateral and yaw mismatches are mostly determined by the road shape [Bosetti et al., 2014] and result in more similar experiences across all drives. Furthermore, although there were only two distinct braking maneuvers in the maneuver-based conditions (DEC1 and DEC2), this does not mean that there was no longitudinal maneuvering present in the other maneuvers. In fact, as is visible in Figure 7.5a, the longitudinal specific force mismatch was also present during corner maneuvers, where participants braked into and accelerated out of the corner. Thus, the ratings of these corner maneuvers should also partially consist of a response to the longitudinal specific force mismatch.

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

7.3.6 Participants and Procedures

The experiment was performed by forty-two participants to ensure a large enough sample size [Cai and Wang, 2024] and to account for possible dropouts due to simulator sickness and/or technical problems. All participants were BMW employees and had a European car driver's license B for at least five years (M = 14.5 years, SD = 9.1 years) and an average yearly driven distance of M = 16,278 km (SD = 16,408 km). The average age was M = 32.9 years (SD = 9.4 years). Thirty drivers had previous experience in driving simulators. All drivers provided informed consent and the experiment was approved following BMW's internal ethics review procedures. Due to drop-outs (technical issues or simulator sickness), 36 complete data sets were obtained. The incomplete data sets of the drop-outs are not considered in further analysis.

All experiment sessions were ran by a single experimenter to ensure consistency in the interaction with the participants. All participants completed the experiment in a single session. Before entering the simulator, they all read a written briefing that explained the three rating methods, as well as the rating scale. All drivers performed one training

drive for each of the three conditions, to get accustomed to the simulator, the sensation of motion, and the rating methods. The training drive of the OLCT condition contained inverted longitudinal (f_x) motion, to create large false cues and anchor the highest value of the rating scale (10), as in [Kolff et al., 2024b]. The OLCT and OLMB training drives were not based on the participant's own CLMB training drive, but used a pre-recorded drive, such that the anchoring of the rating scale was identical for all participants.

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

For the maneuver-based ratings, drivers were asked to give their impression of the maneuvers (Figure 7.5) using the rating knob. Drivers were instructed to rotate towards their intended rating, leave the rating at this value for at least two seconds, and then rotate back to zero. The selected maneuvers were spaced to give drivers enough time to give their rating and refocus on the driving task.

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

7

7.4 Results

7.4.1 Modeling of Continuous Ratings

Figure 7.6 shows the measured median continuous ratings (blue) over all drives. The individual rating data are included in Appendix D.3. Given that the rating scale runs from 0 (congruent motion) to 10 (highly incongruent motion), ratings are generally low (< 2), i.e., the MCA setting was rated well. The rating peaks generally coincide with the end of the maneuver (vertical line), showing the lagged response to the incongruences, as expected from the estimated rating dynamics represented in the rating model in Eq. 7.1. The figure also shows the p-CWA model of Kolff et al. [2024] (Chapter 4 in grey, which predicts the peaks of the continuous ratings quite well, although the quality of the fit is low at VAF= 11.0%. This VAF excludes the initial acceleration (between 0 and 20 s) and final deceleration 'DEC2' sections, as these were generally ignored by most participants; it might have been unclear to them that these were also to be rated. A second point of interest lies at the plateau between 90-110 s which is, as explained in Subsection 7.3.5, caused by the drive matching approach. For further model calculations, the modeled ratings of the rating model are calculated for each drive separately, such that this plateau is not present (but cannot be compared in a single figure).

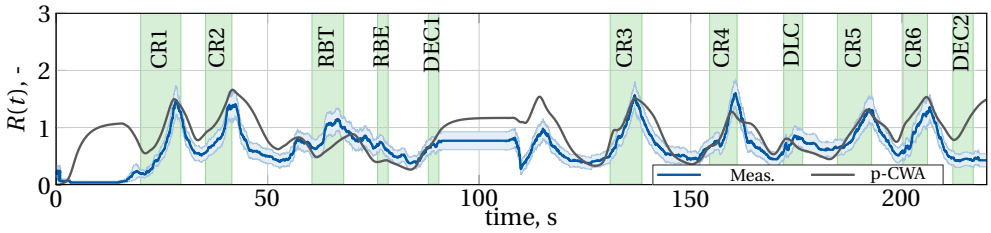


Figure 7.6: Mean continuous motion incongruence rating (blue line) and standard error (blue shaded area) over all OLCT drives as a function of equivalent time in seconds (participant 1, run 1). The predicted rating based on the p-CWA model of Kolff et al. [2024] (Chapter 4) is displayed by the grey line. Green areas are the maneuvers, although these were not highlighted to the participants in this OLCT condition.

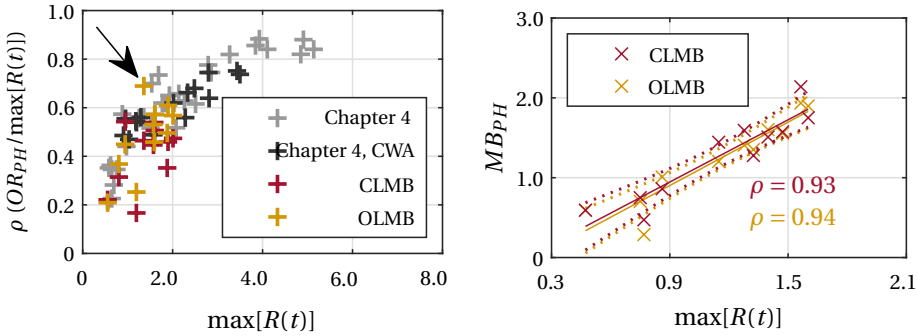
7.4.2 Rating Relationships

Relationship Overall/Continuous Ratings

A relationship linking the continuous rating to the overall ($OR_{PH} = f_{OR_{PH}}[R(t)]$) [path (2a) in Figure 7.2] is investigated. For the overall ratings, these fits are determined for each maneuver separately. Thus, each maneuver has a regression coefficient as to how much it correlates to the overall rating. As this requires a single data point for the continuous ratings in each maneuver, the continuous ratings are summarized through three methods: i) the maximum of $R(t)$ of the maneuver (CLMB: $\rho = 0.46$, OLMB: $\rho = 0.69$, ii) the mean (CLMB: $\rho = 0.46$, OLMB: $\rho = 0.65$), and iii) the median (CLMB: $\rho = 0.40$, OLMB: $\rho = 0.44$). A value closer to 1 indicates a stronger linear relationship, such that the maximum best explains the relationship between the rating methods. These values correspond to the roundabout ('RBT'), for which the correlation was always highest. Figure 7.7a shows how well the overall rating correlates to the maximum rating of each maneuver, expressed as the maximum continuous rating of that maneuver. Here, the grey values show the correlation values as determined by Kolff et al. [2024], the dark grey indicates such data points that correspond to a CWA condition. The red (CLMB) and orange (OLMB) points indicate the present study. To obtain an explicit predictive relationship, the regression fit with the highest Pearson correlation ($\rho = 0.69$, indicated by the arrow in Figure 7.7a) in the OLMB condition is taken: $OR_{PH} = 0.79 \cdot \max[R(t)] + 1.63$. This is similar to the relationship determined in Kolff et al. [2024] (Chapter 4), where the relationship $OR_{PH} = 0.8 \cdot \max[R(t)] + 2.0$ was found.

Relationship Maneuver-Based/Continuous Ratings

To investigate the relationship between continuous and maneuver-based ratings ($MB_{PH} = f_{MB_{PH}}[R(t)]$), [path (2b) in Figure 7.2], also the Pearson correlation is calculated. Here, the applied method differs from the overall rating. As a single data point exists in each maneuver for both the continuous and maneuver-based rating methods, a single regression fit can be made on all data points of the various maneuvers together. The continuous ratings are again summarized through three methods: i) the maximum of $R(t)$ in that maneuver (CLMB: $\rho = 0.93$, OLMB: $\rho = 0.94$, ii) the mean (CLMB: $\rho = 0.80$, OLMB: $\rho = 0.76$), and iii) the median (CLMB: $\rho = 0.59$, OLMB: $\rho = 0.49$). Similar



(a) Overall ratings

(b) Maneuver-based ratings

Figure 7.7: Correlations between the maximum of the continuous ratings and the overall ratings (a) / maneuver-based ratings (b). The data points represent each maneuver as defined in Figure 7.5. In (a), the arrow indicates the maneuver ('RBT') with the highest correlation. In (b), the lines are the regression fits, the dotted lines represent the 95% confidence bounds.

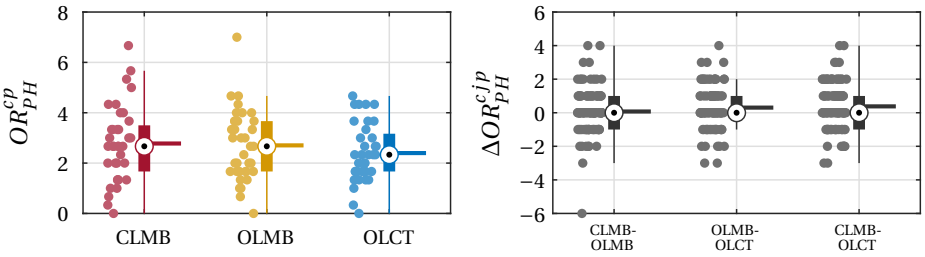
to the overall ratings, it is the maximum of the continuous rating in the maneuver with the highest Pearson correlation (Figure 7.7b) that is the best predictor for the maneuver-based ratings: $MB_{PH} = 1.32 \cdot \max[R(t)] - 0.29$.

7.4.3 Equivalence of CL/OL ratings

Overall Ratings

The overall rating distributions are shown in Figure 7.8a. The OLCT is also shown for reference, as the overall ratings were recorded in all three conditions. The box plots show the median (circles), the box edges indicate the 25th and 75th percentiles, and the whiskers show the range of the non-outlier data points. All individual data points are plotted as dots. The horizontal bars represent the means of the distributions. The data are normally distributed; the means for the CLMB, OLMB, and OLCT conditions are 2.78, 2.70, and 2.40, respectively, showing that the OLCT condition was rated slightly lower. The Bayes factor of the single effect between the CLMB and OLMB conditions [path (3a) in Figure 7.2] is $BF_{incl} = 0.263$ (Table 7.2 under 'OR_{PH}'), indicating moderate evidence of equivalence (< 0.3) [Jeffreys, 1961]. Note that when including the overall ratings obtained in the OLCT condition, the Bayes factor increases to $BF_{incl} = 0.419$, providing no more evidence of equivalence. However, as the prime focus here is on the comparison between CLMB and OLMB, this does not affect any further conclusions on equivalence between closed-loop and open-loop driving.

Even though the CLMB and OLMB overall ratings are equivalent, individual differences can still be present. For example, two drivers could have rated the two conditions differently, but in exactly opposite ways. Although this would lead to equivalent data, it would ignore insights into individual differences. Figure 7.8b shows the distributions of ΔOR_{PH}^{cjp}, i.e., the difference in overall ratings per condition pair of each individual run pair. With the presence of the OCLT condition, this results in three ΔOR_{PH}^{cjp} distributions. As this only has one distribution per condition pair, no statistical test is possible. The



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

Figure 7.8: Overall ratings of the three tested conditions.

Table 7.2: Bayes factors of the driving method, maneuver, and their interaction, for the overall (OR_{PH}) and maneuver-based (MB_{PH}) ratings. Bold values indicate evidence of equivalence (< 0.3).

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

horizontal bars indicate the means: 0.074 for CLMB - OLMB (median = 0), showing that individuals rated the CLMB and OLMB conditions the same. Furthermore, for OLMB - OLCT the mean is 0.31 (median = 0), and for CLMB - OLCT: 0.38 (median = 0). These mean values show that OLCT was also rated slightly lower *within* individuals.

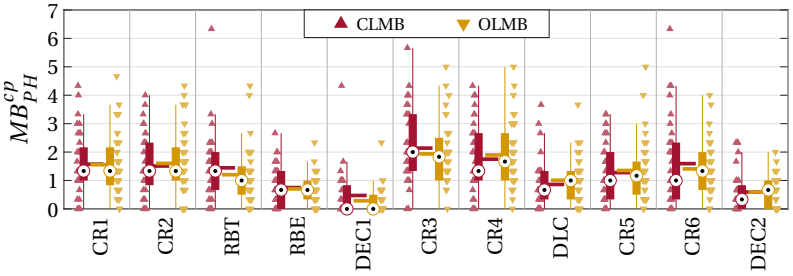
7

Maneuver-based Ratings

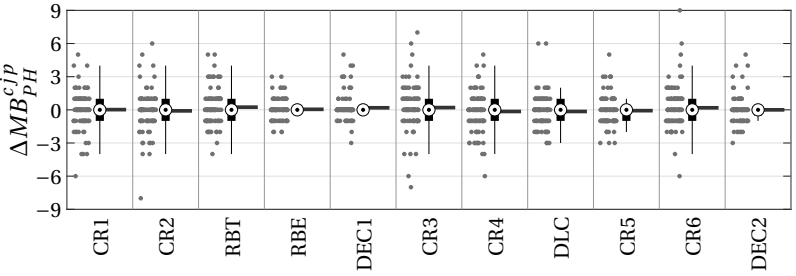
Equivalence of the maneuver-based ratings is investigated next [path (3b) in Figure 7.2]. Figure 7.9a shows the distributions of the maneuver-based ratings of the CLMB and OLMB conditions for each maneuver. Differences exist between the maneuvers, with CR3 the worst rated maneuver (i.e., the highest means). The corners (involving lateral and yaw motion) are generally rated the worst (e.g., CR3 and CR4), whereas maneuvers involving longitudinal motion (DEC1 and DEC2) are rated best.

To investigate the equivalence of the two conditions, the Bayes factors are calculated. The results are shown in Table 7.2 under 'MB_{PH}'. Three possible effects are analyzed for the MB_{PH} data: 'CLMB/OLMB', 'Maneuver' and 'CLMB/OLMB×Maneuver', where the latter represents the interaction effect. For the maneuver effect, the BF_{incl} is $1.018 \cdot 10^{14}$, indicating extremely decisive evidence (> 30) [Jeffreys, 1961] that the maneuvers were rated differently.

In contrast, when considering CLMB/OLMB, $BF_{incl} = 0.143$, providing moderate to strong evidence [Jeffreys, 1961] that the maneuver-based ratings of the two conditions are equivalent, supporting the earlier findings on equivalence of the overall ratings. For the combination of the two effects, no interaction effect exists between the CLMB/OLMB



(a) Distributions of the maneuver-based ratings.



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

Figure 7.9: Maneuver-based ratings for CLMB and OLMB conditions per maneuver.

and the maneuvers ($BF_{incl} = 0.023$). This indicates that the equivalence within the driving method does not depend on the (type of) maneuver. Thus, although the maneuvers are rated differently, these differences are *equivalent* in the CLMB and OLMB conditions.

Similar to the analysis on the overall ratings, Figure 7.9b shows the distributions of ΔMB_{PH}^{cjp}, i.e., the difference of each individual run pair. All medians are 0, and the means are generally very close to 0 (highest $\Delta MB_{PH}^{cjp} = 0.24$, for 'RBT'). This provides further evidence that the drivers rated both conditions equivalently.

7.4.4 Rating Prediction Framework Evaluation

The three steps defined in Figure 7.2 have now been evaluated. First, due to their equivalence, maneuver-based ratings of open-loop drives can be used to predict ratings of closed-loop drives (see red and orange data in Figure 7.10, representing their means). Second, using the estimated regression fits that relate the overall and maneuver-based ratings to the (measured) continuous ratings, both the overall and maneuver-based ratings can be predicted (blue). This holds for both ratings of open-loop and closed-loop driving due to their equivalence. Third, the continuous ratings can be predicted using a rating model, based on objective mismatch signals (grey). Therefore, the steps combined allow for predicting maneuver-based and overall ratings of closed-loop drives using a continuous rating model. Between the predicted maneuver-based ratings of the rating model and the measured closed-loop ratings, the deviations are smaller than half a rating point. Considering a ten-point rating scale, where only steps of 1 were possible,

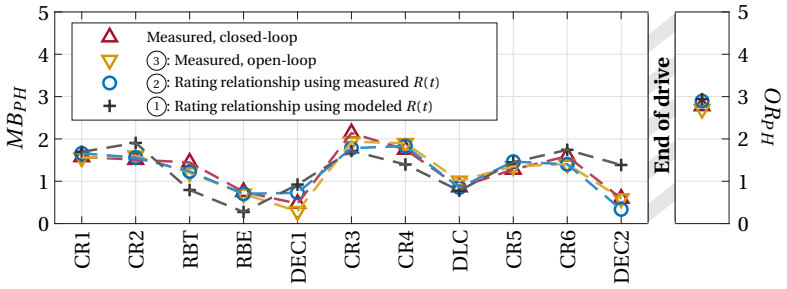


Figure 7.10: Steps of determining maneuver-based (left) and overall (right) ratings.

these errors can be considered acceptable. Exceptions are 'RBE' and 'DEC2', where the differences are 0.65 and 0.79, respectively. For the overall post-hoc ratings, the rating predictions also work well, with a difference between the measured closed-loop (red) and the modeled (grey) ratings of 0.16.

7.5 Discussion

7.5.1 Model Predictions

The model proposed by Chapter 4 was used to predict the measured continuous ratings as a function of the objective mismatch between vehicle reference and simulation motion. Using the same model parameters of Kolff et al. [2024] resulted in a reasonably accurate prediction of the ratings. Although the VAF was low, the resulting predictions of the maneuver-based and overall ratings were accurate. Between the present work and the model of Chapter 4, the scenario, the rating set-up, instructions, and the MCA were the same. However, the simulator, the MCA parameters, and the participant group were different, which may have affected the ratings. Overall, the presented results show that the model still provides accurate results across these variables when the averages of these participant groups are considered. For these three variables there is thus no combined effect. This shows that the rating prediction methodology is effective even across these experiment variables and can thus be applied for predictions of motion cueing quality of future, as of yet untested, driving simulator studies. This highlights the validty of the model as a *predictive model* between experiments as well.

Especially the difference in simulator is notable. The nine DoF Sapphire Space simulator used in the present study was significantly larger than the nine DoF Ruby Space simulator, on which the model of Kolff et al. [2024] was developed. The smaller workspace of the simulator in Kolff et al. [2024] resulted in larger mismatches compared to the present experiment, on which the rating model was fit. This thus also shows that the model still works when considering different ranges of incongruent motion. This is an important quality, as it shows the general applicability of the model across various simulators, which is an important property considering the various simulators to choose from at BMW's Driving Simulation Center. For smaller, hexapod-only systems, the rating model, including its ability to predict ratings of closed-loop driving, can be further tested by applying it to predict the motion cueing quality of practical driving experiments. For ex-

ample, the effectiveness of the rating model can be confirmed by comparing predictions and actually obtained overall post-hoc ratings, as the latter can be obtained with limited interference in the experiment itself.

Predicting ratings for significantly smaller simulators (e.g., small hexapods) might in fact prove to be the largest future difficulty, especially for systems that cannot approximate the high cueing quality of the Sapphire Space and the Ruby Space. As the applied rating scale has a fixed lower anchoring ('no incongruence = 0'), but no upper anchoring ('large incongruence = 10'), it is possible that the rating that drivers associate with 'large incongruence' can depend on the intensity of the incongruences presented in the experiment. Cleij [2020] showed that in some cases, transferability between experiments can be an issue if the difference in presented motion is too large (e.g., between a low- and high-fidelity simulator). She introduced a *MTP* to apply a linear scaling between the acquired ratings, normalized for the presented motion in each experiment. A further investigation on transferability between experiments is therefore suggested, in which larger differences between the motion are present, as this would need to be corrected for in the rating model.

Another crucial direction for future work is to investigate the validity of the findings and the rating model prediction in different driving scenarios. Here, the first step would be to test a different urban route, as this might alter the balance of the presented mismatches and therefore require the introduction of an MTP. Second, extending the results to completely different scenario types (e.g., highway or rural) is another important step. As discussed in Kolff et al. [2024], different scenario's can, for example, induce more interaction with surrounding traffic, which may induce different types of motion (e.g., more lane changes in highway scenarios). Here, maneuvers may be more difficult to rate, as anticipating responses to traffic is more difficult than the road-geometry driven maneuvers of an urban scenario.

7.5.2 Relationships between Rating Signals

To understand how the overall (OR_{PH}) and maneuver-based (MB_{PH}) ratings relate to the continuous ratings ($R(t)$), it was determined which metrics best correlate. Analyzing the correlation between the maximum of the continuous ratings per maneuver and the overall ratings, it can be concluded that the higher the maximum of the continuous ratings, the more these ratings correlate with the overall ratings. This reproduces findings by Kolff et al. [2024]. The point with the highest correlation ($\rho = 0.69$) was the roundabout maneuver 'RBT'. The analysis between the maneuver-based and continuous ratings shows a similar result: the maneuver-based ratings are highly correlated with the maximum of the continuous ratings in that maneuver ($\rho = 0.94$). These results show that maneuver-based and overall ratings can be predicted using continuous ratings.

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

cueing. Similarly, it is suggested to extend the findings on correlations towards other scenarios. As the urban driving scenario generally results in reliable rating data [Kolff et al., 2024b], less strong relationships might be present in scenarios that inherently have a lower reliability, such as rural [Ellensohn et al., 2020] or highway scenarios.

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

7.5.3 Equivalence of Closed- and Open-loop Ratings

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

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

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

Even though the ratings were equivalent, note that the underlying perception does not necessarily have to be equivalent as well. It has been shown [Nesti et al., 2016; Valente Pais et al., 2012] that perceptual thresholds can in fact change under closed-loop and open-loop single-axis settings, hinting at differences in perception. However, even if these perceptual differences would be present in multi-axis car driving simulations, the equivalent ratings show that these differences are small enough to not be of practical significance.

A point of attention lies within the fact that the participants had to rate their own drives, also in the open-loop conditions. This was a crucial choice, as it allowed for the explicit comparison between open-loop and closed-loop driving. Although participants were not told that they would rate their own drives in the open-loop conditions, a potential bias could occur when participants recognize their own drives: then their rating could be affected by their memory of what the motion felt like in the closed-loop con-

dition. To mitigate this, the order in which the three drives were presented in the open-loop conditions was different than that of the closed-loop drives. Furthermore, while the vehicle's trajectory was replicated directly, the traffic in the simulation was still random every time.

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

7.6 Conclusions

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

8

Incongruences and Simulator Sickness

Key findings

- Continuously obtained sickness ratings support that the occurrence and severity of simulator sickness and can be traced back to mismatches in simulator motion of specific maneuvers.
- In urban driving simulations, mismatches in the reproduction of the lateral specific force induce the highest subjective motion incongruence ratings, followed by the longitudinal specific force, and then the yaw rate, reproducing the order found in Chapter 4. The order is longitudinal specific force, yaw rate, and then lateral specific force for simulator sickness.
- False cues induce the most simulator sickness and highest subjective motion incongruence ratings, followed by missing cues, and then scaled motion. For false and missing cue motion these ratings are disproportionally worse than what the linear mismatch models of Chapter 4 predict.

This chapter is based on the following publication:

Title:	Effect of Motion Mismatches on Ratings of Motion Incongruence and Simulator Sickness in Urban Driving Simulations.
Authors:	M. Kolff, C. Himmels, J. Venrooij, A. Parduzi, D.M. Pool, A. Riener, and M. Mulder.
Journal:	Transportation Research Part F: Traffic Psychology and Behaviour, accepted.

Abstract

This chapter investigates the effects of motion mismatches on simulator sickness and subjective ratings of the motion. In an open-loop driving simulator experiment, participants were driven through a recorded urban drive twelve times, in which mismatches were induced by manipulating the following three aspects in motion cueing: (i) mismatches in specific vehicle axes, (ii) mismatch types (scaling, missing, and false cues), and (iii) inconsistent scaling between different motion axes. Subjects (N=52) reported simulator sickness post-hoc (after each drive), as well as continuously during each drive, a first in simulator sickness research. Furthermore, subjective post-hoc motion incongruence ratings on the quality of the motion were extracted. Results show that longitudinal motion mismatches lead to the most simulator sickness and the highest ratings, followed by mismatches in lateral motion, then yaw rate. False cues induce the most sickness, followed by missing and then scaled motion. Inconsistent scaling between the axes has no significant effect. The continuous sickness ratings support that the occurrence and severity of simulator sickness are indeed related to mismatches in simulator motion of specific maneuvers. This chapter contributes to an improved understanding of the relationship between motion mismatches and simulator sickness, allowing for more targeted motion cueing strategies to prevent and reduce sickness in driving simulators. These strategies may include the appropriate selection of the simulator, the motion cueing, and the sample of participants, following the presented results.

8.1 Introduction

SIMULATOR SICKNESS poses a crucial issue in driving simulator when conducting user studies [Caird and Horrey, 2011; de Winter et al., 2012]. Although often considered a form of motion sickness [Johnson, 2007], simulator sickness can also occur without actual physical movement, e.g., in static driving simulation. Symptoms of simulator sickness include dizziness, headache, sweating, stomach awareness, and even nausea [Reason and Brand, 1975], making it an unpleasant experience. Apart from leading to a decreased sense of presence [Almallah et al., 2021], simulator sickness can cause subjects dropping out of experiments, resulting in incomplete data sets, requiring larger samples and increasing the costs of studies. It also negatively affects the validity of simulator studies, as subjects may behave systematically different when experiencing symptoms of simulator sickness [Cobb et al., 1999; Igoshina et al., 2022]. For these reasons, simulator sickness is generally to be avoided and driving simulations that are expected to induce a high degree of sickness are ideally identified beforehand.

The occurrence of simulator sickness in driving simulators has been shown to depend on a variety of experiment variables, such as the participant age [Keshavarz et al., 2018], as well as the driving scenario. For example, both Klüver et al. [2015] and Mourant and Thattacherry [2000] found simulator sickness to be more severe when simulating driving on rural roads compared to highway driving. Similar results were obtained by Himmels et al. [2022], who additionally found that urban scenarios induce even more sickness than rural scenarios. Real-life urban driving itself may already cause a high degree of *motion sickness* [Irmak et al., 2021; Salter et al., 2019; Turner and Griffin, 2000]. An additional issue for urban *simulations* is that the dynamic maneuvers involved, such as sharp corners and frequent acceleration/deceleration, come with strong visual stimuli, while the corresponding physical motion is difficult to reproduce on the motion platform. This can induce mismatches between expected and perceived motion. If a driver notices these mismatches, the motion is defined as *incongruent* [Cleij et al., 2018].

Subjective ratings on motion incongruences are valuable information to obtain, as they can be used to identify the most critical mismatches in the simulator motion and through mitigating these mismatches the simulation realism can be increased. Although it does not necessarily occur to all drivers, one of the main theories in motion sickness research is that simulator sickness is the result of sensory conflicts [Bos, 2011; Reason, 1978], and can arise in the presence of incongruent motion. If so, the question arising is then *which* mismatches induce incongruences and simulator sickness the most, such that these can be systematically avoided. With knowledge on both phenomena, studies inducing motion incongruences and/or a high degree of simulator sickness could then be identified before-hand, such that the most appropriate simulator, MCA, and subject sample size can be selected.

In literature, apart from investigating the effect of scenario choice, most investigations focus on the benefits of particular motion (sub)systems, as these designs often have the specific potential to reduce mismatches in a certain simulator DoF. For small, 3-DoF systems, Parduzi [2021] found no difference in simulator sickness compared to a static simulator. Zöller [2015] even found an *increase* in sickness in a 3-DoF driving simulator, which was attributed to the limited motion space. More recent work showed beneficial effects of using larger hexapod motion system with Y-drives (although small)

[Klüver et al., 2015] or even XY-drives [Himmels et al., 2022b]. Although these studies provide insight in general qualities of various motion systems, the surplus in available motion space does not inherently lead to smaller mismatches. For example, simulators combined with redundant rail systems, such as the BMW Group's Sapphire Space simulator, offer large potentials to drive urban scenarios with larger motion scaling factors. However, they also have the potential to do *more wrong*, as false cues that might be negligible on a small system, also become enlarged using higher scaling factors. Clearly, the top priority of any motion cueing system must be to minimize all cueing errors. However, one could argue that the better the motion cueing can potentially be, it may be more important to focus on mitigating what the simulator could be doing wrong instead of further optimizing what it is doing right.

Crucial for the prediction of simulator sickness would be to have a more refined understanding of *how* the mismatches between expected and simulator motion actually contribute to the development of sickness symptoms. General scenario descriptions such as "rural", "highway", or "urban" are not very informative about the tested route. To transfer (often expensively) gained knowledge to scenarios that were not tested before, one must be able to attribute the emergence of simulator sickness to much more specific properties *during* the scenario, such as mismatches in inertial motion while driving through curves, accelerating/decelerating, or along roundabouts. Current sickness models lack any distinction between the different types of errors (e.g., scaled, missing, and false cues) that may occur in such maneuvers: all different types of errors are treated as equally sickening [Irmak et al., 2023]. Furthermore, these models do not include the relative importance of the vehicle axes, such as longitudinal, lateral, and yaw motion.

Recent work [Kolff et al., 2024b] (Chapter 4) found that in the evaluation of the Perceived Motion Incongruence (PMI) (i.e., a driver's opinion on the (in)congruence of the simulator motion), mostly the accurate reproduction of the lateral and longitudinal specific force channels are important (in this order), with an additional, minor, role of the yaw rate. If simulator sickness is caused by inertial motion mismatches, this might suggest a similar division of the vehicle motion channels on simulator sickness. Currently, most research deals with the topic at hand by acquiring sickness evaluations after a drive ("post-hoc") [Himmels et al., 2022b; Mourant et al., 2007]. Such evaluations are not detailed enough to link to specific mismatches during the drive. Continuous sickness ratings have been used previously to overcome these issues [Irmak et al., 2021; Qiu et al., 2023]. However, a systematic investigation on the importance of mismatches in different vehicle axes as contributors to simulator sickness does not yet exist.

This chapter aims at improving the understanding and the predictive capabilities regarding simulator sickness and motion incongruence ratings based on objective mismatches of the simulator motion. In a dedicated driving simulator experiment, 52 participants were driven as passengers ("open-loop") through an urban simulation on BMW's largest driving simulator "Sapphire Space". Twelve variations of motion cueing were tested, inducing scaled, missing, and false cue motion in the vehicle's three main axes: longitudinal (f_x), lateral (f_y), and yaw (ω_z). Subjects provided a MISC rating [Reuten et al., 2021] on their level of simulator sickness, as well as a subjective rating on the (in)congruence of the motion after each drive ("post-hoc"). Subjects also identified their current level of simulator sickness continuously *during* each drive using a rating knob,

resulting in a continuous measurement, similar to continuous ratings of motion incongruences [Cleij et al., 2018] but a first in simulator sickness research. This allows relating sickness increments to specific events in the simulation, such as different driving maneuvers. An investigation of the importance of (1) the three main axes, (2) the criticality of scaling, missing, and false cue motion, and (3) the importance of axis inconsistency for simulator sickness is performed. The main contribution of this chapter lies in providing a better understanding of how motion mismatches cause simulator sickness and motion incongruence to increase.

The chapter is structured as follows. Section 8.2 introduces the considered methods. Section 8.3 describes the performed experiment. Its results are then described in Section 8.4. This is followed by a discussion in Section 8.5 and conclusions in Section 8.6.

8.2 Methods

8.2.1 Perceived Motion Incongruence

During the simulation, the vehicle model produces the reference motion (specific forces and rotational rates) that is to be reproduced by the simulator. This motion, described by $\tilde{S}_{veh}(t)$, as shown in Figure 8.1, is fed to the *Motion Control System*. This block consists of two parts, the Motion Cueing Algorithm (MCA) and the Motion System. The latter is the physical component of the simulator that produces the actual inertial motion. It is driven by the MCA, which, using knowledge of the configuration and dimensions of the motion system, converts the vehicle motion to simulator motion that can be reproduced by the motion system. Therefore, using the input signals $\tilde{S}_{veh}(t)$, the *Motion Control System* gives the output $\tilde{S}_{sim}(t)$. Differences between the vehicle reference and simulator motion signals are then the objective mismatches, i.e., $\Delta\tilde{S}(t) = \tilde{S}_{veh}(t) - \tilde{S}_{sim}(t)$.

The blocks indicated by *Perception of Motion* then describe how humans perceive the simulator motion. They can notice a difference (Δ) between the physical motion they perceive in the simulator and the motion they would expect to feel from the simulated vehicle. This is known as the PMI [Cleij et al., 2018; Kolff et al., 2024b], see Figure 8.1. As the driver does not exactly know what the vehicle motion would feel like in a particular situation (i.e., the signal S_{veh} is unavailable to them), they must use their *internal representation* [Stassen et al., 1990] of the vehicle motion based on previous experience and the non-motion cues (e.g., visuals) provided in the simulation. Individual variations can arise due to different familiarity or experience with the simulated vehicle [Kolff et al., 2024b] (Chapter 4). Note that thus both the expected motion and the vehicle model motion $\tilde{S}_{veh}(t)$ are approximated versions of the real vehicle's motion S_{veh}.

Although ideally the PMI would be measured directly, this is not possible, as it is "internal" to the driver. Instead, a Motion Incongruence Rating (MIR) was asked from the drivers [Cleij et al., 2018]. This is a subjective rating that represents the PMI. The PMI and MIR can differ, as the latter can be affected by the *rating response* (Figure 8.1). This can be affected by a variety of factors concerning the experiment [Kolff et al., 2024b] (Chapter 4), such as the participant's understanding of the rating method, as well as the (participants response to the) type of rating method used. In the present case, the rating aims to summarize the PMI across the whole drive. Such *overall ratings*, denoted OR_{PH} (see Figure 8.1), are known to correlate well with the worst considered maneuver during

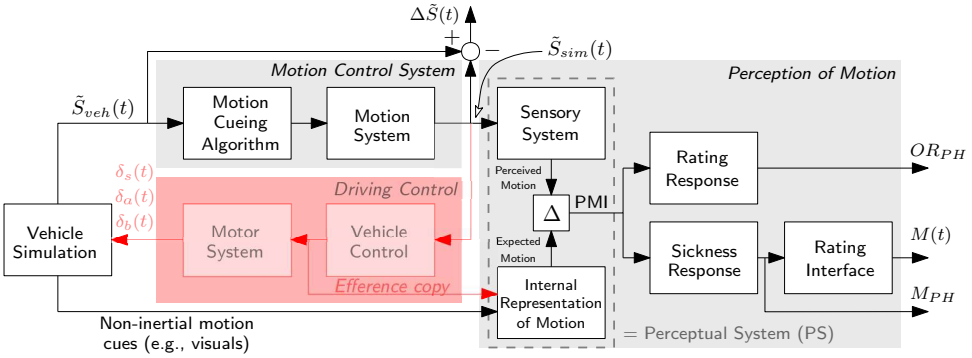


Figure 8.1: Block diagram of the driving and rating tasks, adapted from Kolff et al. [2024] (Chapter 7). The part "Driving Control" (red) is only present in closed-loop driving, and not active in the present experiment.

a drive [Cleij et al., 2018; Kolff et al., 2024a; Kolff et al., 2024b]. The subscript $_{PH}$ denotes that the rating was acquired at the end of each drive, i.e., *post-hoc*. In the current experiment, the rating values were on a semantic differential scale, ranging between values of 0 (fully congruent) and 10 (highly incongruent), with steps of 1, to remain close to the work of Kolff et al. [2024] and Kolff et al. [2024] (Chapter 7).

Additionally, in closed-loop control (when driving the car oneself), the red path in Figure 8.1 would be active, denoted by *Driving Control*. In closed-loop control, an *efference copy* [Kolff et al., 2024a; Mulder et al., 2022] of the intended control actions is used to form the expected motion. The expected motion is then not only based on what drivers expect to feel from the non-inertial motion cues, but also what they expect to feel as a result of their own intended control actions of the steering wheel, accelerator pedal, and braking pedal deflections ($\delta_s(t)$, $\delta_a(t)$, and $\delta_b(t)$ respectively). In "open-loop" driving (being driven as passengers, as in the present experiment), the lack of driving control implies that no efference copy is present. This means that the internal representation of motion, and in turn the expected motion, is only fed by the non-inertial motion cues. Kolff et al. [2024] (Chapter 7) has shown that closed-loop and open-loop driving are rated equivalently in an urban scenario.

8

8.2.2 MIR Rating model

A rating model is used to predict the ratings based on objective mismatch signals. The latter are defined as the differences in inertial motion (specific forces and rotational rates) between the vehicle motion $\widetilde{S}_{veh,m}(t)$ and the simulator motion $\widetilde{S}_{sim,m}(t)$, i. e., $\Delta\widetilde{S}_m(t)$, with $\widetilde{P}_m(t) = K_{\widetilde{P}_m}|\Delta\widetilde{S}_m(t)|$. Here, m represents the mismatch type (specific force or rotational rate) and direction, e.g., $m \in [f_x,\ f_y,\ \ldots,\ \omega_z]$.

In Kolff et al. [2024] (Chapter 4) a linear model was proposed that predicts the continuous rating of the *average* participant. Its structure consists of a first-order low-pass filter transfer function $H_m(j\omega)$ between the absolute mismatch signal $\widetilde{P}_m(t)$ and a modeled rating signal $\tilde{R}(t)$:

$$\hat{\tilde{R}}(j\omega) = \left(\frac{\omega_c}{j\omega+\omega_c}\right)\sum_m K_{\widetilde{P}_m}\hat{\widetilde{P}}_m(j\omega), \tag{8.1}$$

with the low-pass filter's cut-off frequency ω_c and the gains of the several mismatch channels $K_{\tilde{P}_m}$. The $\widehat{(\cdot)}$-terms indicate the Fourier transforms. The low-pass filter represents the participants' lagged response to the mismatches. In Kolff et al. [2024] (Chapter 4) it was shown that the continuous ratings of an Oracle MCA condition as measured in that study could be largely explained when considering the longitudinal specific force mismatch $\tilde{P}_{f_x}$, lateral specific force mismatch $\tilde{P}_{f_y}$, as well as the yaw rate mismatch $\tilde{P}_{\omega_z}$ (i.e., $m \in [f_x,\ f_y,\ \omega_z]$), with the parameters: $\omega_c = 0.52$ rad/s, $K_{f_x} = 0.62$, $K_{f_y} = 1.11$, and $K_{\omega_z} = 1.08$. This model, denoted "p-ORC" was fitted on rating data of the same Oracle algorithm as in the present chapter. It is used instead of the more general "p-ALL", which was fitted on a combination of Oracle and CWA rating data, as that model does not incorporate a yaw rate mismatch component.

Kolff et al. [2024] (Chapter 4) also investigated the relationship between continuous and post-hoc rating signals, from which it was found that the latter is well described by the most incongruent point in the simulation (i.e., the point with the highest continuous rating). Furthermore, the explicit rating relationship $OR = 2.0 + 0.8 \cdot \max[R(t)]$ was found. Thus, using simulations of the continuous rating signals, a prediction of the overall post-hoc can be made.

8.2.3 Simulator Sickness

The second quantity evaluated by the subjects was their own perceived level of simulator sickness. One of the main theories in motion sickness research considers sensory-expectancy conflict as the prime cause for sickness [Bos, 2011; Reason, 1978]. Here, the conflict refers to a mismatch between the sensed sensory signals and the expected sensory signals [Irmak et al., 2023]. Considering the similarity with PMI here, it is likely that in the context of (driving) simulation, simulator sickness might *result* from PMI. However, simulator sickness is a phenomenon that does not necessarily occur to everyone. Not only can individuals have different susceptibility to motion and/or simulator sickness (some get sick easily, some never), expectation of the motion has also been shown to affect motion sickness [Kuiper et al., 2020]. Therefore, individuals with a lot of driving experience might get sick earlier than those without, as they might notice the motion to be incongruent while their less-experienced colleagues may not. The "Sickness Response" block in Figure 8.1 represents this individual response.

For measuring the participants' level of simulator sickness, the Motion Illness Symptoms Classification (MISC) [Bos et al., 2005; Wertheim et al., 2001] is often used. This 11-point symptom-based questionnaire ranges from 0 (no sickness) to 10 (vomiting), see Table 8.1. Its intermediate values represent different levels of motion sickness, including dizziness, headache, sweat, or stomach awareness [Reason and Brand, 1975]. It relies on the fact that stronger sickness symptoms (nausea, vomiting) are usually preceded by lighter symptoms (headache, sweating, etc.) [Reason and Brand, 1975]. Reuten et al. [2021] found that there is a particular order in which sickness symptoms develop, confirming the progressiveness of the symptom-based MISC. MISC ratings were further found to be highly correlated to unpleasantness ratings [de Winkel et al., 2022; Reuten et al., 2021]. The MISC is a more refined and single-item alternative to, for example, the sixteen-item Simulator Sickness Questionnaire (SSQ) [Kennedy et al., 1993], which is too extensive for real-time use. The MISC is easy to understand and can be used with little

training [Bos et al., 2010].

Table 8.1: The Motion Illness Symptoms Classification (MISC), adapted from Bos et al. [2005].

Symptom		**Score**
No problems		0
Slight discomfort but no specific symptoms		1
Dizziness, warm, headache, stomach awareness, sweating, etc.	vague	2
	some	3
	medium	4
	severe	5
Nausea	some	6
	medium	7
	severe	8
Vomiting		10

The MISC-value is usually given verbally after each drive, i.e., post-hoc, as will also be done here. This results in values of M_{PH}, as shown in Figure 8.1. Another possibility is to ask for MISC values during a drive after a given interval, like every minute [Diels et al., 2023; Hogerbrug et al., 2020] or every 30 or 40 seconds [Irmak et al., 2022; Irmak et al., 2021]. A benefit of the latter approach is the higher temporal resolution, as sickness symptoms might build up and/or disappear *during* a drive, which cannot be measured using post-hoc ratings. A drawback, however, is that repeatedly asking subjects to rate their well-being, reminds them to think about how they feel, which may lead to higher MISC values. Furthermore, even a 30-seconds resolution might be insufficient to relate specific maneuvers and sickness, especially in very dynamic urban scenarios.

In the present experiment a novel approach to obtaining MISC ratings during the drive was tested. Here, participants could provide the MISC rating at *any* time during the drive, through a rating interface (See Figure 8.1), and were instructed to *only* change their rating when they noticed a change in their well-being. As a novelty in simulator sickness research, but similar to its use in general motion sickness research [Qiu et al., 2023], the resulting 'real-time' rating $M(t)$, yields a continuous rating of simulator sickness, rather than measuring on fixed intervals. Through obtaining the MISC as soon as it changes, it should become more clear which maneuvers or parts of a drive affect simulator sickness. While this approach potentially reduces the impact of measurement interference mentioned above, a potential risk is that, in due time, some participants forget to change the rating.

8.3 Experiment Set-up

8.3.1 Sample

In total, 52 subjects participated in the experiment (43 men, 9 women). All participants were in the possession of a European car driver's license B. The average age was M = 41 yrs (SD = 8 yrs). 43 participants had experience with simulator driving. The experiment was terminated if a participant reached a post-hoc MISC value of six in two succes-

sive drives (*some nausea*), or a value of seven or higher in a single drive (*medium nausea*; for an explanation of MISC ratings, see Table 8.1. This was the case for four participants, all males. In two additional cases the experiment could not be completed due to technical issues. A total of 46 (52 – 6) complete data sets were obtained. The six incomplete data sets were excluded from analysis.

All participants were employees of the BMW Group and participated in the study on a voluntary basis during their regular working hours. Participants were not specifically selected for their own estimated sensitivity to simulator sickness, to obtain a representative group of participants. They were informed about the purpose of the study. The participation could be ended at any time on the participants' initiative without consequences. The experiment was approved following BMW's internal ethics review procedures. All participants provided informed consent.

8.3.2 Scenario

A single recorded drive through an urban environment was used, see Figure 8.2a. This scenario was based on the driven route used in Kolff et al. [2024] (Chapter 4), although a new recording was used and the route was shortened to last 133.8 s. It consisted of several maneuvers typical for urban driving: accelerations ('ACC#'), decelerations ('DEC#'), three 90° corners ('CR#'), and a roundabout ('RBT'). The highest occurring speed in the recording was 50 km/h. Traffic was not simulated, i.e., there were no other road users.

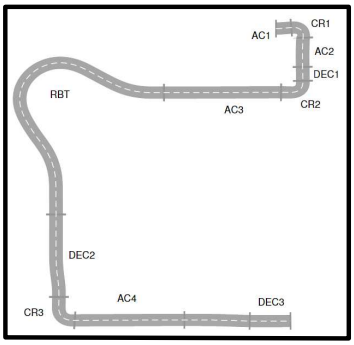
8.3.3 Apparatus

The experiment was performed on the BMW Group's Sapphire Space simulator, see Figure 8.2b. The motion system has nine DoFs, consisting of three subsystems: a 19.14 m × 15.70 m XY-drive, a 1.15 m stroke hexapod and a 360° yaw-drive. The simulator is fully enclosed by a dome, in which a BMW X5 series (G05) vehicle mockup was placed. The steering wheel rotated corresponding to the simulated drive.

The iDrive navigation knob on the center console (Figure 8.2c) was used as the rating interface by the drivers to provide the continuous MISC rating (denoted as $M(t)$). Inside the dome, visuals were rendered using Unreal Engine and displayed using 12 Norxe P1 projectors, resulting in a full 360° projection in the dome. The current MISC rating value in the form of a 'rating bar' was visible in the central field of view in a type of head-up display. The size and color of this rating bar changed (see screenshot in Figure 8.2d) from rating 0 (short, white) to rating 10 (long, red), to make the rating method more intuitive for drivers to use. The current numerical MISC value of the participants was also displayed on the rating bar. The full MISC including verbal anchors was furthermore displayed on the vehicle Central Information Display (CID) throughout the drives. The velocity of the vehicle was visible on the tachometer on the dashboard and in a head-up display alike screen projection, together with the driving direction (arrows).

8.3.4 Independent Variables

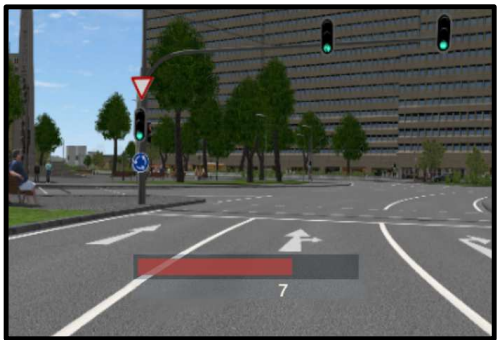
The experiment tested twelve variations of motion cueing using the optimization-based MCA described in Ellensohn et al. [2019]. Here, the simulator motion is optimized beforehand by minimizing the difference between the reference drive and simulator motion along the whole drive. It is referred to as *Oracle* motion cueing due to its knowledge



(a) Top-down view of the driven route, containing corners (CR#), accelerations (AC#) decelerations (DC#), and a roundabout (RBT#).

(b) The Sapphire Space simulator (Image: BMW Group)

(c) The iDrive rating knob, used for the extraction of the continuous Motion Illness Symptoms Classification (MISC) values.



(d) The Motion Illness Symptoms Classification (MISC) value given by the participants as visible in the simulated scene, consisting of a coloured bar and the numeric value (currently set to 7).

Figure 8.2: The experiment set-up.

of all future states (only possible when using a pre-recorded drive) and represents the best-possible motion cueing for a given simulator and recorded drive [Kolff et al., 2022]. The combination of Oracle motion cueing and the Sapphire Space's large and dynamic workspace allows for motion cueing with high scaling factors [Kolff et al., 2022]. For the considered urban scenario, it was iteratively found that values of 0.8 were the highest possible scaling factors in all directions that still fit in the workspace of the simulator. A higher scaling factor would have resulted in the simulator reaching its limits. The corresponding motion was used as the baseline condition 'A'. The other eleven experimental conditions (labeled 'B' to 'L') were variations of the baseline condition 'A', but then containing scaled, missing, and false cue motion in the three main vehicle directions (longitudinal, lateral, and yaw).

To investigate the dependency on these specific vehicle axes, the specific forces and

rotational rates that act on the simulator cabin must be manipulated. These are the outputs of the Oracle algorithm. Because of the presence of the simulator's yaw-drive, a simulator motion in a certain direction in its inertial frame does not necessarily equate to a motion in the same direction as acting on the simulator cabin. For example, a longitudinal simulator motion under a 90° yaw-drive angle will result in a lateral motion acting on the simulator cabin [Kolff et al., 2023]. Thus, to manipulate the forces acting on the cabin, it is not possible to turn off a specific simulator axis. Rather, the non-linear simulator motion must be optimized using Oracle to produce the manipulated motion as its output. Therefore, Oracle was run twelve times separately, once for each condition, using scaling factors on the vehicle reference data (0.4 for scaled, 0 for missing, and −0.8 for false cue motion) in the relevant directions, see Table 8.2. This thus allows for investigating the effect of scaled, missing, and false cue motion in the three main vehicle DoFs. Note that this also includes the overall scaling factor of 0.8. Although Oracle would also be able to scale down the motion itself as a result of its optimization, this 'pre-scaling' of the vehicle reference input was used to ensure that the motion channels that were *not* manipulated in a specific condition were always scaled down using the same factor. Otherwise, depending on the condition, Oracle could have either less or more workspace available in the other directions, resulting in different scaling factors between the conditions in the non-manipulated axes. The results of these optimizations are shown in Figure 8.3. In two conditions ('E' and 'I'), the motion was manipulated in all three axes, such that the axis manipulations are consistent with each other. The scaling factors for the baseline motion (0.8) and scaled motion (0.4) were chosen as they correspond to the upper and lower scaling factor values considered to be acceptable by participants in driving simulation [Berthoz et al., 2013].

Note that only the scaling factors of the f_x, f_y, and ω_z motion channels were varied. The vertical (f_z) channel, as well as the rotational roll (ω_x) and pitch (ω_y) rate channels were not necessarily zero, even in the no motion condition 'I'; their magnitudes are very small (grey lines in 8.3e-8.3d). This resulted in a fairer comparison with the conditions in which the motion was dominant and active. Considering the terrain flatness in the scenario and the corresponding small roll and pitch rotational rates, this is likely to be a negligible effect. In all conditions, Oracle applied tilt-coordination, in which a rotation of the simulator cabin is used to generate a sustained specific force through the gravity vector [Stratulat et al., 2011]. For this to be also perceived as a pure sustained acceleration by the participants, the associated rotational motion must be below the perceptual threshold. Therefore, the Oracle was constrained to keep the rotational rates ω_x and ω_y below <3 deg/s [Reymond and Kemeny, 2000]. As a rotation around the z-axis does not result in a sustained specific force [Kolff et al., 2023] and cannot be used for tilt-coordination, the ω_z channel was not constrained.

8.3.5 Procedures

The participants first completed a pre-questionnaire including demographic questions as well as questions related to simulator sickness experience. The latter was done using the Motion Sickness History Questionnaire (MSHQ) [Griffin and Howarth, 2000]. Participants were then led to the simulator and given a safety briefing. Each subject experienced all twelve cueing variants, resulting in twelve rides of 133.8 s each (total driving

Table 8.2: Scaling factors of the vehicle data in each direction for each condition. Bold values highlight the manipulated motion channels.

	Baseline	Scaling				Missing				False		
Axis	A	B	C	D	E	F	G	H	I	J	K	L
long. (f_x)	0.8	**0.4**	0.8	0.8	**0.4**	**0.0**	0.8	0.8	**0.0**	**−0.8**	0.8	0.8
lat. (f_y)	0.8	0.8	**0.4**	0.8	**0.4**	0.8	**0.0**	0.8	**0.0**	0.8	**−0.8**	0.8
yaw (ω_z)	0.8	0.8	0.8	**0.4**	**0.4**	0.8	0.8	**0.0**	**0.0**	0.8	0.8	**−0.8**

time: about 27 minutes per subject). The order of the conditions was counter-balanced between participants using the Latin-square method to minimize order effects between the conditions. With 46 participants and 12 conditions, the Latin square was unfinished.

During the drives, the participants' task was to continuously evaluate their MISC-level using the rating knob. After each drive (*post-hoc*), participants verbally answered a questionnaire on their MISC level at that moment, as well as their subjective rating on the overall quality of the motion (MIR). The questionnaire further included a question related to the perceived match of motion and visuals, a question on perceived realism, and a question on the sense of presence. Any question wordings and scales, including the MISC, were displayed continuously on the central display inside the vehicle for reference. Only MISC and MIR are further evaluated in the present chapter.

After the sixth drive, a five-minute break was held, in which participants were also allowed to step out of the simulator. This break was included to allow for some relief for the participants' eyes and from the rating tasks, as well as to reduce the level of simulator sickness to make sure that more participants could finish the whole experiment. Due to this break, it is possible that the carry-over effect of simulator sickness that is present between all conditions is not present between the conditions directly before and after the break, thus affecting the simulator sickness results. However, due to the different order in which the conditions were presented to the participants, the condition after which this break occurred differed per participant, which should thus cancel out over all participants.

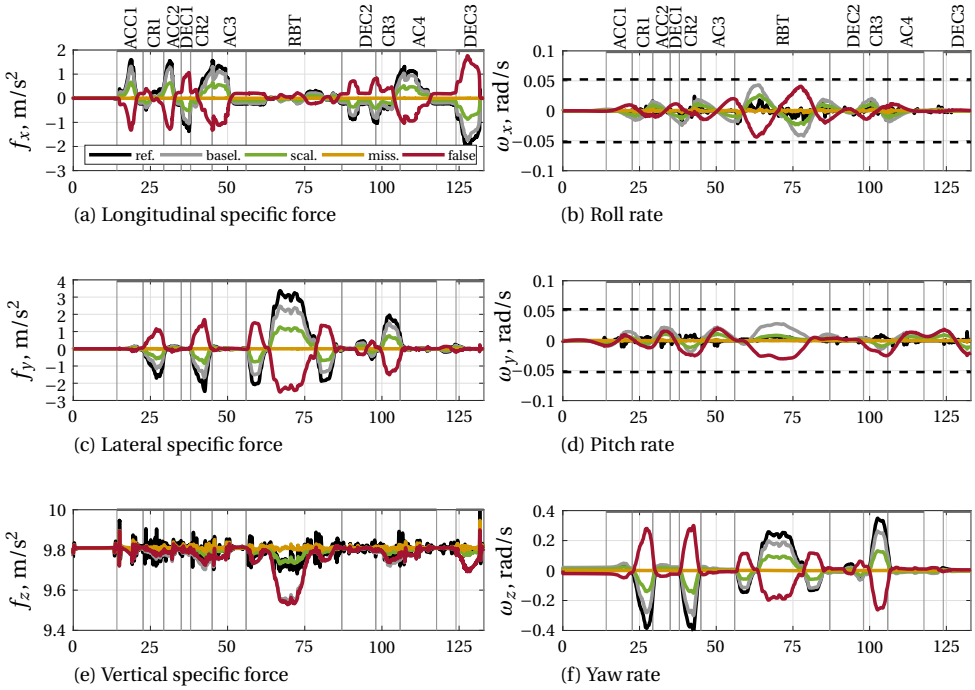
8.3.6 Hypotheses

The two dependent measures in the experiment, MIR and MISC, are clearly related. However, their exact relation is unknown and likely to be nonlinear, as high MIR ratings may not necessarily lead to high MISC ratings (i.e., motion can be mismatched, but not sickening). Furthermore, some participants can be expected to be quite capable of indicating incongruent motion while for them this incongruent motion does not lead to simulator sickness.

The following hypotheses are tested with regard to motion incongruences (MIR):

H1: False cues (conditions J-K-L) receive the highest MIR values, followed by missing cues (conditions F-G-H-I), and then the scaled motion cues (conditions B-C-D-E).

H2: Mismatches in the f_y (lateral), f_x (longitudinal), and ω_z (yaw) channels lead to an increase in MIR, in this order of severity *(confirmation of Kolff et al. [2024] (Chapter 4))*.



(a) Longitudinal specific force (b) Roll rate

(c) Lateral specific force (d) Pitch rate

(e) Vertical specific force (f) Yaw rate

Figure 8.3: Motion mismatch profiles of the various conditions. In this figure the black lines represent the vehicle data, grey lines the baseline simulator output (0.8), green lines the scaled output (0.4), yellow lines the missing output (0.0), and red lines the false direction output (−0.8).

The following hypotheses are tested with regard to simulator sickness (MISC):

H3: False cues (conditions J-K-L) receive the highest MISC values, followed by missing cues (conditions F-G-H-I), and then the scaled motion cues (conditions B-C-D-E).

H4: Mismatches in the f_y (lateral), f_x (longitudinal), and ω_z (yaw) channels lead to an increase in MISC, in this order of severity.

8.4 Results

The experiment had an incomplete factorial design with two varied factors: the 'axis' in which a mismatch was induced, and the 'error type' corresponding to that mismatch. The levels of one factor were not always present in all levels of another factor (e.g., condition 'A' does not have a manipulation in 'axis'). This makes the use of a two-way Analysis of Variance (ANOVA) for statistical analysis of the data infeasible. While in the experimental design, post-hoc MISC and MIR were considered depending on the applied cueing condition, the main interest lies in the effects of 'error type' and 'axis' rather than in the effect of 'cueing condition'. Cueing condition, however, can be considered nested within the factors error type and axis, applying a hierarchic ANOVA.

Table 8.3: Effects considered in the model. The fixed effect is nested within the levels of the random effects.

	effect type	levels
cueing condition	fixed	A-L (see Table 8.2)
axis inconsistency	random	yes, no
axis	random	f_x, f_y, ω_z
error type	random	scaling, missing, false

Table 8.4: Cueing condition as nested in 'axis inconsistency', 'axis', and 'error type'. Empty cells (·) indicate that no level was assigned to the cueing condition, resulting in a different number of observations depending on the random effects included in each tested model.

condition	axis inconsistency	axis	error type
A (0.8 0.8 0.8)	no	·	·
B (**0.4** 0.8 0.8)	yes	Long.	scaling
C (0.8 **0.4** 0.8)	yes	Lat.	scaling
D (0.8 0.8 **0.4**)	yes	Yaw	scaling
E (**0.4 0.4 0.4**)	no	·	scaling
F (**0.0** 0.8 0.8)	yes	Long.	missing
G (0.8 **0.0** 0.8)	yes	Lat.	missing
H (0.8 0.8 **0.0**)	yes	Yaw	missing
I (**0.0 0.0 0.0**)	no	·	missing
J (−**0.8** 0.8 0.8)	yes	Long.	false
K (0.8 −**0.8** 0.8)	yes	Lat.	false
L (0.8 0.8 −**0.8**)	yes	Yaw	false

Although not part of the hypotheses, the variable 'axis inconsistency' was further considered, to exploratively investigate whether there are effects of manipulating one axis in isolation in contrast to manipulating all axes at the same time. Therefore, 'axis inconsistency', 'axis', and 'error type' are considered as (higher level) random effects in a linear model, in which the 'cueing condition' (considered as fixed effect) is nested, see Table 8.3. The levels of the random effects were assigned to the cueing condition levels as described in Table 8.4.

8.4.1 Post-hoc MIR Ratings

The post-hoc overall MIR ratings of the twelve conditions are illustrated in Figure 8.4. For these ratings, linear models were built with several fixed and random predictors: 'condition' (fixed) nested within the random factors 'axis inconsistency', 'axis', and 'error type' (Table 8.3). Then, it was determined which of these predictors explain a relevant share of variance in post-hoc MIR (see Table 8.5).

Here, for the predictors 'subject' and 'condition', the Intraclass Correlation Coefficient (ICC) and R^2 values are displayed. Residual maximum likelihood was applied as method for mixed models, as this is more suitable than maximum likelihood when interested in the relevance of random effects. Note that significance tests are not reported, as the likelihood ratio test's p values are known to be conservative [Pinheiro and Bates, 2000]. Further, step-wise testing has been criticized in the past as a misuse of hypothesis testing [Whittingham et al., 2006]. Instead, the data are analyzed based on Akaike's Infor-

mation Criterium (AIC). Here, the measure of fit is calculated based on deviance, while penalizing complexity [Bolker et al., 2009]. While the AIC avoids step-wise procedures, it may still over fit, and thus suffer from the same issue like the likelihood ratio test's p. Differences in AIC values of 2 or more are considered relevant [Burnham et al., 1998].

The random effects were first added to the model independently, to see whether they can be considered relevant at all [Field et al., 2012]. This step shows only a small decrease in AIC when including 'axis inconsistency' and will therefore not be considered further. A better model fit is obtained when including 'axis' and 'error type' as random factors. Especially the 'axis' provides a large decrease in AIC and is thus the most critical predictor to include in the model. Including both axis and error type (model 5 in Table 8.5) results in the lowest AIC value and is therefore the best model. It performs better than model 1 (fixed-effect only) and the models including only one random effect (models 3 and 4), confirming that there are indeed effects of axis and error type on the MIR, see Figure 8.4a.

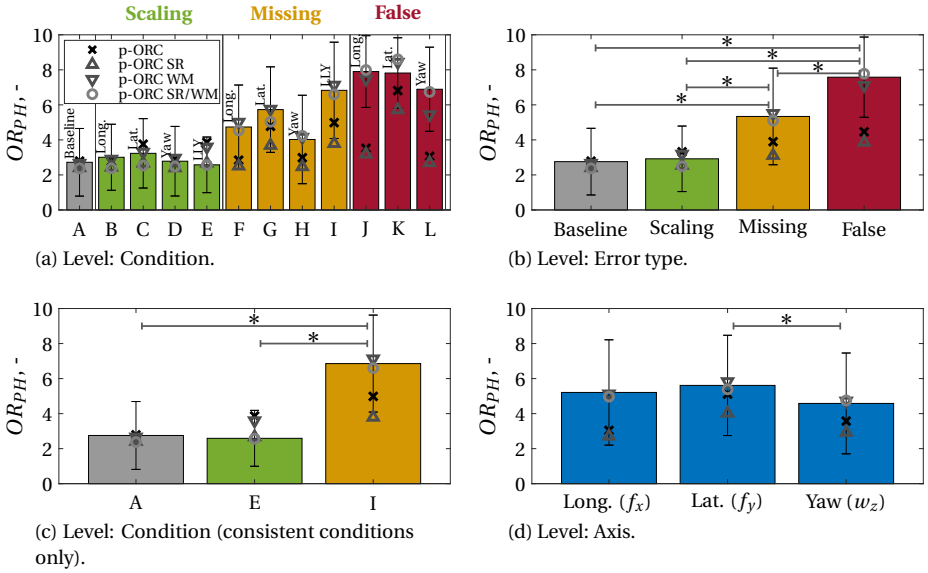
Table 8.5: Model performance comparison with varying predictors, dependent measure: post-hoc Motion Incongruence Rating (MIR). The values of Intraclass Correlation Coefficient (ICC) and R^2 cannot be provided for the random effects models. Model 5 (bold) performs best in terms of the Akaike's Information Criterium (AIC).

model #	predictors	ICC (adj.)	ICC (unadj.)	R^2 (conditional)	R^2 (marginal)	AIC
0	subject	.08	.08	.08	.00	2976.93
1	condition; subject	.24	.13	.59	.45	2575.28
2	condition, subject, axis inconsistency					2574.31
3	condition, subject, error type					2566.34
4	condition, subject, axis					1934.49
5	**condition, subject, error type, axis**					**1925.96**

Figure 8.4b shows the comparison of the error type (levels tested: baseline, scaling, missing, false). The baseline cueing condition A was included as a fourth level. The MIRs were lowest for baseline and scaling, followed by missing and then false cues in ascending order, confirming Hypothesis H1. The effects were subjected to Bonferroni-corrected pairwise post-hoc tests to determine which factor levels of the random effects actually differ. These tests indicated no difference between baseline (A) and scaling, but significant differences across all other conditions. As the number of observations and conditions with each level considered in the post-hoc test vary, the consistent cueing conditions A (0.8 0.8 0.8), E (0.4 0.4 0.4), and I (0.0 0.0 0.0) were also considered separately, see Figure 8.4c. This comparison confirms that there was no significant difference between the baseline (A) and scaled motion (E), but both of them being advantageous over missing motion (I). As there was no condition with full, consistent false cue motion (-0.8 -0.8 -0.8), a similar comparison for false cue motion was not possible.

Finally, Figure 8.4d shows the MIRs grouped per axis. The lateral specific force manipulations result in the highest MIR (i.e., worst ratings), followed by the longitudinal specific force and then the yaw rotational rate. This order is the same as described in Kolff et al. [2024] (Chapter 4) and supports Hypothesis H2. However, the post-hoc tests show a significant difference only between the lateral and the yaw axis. This indicates that a mismatch in the yaw rate (ω_z) affects the MIRs less negatively than a mismatch in

the lateral specific force (f_y) does. Thus, decreasing the lateral specific force mismatches are more important for reducing MIRs than the yaw rate mismatches. There were no significant differences across the mismatches of f_x and f_y or f_x and ω_z. Thus, Hypothesis H2 can only be partially confirmed with statistical significance.



(a) Level: Condition.

(b) Level: Error type.

(c) Level: Condition (consistent conditions only).

(d) Level: Axis.

Figure 8.4: Post-hoc Motion Incongruence Rating (MIR) values and model predictions for the considered levels. Green, yellow, and red indicate scaling, missing, and false cue errors, respectively. Grey (condition 'A') is the baseline. For each condition in Figure 8.4a, the manipulated Degree of Freedom (DoF) is indicated (longitudinal, lateral, or yaw). LLY indicates manipulations in longitudinal, lateral, and yaw directions. The star symbols indicate significant differences.

8

8.4.2 MIR Rating Predictions

The p-ORC model of Kolff et al. [2024] (Chapter 4) is now applied to check its validity under the various motion manipulations. Figure 8.4a shows the model (crosses), predicting the overall ratings per condition. The model overestimates ratings of the scaling conditions and underestimates those of the missing and false cue conditions. Furthermore, more clearly visible in Figure 8.4d, a different balance in the three axes is present than what the rating model predicts. Based on the presented findings of the experiments' measured MIRs, two improvements to the rating model can be proposed.

First, considering the finding that there is no significant difference between the baseline (scaling of 0.8) and scaling motion (scaling of 0.4), which is in line with the preferred scaling range of 0.4 – 0.8 found by Berthoz et al. [2013], an adaptation to the input of the rating model can be made. Instead of taking the full vehicle reference in the calculation of $\tilde{S}_m(t)$ in (8.1), a reference scaling of $0.6\tilde{S}_m(t)$ is applied. This results in the new model "p-ORC SR", with SR indicating "Scaled Reference". This model is indeed better able to predict the scaled conditions (as indicated by the triangles, also visible in Figures 8.4c and 8.4b), but underpredicts the error type and axis levels.

The second improvement applied to the model was therefore to consider additional gains on the missing and false cue conditions, i.e., K_{mc} and K_{fc}. The rationale is that humans might rate these types of motion disproportionally worse than that the linear rating model predicts [Cleij, 2020]. These gains were only applied on the mismatch signals in which the missing or false cue motion was active. The gains of the three axes are also adjusted, i.e., K_{f_x}, K_{f_y}, and K_{ω_z} to find updated values compared to Kolff et al. [2024] (Chapter 4). Together, the gains are optimized to minimize the average quadratic difference between the measuring and predicted post-hoc overall ratings, i.e., $\sum(OR_{PH} - \widetilde{OR}_{PH})^2 / N_{cond}$, with $N_{cond} = 12$ the number of conditions. The optimization was performed using a Nelder-Mead algorithm in Matlab. As this procedure does not guarantee finding a global optimum [Kolff et al., 2024b], the procedure was repeated 50 times with randomly generated initial values of the five gains. To simplify the procedure, the missing and false cue gains K_{mc} and K_{fc} were only allowed to attain more physically interpretable values, i.e., rounded to half values, e.g., $K_{mc} \in [0.0,\ 0.5,\ 1.0,\ \ldots\ 4.0]$.

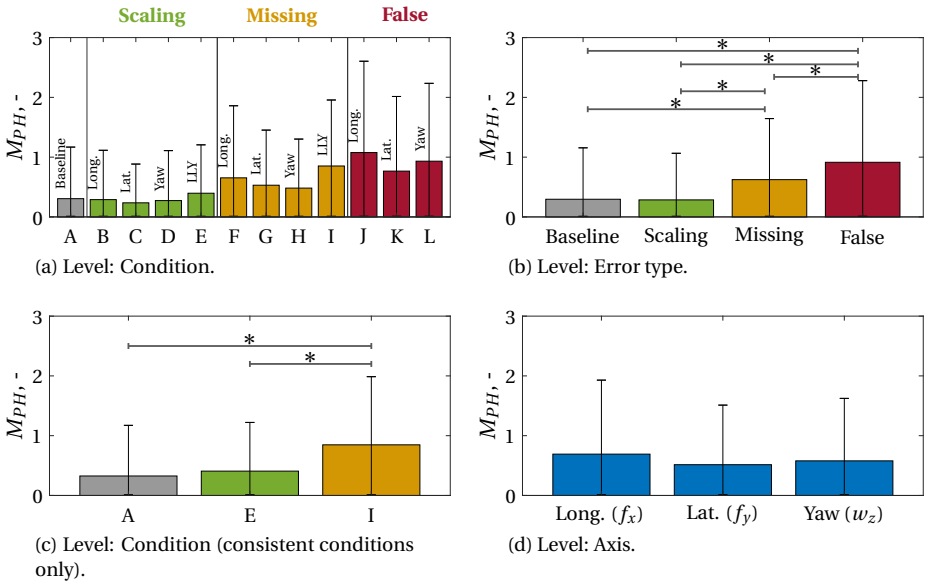
The resulting weighted mismatches (WM) model also includes the SR logic and is therefore denoted as p-ORC SR/WM. This model, with updated weights, indeed performs much better in predicting the overall post-hoc ratings, as visible by the circles in Figure 8.4a. Table 8.6 shows the parameters of the various models as well as the fit quality, confirming that the model p-ORC SR/WM strongly outperforms both the p-ORC and p-ORC SR models. The weights for the missing and false cue motions, $K_{mc} = 3.0$ and $K_{fc} = 3.0$, respectively, were best found to measure the rating data. This indicates that these error types were rated three times higher than the linear rating model p-ORC predicts. The model also outperforms a variant without the SR logic, i.e., fitted to the full reference motion (p-ORC WM).

Table 8.6: Gains and cost function values of the four evaluated MIR rating models. The gains of the models p-ORC/p-ORC SR come from Kolff et al., 2024b (Chapter 4).

model	ref. scaling	K_{f_x}	K_{f_y}	K_{ω_z}	K_{mc}	K_{fc}	$\sum(OR_{PH} - \widetilde{OR}_{PH})^2/12$
p-ORC	1.0	0.62	1.11	1.08	1.0	1.0	3.82
p-ORC SR	0.6	0.62	1.11	1.08	1.0	1.0	5.55
p-ORC WM	1.0	1.12	0.73	3.49	2.0	2.0	0.33
p-ORC SR/WM	**0.6**	**1.05**	**0.65**	**4.51**	**3.0**	**3.0**	**0.20**

8.4.3 Post-hoc MISC Ratings

The post-hoc MISC data were subjected to the same analysis steps as the post-hoc MIR ratings. An analysis on the AIC of the various model predictors (Table 8.7) shows that, similar to the post-hoc overall MIR analysis, the model fit did not improve significantly when adding 'axis inconsistency' as random effect. The fit did improve when adding 'error type' and 'axis' independently compared to the fixed effect-only model. Similar to the post-hoc overall MIR predictors, the largest AIC reduction of a single predictor is achieved using the 'axis'. Furthermore similar, the random intercept model 5, including axis and error type, has the lowest AIC value and thus performed best. This indicates effects of both error type and axis on simulator sickness.



(a) Level: Condition.

(b) Level: Error type.

(c) Level: Condition (consistent conditions only).

(d) Level: Axis.

Figure 8.5: Post-hoc Motion Illness Symptoms Classification (MISC) values for the considered levels. Green, yellow, and red lines indicate scaling, missing, and false cue errors, respectively. Grey line (condition 'A') is the baseline. For each condition in Figure 8.4a, the manipulated Degree of Freedom (DoF) is indicated (longitudinal, lateral, or yaw). LLY indicates manipulations in longitudinal, lateral, and yaw directions. The star symbols indicate significant differences.

Figure 8.5a shows the post-hoc MISC distributions per condition. Note that here it is clear that, when considering the whole group, the participants did not get very sick, considering $M_{PH} < 3$. The distributions grouped per error type are shown in Figure 8.5b. Here, the groups differ significantly except for baseline and scaling (similar to the MIR results), with baseline and scaling inducing the lowest ratings of simulator sickness, followed by missing and false. Conditions A (0.8 0.8 0.8), E (0.4 0.4 0.4), and I (0.0 0.0 0.0) were again also considered separately, confirming that there was indeed no significant difference between the baseline and scaled motion conditions, but both the baseline and scaled motion conditions being advantageous over missing motion (see Figure 8.5c); Hypothesis H3 can therefore be confirmed. Finally, Figure 8.5d shows the conditions grouped by axis. Here, the longitudinal specific force motion induces most sickness, followed by the yaw rate motion, and then the lateral specific force motion. The post-hoc tests did not yield any significant results, indicating that effects were too small (and hence: practically less relevant) to be detected by the corrected post-hoc tests (Figure 8.5d); Although the nested model analysis from Table 8.7 points at the existence of an effect, Hypothesis H4 cannot be confirmed.

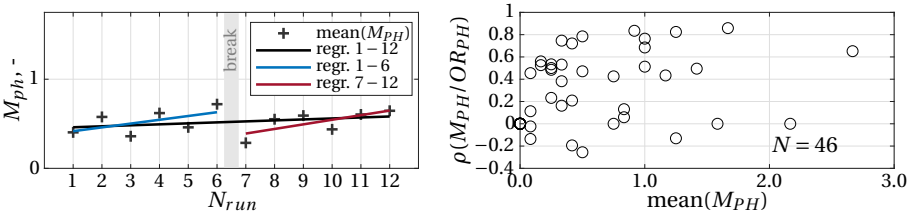
One possible concern in simulator sickness research is the occurrence of carry-over effects between the runs. Figure 8.6a shows the increase in M_{PH} over the twelve runs during the experiment, averaged over the participants. Note that, due to the different order of the experimental conditions presented to the participants, the actual underlying tested condition in each run differs per participant. The black line represents a linear

Table 8.7: Model performance comparison with varying predictors, dependent measure: post-hoc Motion Illness Symptoms Classification (MISC). Intraclass Correlation Coefficient (ICC) and R^2 values cannot be provided for the random effects models. Model 5 (bold) performs best in terms of the Akaike's Information Criterium (AIC).

model #	predictors	ICC (adj.)	ICC (unadj.)	R^2 (conditional)	R^2 (marginal)	AIC
0	subject	.53	.53	.53	.00	1568.49
1	condition, subject	.58	.55	.61	.06	1527.93
2	condition, subject, axis inconsistency					1529.93
3	condition, subject, error type					1497.96
4	condition, subject, axis					1205.59
5	**condition, subject, error type, axis**					**1179.26**

regression to the data points with a slope of 0.011. This implies that over the twelve runs, there is a slight build-up in MISC, resulting in an overall MISC increase of $12 \cdot 0.011 = 0.13$. This low value shows that participants barely got more sick *throughout* the experiment, showing that the carry-over effects between the conditions were indeed limited. However, the break that was taken between runs 6 and 7 is visible, as this holds the largest decrease in M_{PH} between any two runs. When performing separate regression fits for the data before and after the break (blue and red lines), larger slopes are present (0.043 and 0.052, respectively), indicating a stronger, but still limited build-up of the MISC over the conditions.

8.4.4 Relating Post-hoc MIR and MISC Ratings



(a) The post-hoc Motion Illness Symptoms Classification (MISC) averaged over participants for consecutive runs, including regression ("regr.") fits for all runs $(1-12)$, before $(1-6)$, and after $(7-12)$ the break.

(b) The mean post-hoc Motion Illness Symptoms Classification (MISC) given versus the Pearson correlation between the post-hoc MISC and Motion Incongruence Rating (MIR) ratings per subject.

Figure 8.6: Relations of the post-hoc Motion Illness Symptoms Classification (MISC) ratings.

Considering that the extraction of the post-hoc MIR and MISC values occurred at the same point in time, it is possible that participants' answer for both are directly correlated. Therefore, Figure 8.6b shows, for each subject, the Pearson correlation between the post-hoc MISC and MIR values as a function of the average post-hoc MISC value. The figure shows that for subjects that do not get sick (i.e., low M_{PH}, the correlation values show a large range between -0.4 and 1.0. For most subjects (83%) the correlation is

low (< 0.6), indicating that they answered the questions on MIR and MISC in different ways. However, for subjects with higher MISC values, these values correlate better to the MIR, up to a correlation of 1.0, whereas low correlations do not occur anymore. For these subjects, how sick they feel corresponds to how they disapprove the motion. This might indicate that the MIR can indeed be affected by the participants' current level of simulator sickness. The participants with a negative correlation gave lower MIR values for conditions where they indicated to get more sick (higher MISC), although this only happened to five participants and with low correlation values (< -0.2), which might not be of practical significance.

8.4.5 Continuous MISC Ratings

The continuous MISC ratings obtained *during* the simulations are shown in Figure 8.7. The individual rating data are included in Appendix D.4. Note that the subfigure names correspond to the conditions ('A' to 'L'). The lines represent the mean values over all participants; the spreads indicate the 95% confidence intervals. The maneuvers as defined in Figure 8.2a are indicated as a reference. Note that the values are generally low (<2), as they also include the participants that rated their simulator sickness with 0 all the time.

The colors green, yellow, and red indicate the scaling, missing, and false cue conditions, respectively, similar to Figures 8.4 and 8.5. In all figures, the coloured circles indicate the post-hoc MISC values given after each drive, averaged over all participants. As explained in Section 8.3.5, participants were instructed to let the post-hoc MISC value represent the level of simulator sickness at that point in time, i.e., after the run. Therefore, this should be similar to the last continuous rating given, although there is some time (approximately ten seconds) between their extraction. Generally, these values indeed correspond well; in 76% of the cases the same value was given (over all conditions), which shows the validity of the continuous MISC rating method. Note that the continuous rating value was not reset to 0 at the start of each run. Therefore, at the start of each run it starts at the last logged value of the previous run. As is visible in most conditions (such as Figure 8.7K), participants here decrease the $M(t)$ value. This is likely caused by the small 'breaks' between the conditions, lasting around 30 seconds, allowing for some relief and resulting in participants decreasing the rating right at the start of the next run.

Considering the difficulty of statistically analyzing differences in continuous ratings, the analysis on the differences between the conditions is only provided qualitatively. Comparing the conditions without axis inconsistency (Figures 8.7A and 8.7E) shows that whereas lower MCA gains do not result in an increase in simulator sickness, even though mismatches are present, providing *no* motion in the three main axes *does*, see Figure 8.7I. The ratings are even similar in magnitude to the conditions with false cue motion (8.7J-8.7L). For the scaled conditions with axis inconsistencies (Figures 8.7B-8.7D), i.e., conditions were only one axis was scaled down to 0.4, no clear increase in sickness is visible over time. Comparing scaled and missing motion (with axis inconsistencies, Figures 8.7B-8.7D vs. Figures 8.7F-8.7H), the MISC-values *do* increase over time. This shows that providing a form of scaled motion, even in case of inconsistencies, has a beneficial effect on simulator sickness compared to when a motion channel is not active at all. Finally, for the false cue conditions (Figures 8.7J-8.7L), the MISC values are the highest. Here, it is also visible that the MISC-values increase as a function of specific ma-

neuvers. For the condition 'J', with false cue motion in the longitudinal channel, the reported MISC values especially increase after strong longitudinal false cues, for example at 'DEC3', the final braking maneuver. Similarly, conditions 'K' and 'L' increase most after corner maneuvers, such as the roundabout ('RBT').

8.5 Discussion

The presented study investigated the effects of motion mismatches on motion incongruence ratings as well as simulator sickness. Scaled, missing, and false motions were systematically applied to the longitudinal (f_x), lateral (f_y), and yaw (ω_z) channels. Participants rated their level of simulator sickness using the MISC [Reuten et al., 2021; Wertheim et al., 2001], both continuously during the drive as well as directly after (post-hoc). They also rated the incongruence of the motion, using a post-hoc MIR. They also rated their level of simulator sickness using the MISC [Reuten et al., 2021; Wertheim et al., 2001], both continuously during the drive as well as post-hoc.

8.5.1 Post-hoc MIR Ratings

The post-hoc MIR ratings show strong evidence of an effect of error type, with false cues inducing the highest MIRs, followed by missing and scaled cues, confirming Hypothesis H1 with regard to post-hoc ratings. There was no significant difference across baseline and scaled motion. There was evidence of an effect of the axis. The lateral specific force (f_y) mismatch was indeed found to be the most critical for the MIRs, followed by longitudinal specific force (f_x), then yaw rate (ω_z) mismatches. This is in line with the findings of Kolff et al. [2024] (Chapter 4). Significant differences were only found between ω_z and f_y. Concluding, these findings only partially conform Hypothesis H2.

8.5.2 MIR Rating Predictions

Applying the model of Kolff et al. [2024] (Chapter 4) resulted in overall values that underestimated rating of the missing and false cue conditions. The model that best predicts the data uses a combination of vehicle reference input scaling of 0.6 and applying additional weights of 3 to the missing and false cue motion. This thus results in an updated model structure, compared to (8.1):

$$\hat{\tilde{R}}(j\omega) = \left(\frac{\omega_c}{j\omega+\omega_c}\right)\sum_m K_{MFC}K_{\tilde{P}_m}\left(\left|0.6\hat{\tilde{S}}_{veh,m}(j\omega) - \hat{\tilde{S}}_{sim,m}(j\omega)\right|\right), \tag{8.2}$$

with K_{MFC} the gain for the missing and false cue motion. Furthermore, the weights of the three mismatch channels were retuned through a fitting procedure to accurately predict the ratings of the twelve conditions tested in the experiment. Due to the "extreme" conditions, where full false cue motion is not representative of realistic driving simulator experiments, the found mismatch weights might be less representative than the parameters found in Kolff et al. [2024] (Chapter 4), which was further validated in Kolff et al. [2024] (Chapter 7). However, the reference scaling and weights on missing and false cue motion can be used for future experiments while remaining at a linear model structure. Although in the experiment the conditions were explicitly configured to have scaling, missing, and false cue motion, and therefore easy to identify, realistic motion profiles

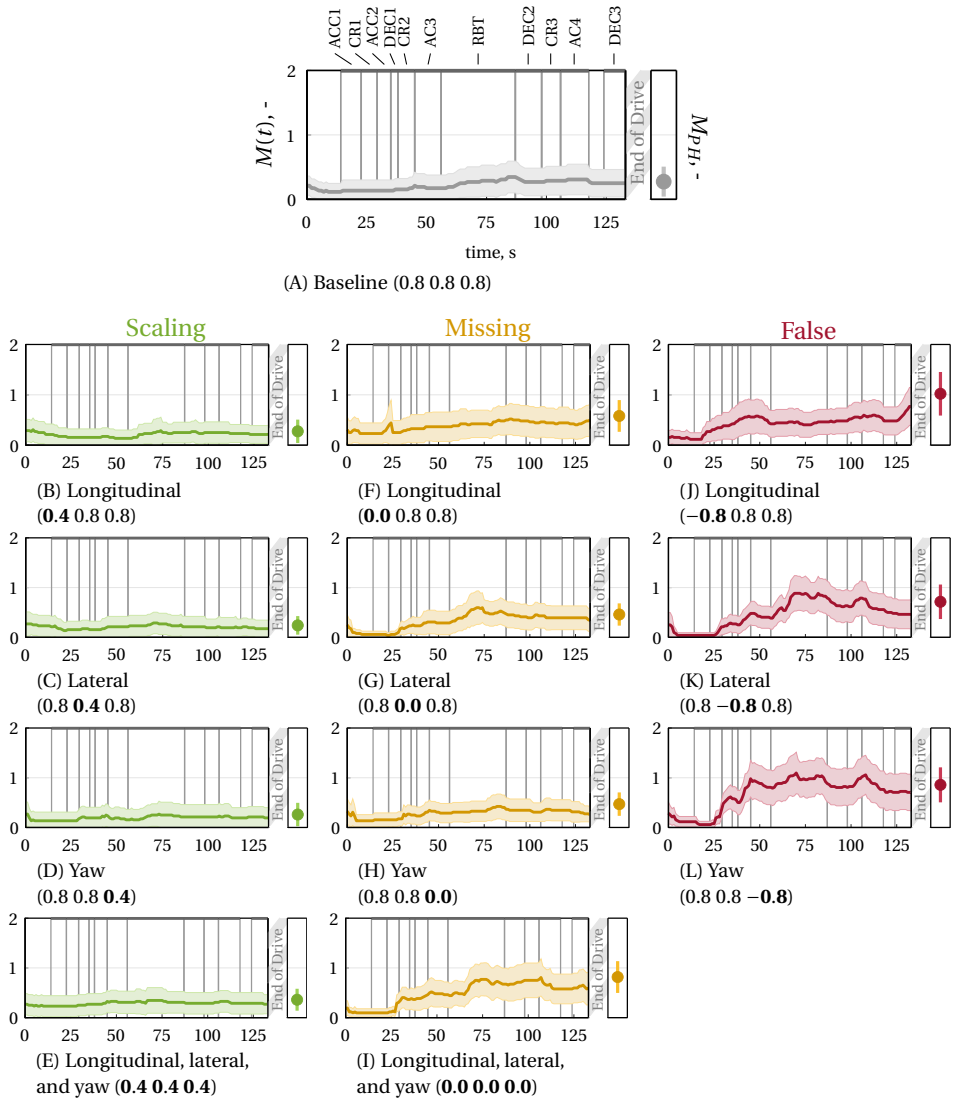


Figure 8.7: Average continuous Motion Illness Symptoms Classification (MISC) ratings as a function of time (left) and Post-Hoc (PH) MISC ratings (right) for each condition. The shaded areas represent the confidence intervals. The circles are the average post-hoc MISC values, with the vertical bars representing the confidence intervals.

will need specialized algorithms to identify which error types are present in the motion, such as proposed by Cleij [2020] and Kolff et al. [2022].

8.5.3 Post-hoc MISC Ratings

The analysis of the post-hoc MISC ratings indicated no relevant effect of axis mismatches. Regarding effects of different error types, false cues had the most negative effect on rated simulator sickness, followed by missing cues, and then scaled motion, confirming Hypothesis H3. Ratings for scaled motion did not significantly differ from those obtained in baseline motion. Although the model comparisons suggest an effect of the axis, the pairwise post-hoc tests were all non-significant, such that Hypothesis H4 cannot be accepted. Longer exposure times or more distinctions between the conditions would likely have been beneficial to find stronger differences in the post-hoc MISC ratings.

The confirmed Hypothesis H3 stated that false cues are worse than missing cues, and these are worse than scaling errors. Similar to the confirmed Hypothesis H1, both hypotheses can be interpreted differently, namely that false cues are not just worse than missing cues, but also *disproportionately* worse. The fact that false cues are rated worse than missing cues might explained by their larger objective mismatch. Hence, it is recommended to further investigate the roles of scaling, missing, and false cues in the case of equal mismatch magnitudes, as this could reveal whether the *type* of error induces relatively high or little simulator sickness and/or MIRs.

8.5.4 Continuous MISC ratings

For the first time in simulator sickness research, continuous MISC ratings of the level of simulator sickness were also recorded. These acquired ratings provided results that seem intuitively correct, namely that the ratings increase for maneuvers relevant for the specific condition (corners for lateral and yaw motion, accelerations/decelerations for longitudinal motion). Considering the similarity of the last given values and the post-hoc MISC – with post-hoc ratings being widely accepted as a measure of simulator sickness – this supports that the simulator sickness is a direct result of the motion mismatches, making the continuous rating method seem valid for measuring MISC values. The continuous MISC rating results furthermore allow an investigation how unpleasantness and/or simulator sickness is directly related to the mismatches of specific maneuvers. Reducing these mismatches will then automatically reduce simulator sickness. For experiments focusing on the modelling of motion or simulator sickness, such as Irmak et al. [2023], the continuous rating method thus also provides a useful alternative tool for obtaining high-resolution data on sickness.

8.5.5 Implications

The presented quality levels in terms of MIR and MISC provide useful insights for planning future simulator experiments. Missing motion was shown to be the second strongest contributor to simulator sickness and the MIR. This suggests that dynamic simulators are beneficial over static ones for a dynamic urban scenario. No significant difference was found between scaled and baseline motion. This shows that for the present urban simulation, the scaled motion (scaling of 0.4) is just as good as the baseline motion (scaling of 0.8), which was the maximal motion possible in the simulator

workspace. This corresponds to the range of acceptable scaling factors (0.4 – 0.75) found by Berthoz et al. [2013]. Even if the baseline motion is possible on a given simulator (Such as BMW's Sapphire Space simulator), working with a lower scaling factor might be beneficial when designing motion cueing for a driving simulator study. Possible inaccuracies in the motion cueing or the vehicle model might be less amplified, making the motion cueing easier to design or tune. Thus, for a scaling of scaling of 0.4, an acceptable state of motion cueing might be reached earlier and/or with less effort while at the same time reaching a similar level of MIR and MISC compared to applying a scaling of 0.8. For smaller simulators, with allowable scaling factors of 0.2 or 0.1, it is suggested that future work investigates whether similar levels of MIR and MISC uphold, or if these levels tend more towards the higher (i.e., worse) levels observed for the missing motion.

In the present experiment, the false cue motion, in which the motion was inverted, was found to be the strongest contributor to simulator sickness. Hence, it is of utmost importance to minimize false cues when conducting simulator experiments. Even though the false cue conditions themselves are not realistic motion conditions for a whole drive, the false cue motion in the maneuvers (partially) do correspond to washout motion. In washout, the simulator is moved back to its neutral position after a maneuver. This phenomenon, common in filter-based algorithms, is applied in motion cueing to ensure that the simulator is "ready" to cue the next maneuver. The present results regarding simulator sickness highlight the importance of methods to reduce washout effects, such as optimization-based methods that not only better exploit the available motion space, but can also keep the washout motion under the perceptual threshold.

Note, however, that the presented inverted motion is a form of false cue motion, whereas also other forms of false cue motion exist. For example, even though the false cue motion was completely inverted in the experiment, it is still derived from realistic car motion (in the present case, the baseline motion). However, false cue motion could also occur completely random. For example, pre-positioning can be employed by MCAs to move the simulator below the perceptual threshold of the humans to increase the available workspace of the simulator. If this motion *is* perceived, for example by faulty MCA tuning, this can create a false cue without any relation to the currently presented vehicle motion. The lower predictability of this motion might result in more sickness [Kuiper et al., 2020]. Specific future research in this direction is thus recommended.

With regard to axis effects, the acquired findings are especially helpful when planning the use of a dynamic simulator system with one rail (X or Y), which can usually be used to either redundantly provide f_y or f_x motion. When there is no specific dynamic phenomenon that is of interest, the findings suggest using the simulator with a redundant f_y axis to achieve the lowest (i.e., best) motion incongruences in an urban scenario.

The analysis of the post-hoc MISC evaluations is supported by the continuous MISC ratings. Although the latter were only analyzed descriptively, the severity of the various effects in the continuous ratings corresponds to the post-hoc ratings. Therefore, the continuous rating method can be considered promising with regard to identifying the maneuver-specific causes of sickness. The method can therefore be recommended for use in future investigations, as it reduces the workload compared to extraction MISC-values at a fixed point in time, such as in Diels et al.; Hogerbrug et al.; Irmak et al. [2023;

2020; 2021]. Furthermore, future work can explicitly look into modelling the continuous ratings, similar to models applied to predict MIR [Cleij et al., 2018; Kolff et al., 2024b]. Considering the strong relationship between the continuous and post-hoc MISC, this would be a useful method to be able to predict post-hoc MISC ratings. This would then allow for identifying simulators, MCA settings, and scenarios that induce a high degree of simulator sickness prior to a driving simulator experiment.

This also highlights another strength of the continuous method and the importance of the post-hoc MISC interpretation: post-hoc MISC ratings correspond to the state of the driver at the end of each drive, but do not necessarily represent sickness experienced *throughout* the drive, which can vary significantly over time. The continuous evaluation method provides the benefit of being useable in closed-loop (participants driving themselves) driving experiments, compared to the interval-based methods [Diels et al., 2023; Hogerbrug et al., 2020], as it can be changed on the participant's own initiative and induces little intervention and/or distractions to any other tasks, such as driving.

A final point of consideration considers the extraction of the post-hoc MIR and MISC values. It was found that the worse subjects feel, the better the post-hoc MIR values correlate to the post-hoc MISC. It is possible that the adapted physical state affects the perception of motion, leading them to consider more motion as incongruent motion. In that case, the MIR might become more of a representation of the MISC and less of the actual quality of the motion. Further research in this direction is recommended.

8.5.6 Limitations

Participants only drove open-loop (i.e., driven as passengers) throughout the whole experiment. Despite the fact that Kolff et al. [2024] (Chapter 7) showed that similar MIRs can be obtained with closed- and open-loop driving, this does not necessarily apply to MISC ratings. Closed-loop driving is known to induce less motion sickness in closed-loop driving [Rolnick and Lubow, 1991] compared to open-loop. As a consequence, closed-loop driving experiments might induce less simulator sickness than the presented results of this chapter suggest. Whether this also holds in simulators rather than real vehicles is unknown, although Kolff et al. [2024] (Chapter 7) suggests here there are no differences in closed-loop and open-loop driving. Furthermore, part of the motivation of the present study is to support the experiment design of other, upcoming driving simulator studies, such as simulator selection. It therefore had a focus on understanding relative differences between the tested conditions, which are likely similar in active driving. Having less simulator sickness in closed-loop driving than the open-loop results suggest would also make proper sample size selection more conservative.

The open-loop driving also allowed for testing a baseline variant (A) that was theoretically the best possible motion, as the simulator motion fully exploited the available workspace. This was only possible because all future vehicle states were known in advance, as a result of using a pre-recorded (open-loop) drive, allowing for fully optimizing the simulator motion in its available workspace. In closed-loop driving, future vehicle states are not known, as the speed, curvature, and lane position that drivers choose are not deterministic [Kolff et al., 2024a] and can only be predicted statistically [Eppink et al., 2023]. This would result in all drives being inherently different from each other.

Full optimization based on the whole workspace is then not possible, because extreme workspace positions must be avoided such that unpredictable actions from the driver still do not result in reaching the simulator workspace limits. As a result, closed-loop driving generally only allow less strong motion, such as by using smaller scaling factors. Consequently, the representation of the baseline motion tested in the experiment might not be representative of the baseline motion possible in closed-loop driving.

It should be further noted that the applied experiment design was not fully factorial, as this would have required a cueing condition with inverted motion in all three axes, where too many dropouts were feared. Scaled, missing and false motion may be differently relevant depending on the axis in which the incorrect motion is applied, while interaction effects could not be explicitly tested in the mixed effects models. Tentatively, there is no indication of the existence of such interaction effect (see Figure 8.5a). The qualitative data of the MIR, however, do suggest that the negative effects of missing and false motion in the ω_z axis may be most severe.

One limitation of the continuous MISC rating task is its passivity. Some subjects might have simply forgotten to change their continuous MISC rating, as they were not explicitly asked to do so while driving, such that lower values were recorded than actually was the case. This might also explain why in some conditions (mainly the false cue variants), the post-hoc MISC ratings were slightly higher than the continuous ratings at the end of each run. In the ideal case, these ratings would be the same, as they should measure the same simulator sickness phenomenon. Another explanation for their difference might be that simulator sickness symptoms might continue to get worse after each simulation (i.e., between the extraction of both rating methods). In contrast, symptoms of simulator sickness also could have decreased due to a relief from the sickening motion, although the observed higher values of the post-hoc MISC ratings compared to the last measured continuous ratings do not support this hypothesis.

A further point of consideration is that the measured continuous ratings also contain "fast responses", which can be attributed to feelings of discomfort. This does not have to be a specific limitation of the continuous rating method, as in fact, the same phenomenon can be present in post-hoc MISC ratings. An example of this is visible in Figure 8.7J, in which the post-hoc MISC value is dominated by the last maneuver. This also shows a prime benefit of the continuous ratings, as they provide insight in how the post-hoc ratings are potentially 'biased' by what occurred at the end of a run. For example, in the analysis of the post-hoc MISC, it was found that manipulations in f_x induce the most simulator sickness. Considering Figure 8.7J, this is not necessarily representative of how sickening the whole drive is. For the f_x channel, the post-hoc MISC value is likely largely influenced by the last maneuver, the false cue deceleration cue ('DEC3').

Last, the short duration of the drive may have been a real limitation of this study. With one run lasting 134 s, the drive may have been too short to evoke strong symptoms of simulator sickness, with the frequent breaks in which the post-hoc ratings were collected further mitigating symptoms. There were, however, several considerations for choosing shorter drives with moments of recovery in between. It ensured that symptoms did not get too severe, which allowed for more testing and reduced the risk of carry-over effects to subsequent conditions. Indeed, it was shown that these carry-over effects were limited, although truly eliminating these is impossible for a single-session experi-

ment. More incongruent motion might have extended the validity of the findings, but also likely would have increased the more severe and long-lasting sickness symptoms, but potentially would also increase the risk of carry-over effects. Another consideration was to remain close to realistic experiments, which often contain many short drives, for which the experiments are intended to be used. The limited length and lack of carry-over effects might have been a reason that generally low MISC values were obtained. This might indicate that the presented results are valid in this range. More research is required for investigating the validity under more severe simulator sickness symptoms.

8.6 Conclusion

This chapter investigated the effect of simulator motion mismatches on ratings of simulator sickness and motion incongruence in urban driving simulations. In a driving simulator experiment, subjects were driven as passengers ("open-loop") and experienced baseline, scaling, missing, and false cue motion mismatches in manipulations of the longitudinal (f_x), lateral (f_y), and yaw (ω_z) directions. The subjects evaluated their level of simulator sickness using the Motion Illness Symptoms Classification (MISC), both continuously during the drive as well as afterwards (post-hoc). The Motion Incongruence Rating (MIR), their subjective rating on the (in)congruence of the motion, was only extracted post-hoc.

For both the post-hoc MISC and MIR, it can be concluded that false cues produce significantly higher (i.e., worse) ratings than missing cues. The post-hoc MIRs are disproportionally higher ratings than the linear rating model of Kolff et al., 2024b (Chapter 4) suggest, requiring additional weighting on missing and false cue motion. Furthermore, scaling errors do not result in higher MISC and MIR than the baseline motion. The order found by Kolff et al., 2024b (Chapter 4) was confirmed with regard to MIRs: the reproduction of the lateral specific force motion is most important for minimizing motion incongruences, followed by the longitudinal specific force, and then the yaw rate. While there were no significant differences of the manipulated axis for simulator sickness, here the order was longitudinal specific force, yaw rate, and then lateral specific force.

The continuous MISC measurements furthermore support that the level of simulator sickness depends on the mismatches of the simulator motion. This also highlights the usefulness of this novel evaluation method to attribute the emergence of simulator sickness to explicit driving maneuvers during a simulation. Through this better understanding of the effect of mismatches on motion incongruence and simulator sickness, both can be more accurately avoided in the development and tuning of future urban driving simulations.

Improving Motion Cueing Quality

9

Motion Cueing Selection

Key findings

- Including the potential quality that each motion cueing method can reach leads to a fairer and more effective comparison between the cueing methods.
- Next to the achieved quality of a simulator's motion, including the potential of the quality, cost, and tuning metrics relevant for an experiment greatly determines the quality of a motion cueing method (simulator and MCA) and should thus guide the method's selection process.

Abstract

In driving simulation, the trade-off between different motion cueing methods (the motion cueing algorithm and the simulator) is typically made with an exclusive focus on the quality of the motion provided. In practice, however, many other quality, cost, and tuning characteristics may also affect the motion cueing quality and should therefore be included in the trade-off. This chapter provides an example assessment of how such metrics (quality of the motion, energy consumption, tuning effort, cost, and noise) can be evaluated based on the potential motion that a given MCA and simulator combination can reach once the tuning is performed. This potential motion was estimated using a non-linear optimization of the simulator motion based on a baseline tuning, defining an upper limit that an MCA may reach once the tuning would be performed, leading to a fairer and more effective trade-off. Example assessments of two real driving simulator experiments with motion cueing show that the best-suited simulator indeed differs when including the proposed metrics. This assessment may serve as an example for a more accurate, better-balanced, multi-faceted trade-off, and effective selection of motion cueing algorithms and simulators, improving the quality of future driving simulation experiments.

9.1 Introduction

SELECTING the proper motion cueing settings is an important, but difficult process in driving simulator experiments, for which few guidelines exist. Improper settings create differences between what a driver perceives in the simulator and what they would expect from the real vehicle, known as *incongruences*. The goal of the MCA, which drives the simulator's motion system, is to avoid these incongruences as much as possible, while remaining in the constrained workspace of the simulator. A proper definition and assessment of the *motion cueing quality*, suitable for the experiment in question, is thus paramount as the guiding principle in the comparison of MCAs and simulators.

Most analyses on motion cueing quality solely focus on the quality of the generated motion accelerations and rotations. As a consequence, many other properties of MCAs and/or simulators (e.g., tuning effort, energy consumption) are often not considered when analyzing the motion cueing quality. In contrast, Nahon and Reid [1990] described a qualitative comparison between three filter-based MCAs by comparing several metrics, such as the number of tuneable parameters as a measure of tuning effort, and the computational load. Their central notion, that a trade-off is more effective and fair when considering all relevant metrics for an experiment, gives rise to the question *which* metrics are then of actual importance to a specific driving simulator experiment, and how the identification of these metrics can lead to an improved trade-off of the available motion cueing methods. To this end, Chapter 2 introduced a variety of quality, cost, and tuning metrics, as well as a survey method to identify the most important metrics for each experiment.

Chapter 2 also described that a trade-off is even more fair if the *potential* of the motion cueing quality that each motion cueing method can reach is considered. Only then can it be evaluated whether it is worth the time and cost investment to tune an MCA for a specific simulator. Measuring this potential is currently not possible, however, precisely because the tuning has not yet been performed. It would be highly time- and cost-inefficient to tune all available motion cueing methods. The potential must therefore be estimated without knowing the exact tuning parameters yet. This also requires models of the various quality, cost, and tuning metrics, to be able to express how these metrics would change between the currently available tuning parameters and the potential that the motion cueing method can reach. Especially the metrics that depend on the human's subjective evaluation, such as the quality of the motion and simulator sickness, are inherently difficult to predict. It is here where the motion incongruence rating models developed in Chapters 4 to 6 provide a crucial contribution. Only with such models can the potential of the motion cueing be accurately included in terms of their impact on subjective metrics.

The contribution of this chapter, and the final step in this dissertation, is to demonstrate an example assessment of motion cueing methods for two real driving simulator experiments that have been performed at BMW. Simplified calculation methods for the most important metrics of the experiments are proposed and used to extend the concept of "motion cueing quality" among various quality, cost and tuning metrics. Both the exact simulator characteristics and MCA properties differ from their true values, however, due to data protection regulations. However, the goal of the chapter is not to define absolute metrics on the exact assessment of the two experiments in question, but rather to

present a methodological framework on how motion cueing quality should be assessed for future driving simulator experiments. As part of this, a method to estimate the potential of the motion cueing method is proposed, based on an optimization to fully exploit the remaining workspace. Combined, the chapter thus shows how the inclusion of quality, cost, and tuning metrics, as well as their potential for MCAs and simulators, may impact the choice of the motion cueing method.

The chapter is structured as follows. First, Section 9.2 describes the two driving simulator experiments, the proposed metrics, and the method for the potential estimation step. Section 9.3 describes the results of the analysis. This is followed by a discussion in Section 9.4 and conclusions in Section 9.5.

9.2 Methods

9.2.1 Experiment Use-cases

To demonstrate the calculation of relevant quality, cost, and tuning factors, together with the estimation of the cueing potential, two driving simulator experiments are analyzed.

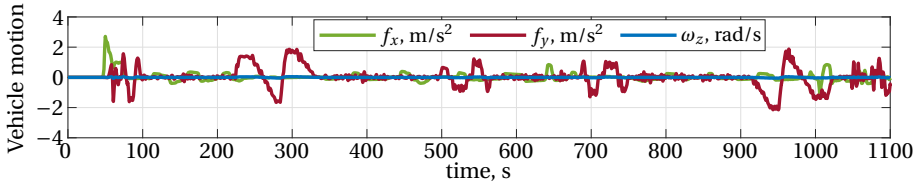
9.2.1.1 Experiment A: Fatigue in Automated Driving

The first experiment under consideration concerns the role of fatigue when driving with an ADAS system, and has previously been described in Chapter 2. First, participants drove for 15 minutes on a German highway scenario. After that, the autonomous driving systems were engaged until the participant reported a high level of drowsiness. This phase could take up to 90 minutes. For this reason, the noise of the simulator was to be reduced as much as possible, as this could affect the fatigue and attentiveness of drivers. After the first phase, a second manual segment was performed, lasting another 15 minutes. The three main vehicle signals for a representative reference drive are shown in Figure 9.1a. The use-case contains longitudinal motion (braking/accelerating) and lateral motion (overtaking and changing to a different highway, for this trace between 200-300 s and 900-1,000 s).

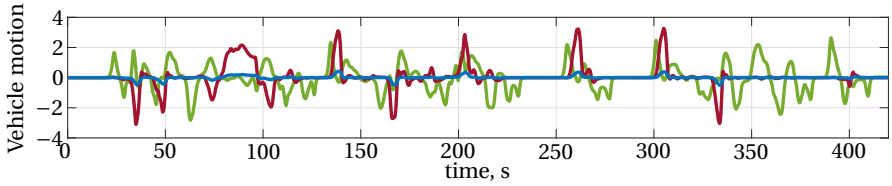
The goal here is to find the most suitable simulator for the experiment use-case. As determined in Chapter 2, the most important metrics in the experiment were the quality of the motion, stability, simulator sickness, tuning effort, and tuning complexity. To represent the wishes of the organization, the financial cost and the energy consumption are also considered.

9.2.1.2 Experiment B: Motion System Comparison Study

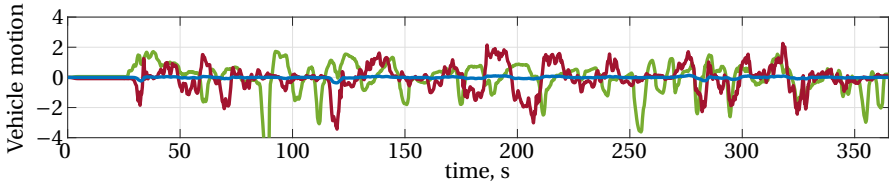
The second experiment under consideration performed a “motion system comparison study” of motion systems in an urban, rural, and highway simulation. The goal here is to identify the most suitable simulator for all three scenarios, although allowing for parameter differences in the MCA. The vehicle reference signals of three representative reference drives (urban, rural, and highway), gathered from a similar, representative experiment, are shown in Figures 9.1b to 9.1d. Especially the highway scenario (Figure 9.1d) contains little motion, as only slight accelerations and decelerations are present. For this experiment, the most important identified metrics through the survey of Chapter 2 include the quality of the motion, simulator sickness, tuning effort, energy consumption, and financial cost.



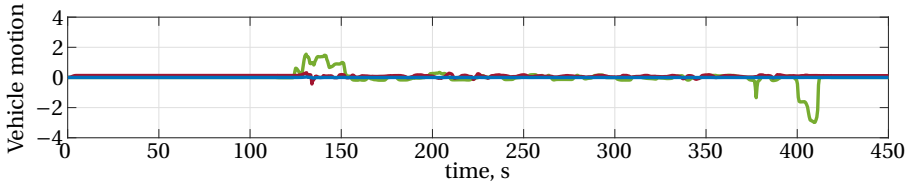
(a) Experiment A: Fatigue in Automated Driving



(b) Experiment B: Motion System Comparison Study, urban



(c) Experiment B: Motion System Comparison Study, rural



(d) Experiment B: Motion System Comparison Study, highway

Figure 9.1: The vehicle reference longitudinal and lateral specific forces, as well as the yaw rates for the two considered experiments.

9.2.2 Simulators

The simulators under consideration are the same ones as introduced in the introduction chapter of this dissertation:

- The *Green Stage* (Cruden B.V., 2021, Figure 9.2a) is a static simulator. Its visual system consists of ten displays, providing a field-of-view of approximately 180°.
- The *Ruby Space* (VI-Grade, 2015, Figure 9.2b) is a 9-DoF system (hexapod on tripod). The tripod can rotate ±25° and also has a 1.5 m workspace in both x and y-directions.
- The *Vega Vector* (Cruden B.V., 2021, Figure 9.2c) 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* (Cruden B.V., 2021, Figure 9.2d) is similar in design as the Vega Vector, but has an additional ±180° yaw-drive underneath, resulting in a 7-DoF system. The LED-wall is also similar to the Vega Vector, but covers the full 360° horizontal field-of-view.
- The *Sapphire Space* (Van Halteren Technologies B.V. and AVSimulation, 2021, Figure 9.2e) is BMW's largest simulator (9-DoF). It includes a large 6-DoF hexapod (total stroke of 1.15 m) with a single DoF ±175° yaw-drive on top. Its XY-drive underneath allows additional movement over an area of 19.14 m × 15.7 m. Visuals are projected by a full 360° projection system inside the enclosed dome.

9.2.3 Motion Cueing Algorithms

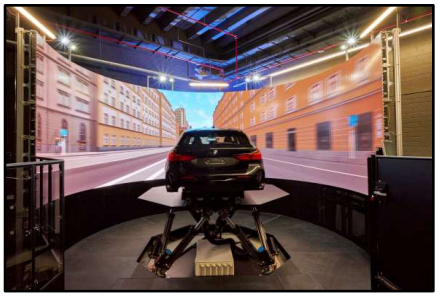
Two MCAs are considered for the analysis in this chapter. First, a typical CWA is included [Conrad et al., 1973; Reid and Nahon, 1985], as also used in Chapters 4-7. Due to the worst-case tuning and the CWA's "blindness" to future states, it cannot utilize a simulator's full potential. Typical for CWA motion cueing is the need for tuning its large number of parameters, a time-consuming process.

Second, the "Oracle" approach is considered as MCA, as previously used in Chapters 4-6 and 8. This method employs a non-real-time optimization-based cueing strategy with infinite prediction horizon. This MCA calculates the simulator's motion by minimizing a cost function. The algorithm contains the full non-linear kinematic descriptions of the analyzed simulators and is based on the implementation of Ellensohn et al. [2019]. The Oracle algorithm and its output motion can only be used in simulations where participants are driven through the environment ("open-loop") rather than driving themselves, as for real-time driving the future states are inherently unknown.

Considering the high focus on automated driving in Experiment A, the open-loop Oracle strategy is a considered a possible MCA here, next to the CWA (possible in both open- and closed-loop simulations). For Experiment B, where closed-loop driving is a vital part of the experiment, only the CWA is considered as the MCA.

(a) Green Stage (static).

(b) Ruby Space (tripod, hexapod).



(c) Vega Vector (hexapod).

(d) Sirius Vector (yaw-drive, hexapod).

(e) Sapphire Space (xy-drive, hexapod, yaw-drive).

Figure 9.2: The five simulators considered in the evaluation. Image courtesy: BMW Group.

Table 9.1: Parameters of the incongruence rating models in the three scenarios types.

model	ω_c [rad/s]	K_{f_x} [-]	K_{f_y} [-]
p-ALL rural A	0.35	1.14	1.00
p-ALL urban	0.36	0.91	1.50
p-ALL highway	0.21	1.09	1.46

9.2.4 Motion Cueing Quality Metrics

A set of quality, cost, and tuning metrics is defined that represents the motion cueing quality. It is important to note that these metrics are not intended as absolute measures, but rather serve as examples of how their implementation may be realized. Each of the metrics has been discussed in Chapter 2, such that the focus here lies on providing examples on how the metrics can be evaluated.

9.2.4.1 Motion Incongruence Ratings

To evaluate the various combinations in terms of the motion cueing quality, predictions on overall subjective ratings are made based on the linear rating model introduced in Chapter 4. This method relies on two steps.

First, a prediction of the continuous ratings [Cleij et al., 2018] is made based on the mismatches between vehicle reference and simulator motion. Chapter 4 proposed a linear model that predicts the continuous rating of the average participant using low-pass filter transfer dynamics $H_m(s)$ between the measured mismatch signals $\widetilde{P}_m(t)$ (inputs) and a modeled rating signal $\widetilde{R}(t)$ (output), with m the mismatches:

$$\widehat{\widetilde{R}}(j\omega) = \left(\frac{\omega_c}{j\omega + \omega_c}\right) \sum_m K_{\widetilde{P}_m} \widehat{\widetilde{P}}_m(j\omega) \tag{9.1}$$

Chapter 4 showed that in urban simulations, the continuous ratings can be largely explained when considering the mismatch channels $\widetilde{P}_{f_y}$ and $\widetilde{P}_{f_x}$. Similarly, Chapters 5 and 6 validated the model for highway and rural simulations, respectively. The different model parameter values that represent these scenario types are shown in Table 9.1.

The second step is the conversion to predictions of overall ratings (OR), which represent the whole drive. This rating is predicted using the most incongruent point occurring in a drive, i.e., the maximum of the modeled continuous rating:

$$OR = 2.0 + 0.8 \cdot \max[\widetilde{R}(t)] \tag{9.2}$$

This allows for comparing the various motion cueing methods of the experiment using a single scalar value.

9.2.4.2 Tuning Effort

Tuning of an MCA is an important aspect of the motion cueing process, as the output motion depends on the selection of the parameters. Especially for filter-based algorithms such as the CWA, few guidelines exist for tuning, such that it is often an iterative, trial-and-error process. As a measure of the tuning effort and complexity, the number of parameters is used (Chapter 2). As described in Chapter 2, the number of parameters of

Table 9.2: Example weights of the motion subsystems in kg. Actual weights differ.

	hexapod	x-drive	y-drive	yaw-drive	tripod
Green Stage	-	-	-	-	-
Ruby Stage	500	-	-	-	1,500
Vega Vector	1,500	-	-	-	-
Sirius Vector	1,500	-	-	1,000	-
Sapphire Space	2,000	60,000	5,000	1,000	-

an MCA generally depends on the number of DoFs of the simulator. For the CWA under consideration the number of parameters can be expressed as $n_p = 23 + 2b_{\psi_d} + 3b_x + 3b_y$, with the booleans b_{ψ_d}, b_x, and b_y indicating the presence of a yaw-drive, x-drive, and a y-drive, respectively. As an example, the tuning effort F may be expressed by:

$$F\ [\text{hrs}] = \sqrt{\frac{n_p}{2} \cdot (OR_{T_0} - OR_{T_1})}, \tag{9.3}$$

with the (predicted) overall ratings OR of the potential and baseline tunings T_1 and T_0, respectively. Here the rationale is that a tuning set that is still far away from the potential best result, will require more effort to tune. The combined relationship between the number of parameters and deviation from the cueing potential is likely to be nonlinear, as especially in the fine-tuning a significant time investment is required. Vice versa, when starting the tuning process using a baseline tuning with a large tuning potential, improvements are likely quickly made. This relation is thus described by the square root. For the Oracle MCA, the six weights of the three translational and three rotational channels can be chosen, but it generally does not require a lengthy tuning process due to the optimization nature.

9.2.4.3 Noise

Currently, no models exist to predict the noise production of the various simulators. From experience, however, the noise largely depends on the weight of the motion system involved and its motion. The noise is thus estimated through its power by:

$$N(t)\ [\text{W}] = \sum_d m_d |v_d(t) I_d(t)|, \tag{9.4}$$

with v_d the velocity of the respective motion subsystem, m_d its weight, and I the motion of the motion subsystem component in the DoF d in the inertial system. Example values for the weights of the various motion subsystems of the simulators under consideration are shown in Table 9.2. Note that the individual weights need to be added; the noise of the x-drive of the Sapphire Space (the lowest subsystem) is caused by moving all other components on top as well, with a total weight of $(60,000 + 5,000 + 2,000 + 1,000 = 68,000$ kg). In the present application, the average noise occurring during a drive is considered. Other approaches may include the peak induced noise level, or a metric expressing noise in specific frequency bands (e.g., to avoid resonance).

9.2.4.4 Simulator Sickness

Previous research has shown that for highway scenarios, simulator sickness is generally low on all considered systems [Himmels et al., 2022b]. 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 low. No accurate models currently exist to predict simulator sickness due to the difficulty of integrating the visual-vestibular interaction [Kotian et al., 2024]. As Chapter 8 showed incongruent motion to be related to simulator sickness, the motion incongruence ratings are considered as a suitable proxy for simulator sickness.

9.2.4.5 Stability

Stability is considered to be an important metric in the evaluation due to the experiment length. An instability may potentially crash a simulation and will likely require restarting the scenario, such that large parts of the data must be discarded. Stability is difficult to quantify and is thus only described qualitatively here. The use of filters in the CWA generally results in a stable simulation [Nahon and Reid, 1990], although the closed-loop driving may still result in participants driving the simulator into the workspace limits, potentially crashing the simulation. All simulators thus perform equally well here ("medium"). In contrast, the output of the Oracle MCA is fixed (open-loop), such that stability is "high". Although not considered here, an MPC algorithm can suffer from low stability, as its real-time optimization can cause unstable behavior [Fang and Kemeny, 2016].

9.2.4.6 Energy Consumption

As described in Chapter 2, energy consumption is a metric of increasing importance [Meike and Ribickis, 2011] due to the induced financial cost, environmental impact, and hardware limitations. Determining the exact energy consumption of a simulation is a complex task [Ordoñez et al., 2017], made additionally difficult by the various types of motion systems under investigation. As an estimate, the kinematic energy required to move the simulator is calculated. The main benefit of this approach is the high comparability in *relative* energy consumption in terms of the amount of motion. The kinetic energy is calculated by integrating the power over time:

$$E\,[\text{kWh}] = \int P\,dt = m_d \int v_d \left|\frac{dv_d}{dt}\right| dt, \tag{9.5}$$

with v_d the velocity of the respective motion subsystem and m_d its weight. t represents the time in which the simulator is active, i.e., the longer the experiment lasts, the higher the energy consumption. As in the calculation of the noise, the weight moved by each subsystem is the cumulative weight of the subsystem itself and other subsystems on top.

9.2.4.7 Financial Cost

In many cases, financial cost is an important factor in the design of driving simulator experiments. In terms of the motion cueing, cost is determined by preparing and tuning the MCA, as well as the cost of running the simulator (energy consumption, maintenance, etc.) [Bennett, 1995]. To express the cost of the motion cueing, a fictional "cueing currency" ¢ is introduced. In the present applications, the total cost of the simulation

(considering only the motion cueing) consists of the energy cost and the labor costs for tuning the motion cueing:

$$C\,[¢] = E \cdot C_{\text{kWh}} + F \cdot C_{labor}, \tag{9.6}$$

with C_{kWh} the price per kWh and C_{labor} the labor cost in designing the tuning. As an example, the prices are set to $C_{kWh} = 0.50\,¢/\text{kWh}$ and $C_{labor} = 50\,¢/\text{hour}$.

9.2.5 Motion Cueing Quality Potential

To estimate how a motion cueing method will perform *without* performing the actual tuning, an *estimation* of the cueing method's potential must be made. For this, an optimization problem is solved. Generally, a tuning process aims to reduce incongruences, which are predicted using the objective mismatches between the vehicle reference motion $\widetilde{S}_{veh,m}(t)$ and the simulator motion $\widetilde{S}_{sim,m}(t)$, i. e., $\Delta\widetilde{S}_m(t) = \widetilde{S}_{veh,m}(t) - \widetilde{S}_{sim,m}(t)$, with $\widetilde{P}_m(t) = K_{\widetilde{P}_m}|\Delta\widetilde{S}_m(t)|$. Here, m represents the mismatch direction. The simulator motion vector, containing the various mismatch signals, can be calculated as:

$$\widetilde{\boldsymbol{S}}_{sim}(t) = \boldsymbol{T}_{IB}(\boldsymbol{\beta})\boldsymbol{I}(t). \tag{9.7}$$

Here, $\boldsymbol{I}(t)$ represents the motion signals of the simulator in the inertial system. The matrix $\boldsymbol{T}_{IB}(\boldsymbol{\beta})$ describes the transformation from the inertial to the simulator's body system, as a function of the Euler angles $\boldsymbol{\beta} = [\varphi, \theta, \psi]$. The contents of this transformation matrix depend on the kinematic configuration of the simulator, as derived in Chapter 3.

The potential tuning T_1 is now determined by optimizing the gains k_d on the inertial motions $\boldsymbol{I}_{T_0}(t)$ of the baseline tuning T_0 to fully utilize the simulator workspace, i.e.:

$$I_{d,T_1} = k_d I_{d,T_0}, \tag{9.8}$$

where the subscript d indicates the motion in each DoF. The number of gains (i.e., optimization parameters) thus equals the total number of DoFs of a simulator's motion system. The optimization is then performed using rating model as the cost function:

$$J(k_d) = \sum_m \sum_t K^2_{\widetilde{P}_m} \left[\widetilde{S}_{veh,m}(t) - \boldsymbol{T}_{IB}(\boldsymbol{\beta})\, k_d I_{d,T_0}(t)\right]^2. \tag{9.9}$$

Note that the right-hand side represents the signals $\widetilde{S}_{sim,m}(t)$. The resulting nonlinear optimization is then solved using a the "fminsearch" function in Matlab, using a Nelder-Mead algorithm. As this does not guarantee obtaining the global optimum, 30 iterations with randomly generated initial conditions (i.e., the gains on each DoF) are performed. Each optimization is constrained to the workspace limits of the simulator's motion system under consideration, shown in Table 9.3. Furthermore, the angular derivatives $\dot{\varphi}$ and $\dot{\theta}$ are constrained to 3 deg/s [Reymond and Kemeny, 2000] to remain under the rotational threshold when using tilt-coordination.

Note that this method requires a baseline tuning set to begin with (T_0), as it can only increase or decrease the magnitude of an existing motion trajectory. It thus differs from an auto-tuning approach, such as described in Pham and Nguyen [2022].

An example of the potential estimation of Experiment B (Motion System Comparison Study), the urban simulation, is shown in Figure 9.3. The baseline tuning T_0 here came

Table 9.3: Considered workspace limits of the motion systems in the potential estimation.

	hexapod						xy-drive		yaw-drive
	x_h m	y_h m	z_h m	φ_h deg	θ_h deg	ψ_h deg	x_d m	y_d m	ψ_d deg
Vega Vector	0.40	0.40	0.3	15	15	25	-	-	-
Sirius Vector	0.40	0.40	0.3	15	15	25	-	-	180
Ruby Space	0.28	0.25	0.2	15	15	25	0.75	0.75	20
Sapphire Space	1.4	1.2	0.8	15	15	25	9.57	7.85	180

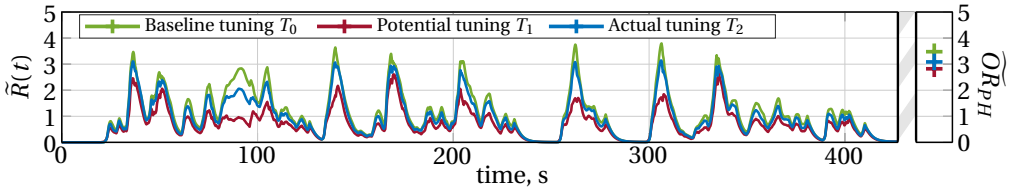


Figure 9.3: Predicted continuous and overall ratings of the baseline, potential, and actual tuning, for the Sirius Vector in the urban motion system comparison experiment.

from a similar urban driving experiment on the Vega Vector simulator. The potential tuning T_1 clearly provides better specific force cueing, as the predicted continuous ratings are lower. For example, the peak near 300 s shows a near halving of the predicted rating. The actual tuning T_2, defined after the assessment described in this chapter, forms a middle ground solution, indicating there was a better tuning to be performed than the baseline. Note, however, that because the maximum value of the predicted continuous ratings is not reduced as much between the T_1 and T_2 tunings (even though the most incongruent point is different, i.e., around 170 s and 320 s, respectively), the differences in the overall ratings are less pronounced.

9

9.3 Results

9.3.1 Fatigue in Automated Driving

The evaluation of the different considered metrics for the experiment A use case is shown in Figure 9.4. Figure 9.4a shows the predicted overall ratings of the CWA for the baseline T_0, potential T_1, and actual T_2 tunings. The Green Stage, as a static simulator, can clearly only produce "no motion". For all other simulators, and especially for the Sapphire Space, a strongly improved potential tuning T_1 exists, based on the initial tuning T_0.

Furthermore, the predicted ratings of the Oracle MCA are shown, which clearly outperforms even the potential tuning of the CWA, as expected [Ellensohn et al., 2019c]. Based on only the predicted ratings as a metric, the Oracle and Sapphire Space combination would be thus best for the given use-case. However, the additional motion in this case comes at a clear downside. As shown in Figures 9.4b and 9.4c, both the energy consumption and tuning effort strongly increase. Both also further increase the financial cost (Figure 9.4d). Finally, the noise power is strongly increased as well (Figure 9.4e).

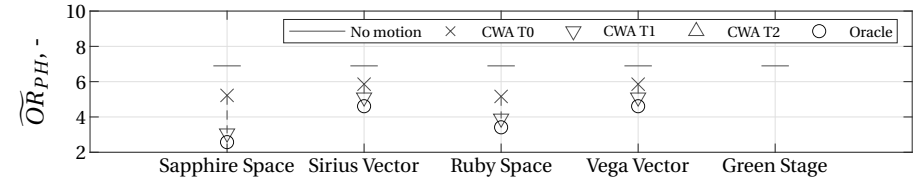
As described in Section 9.2, a potential quantitative selection is not further analyzed. Although important for the presented experiment, the stability of a CWA is generally not problematic. Furthermore, simulator sickness is assumed to be a proxy for motion incongruences, which are already included in the analysis via the predicted overall ratings. Based on the consideration of the combined metrics, and the requirement to perform the experiment partly closed-loop during the manual control phase (for which the Oracle is not suitable), the final experiment in this case in fact used the Vega Vector simulator. The present analysis shows that the Sirius Vector does indeed not provide a benefit over the Vega Vector, whereas the Sapphire and Ruby Space provide significant drawbacks in the metrics other than the motion cueing quality. From this argumentation, it thus follows that the Vega Vector was indeed a logical choice. However, considering that the Oracle clearly outperforms the CWA in terms of the predicted ratings and the financial cost (due to the lack of required tuning), the results suggest that Oracle may be a viable consideration. Considering the main focus on the autonomous driving part, in which the Oracle *is* a suitable solution, it may be opted to make an adaptation to the experiment itself to only offer autonomous driving. This would, in combination with the Vega or Sirius Vector, make the Oracle MCA the superior choice in the motion cueing method.

9.3.2 Motion System Comparison Study

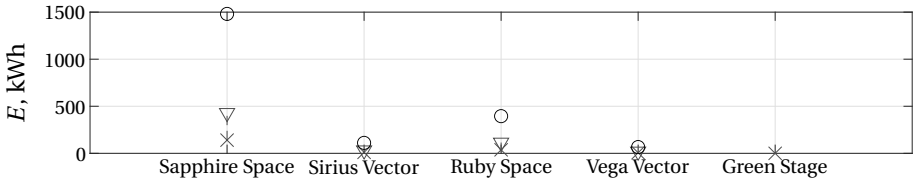
The metric evaluations of the second experiment are shown in Figures 9.5 to 9.7. Here, Oracle is definitely not feasible and is therefore not considered as an option. Similar trends as in the first use case are visible, where the potential of the Sapphire Space clearly outperforms all other simulators, while coming at the cost of a higher energy consumption, tuning effort, cost, and noise power. Note that especially in the highway simulation (Figure 9.7a), there are only small differences between the predicted overall ratings of the various tuning configurations, explained by the lack of maneuvers with strong motion (see Figure 9.1d).

In reality the simulator was already known beforehand (Sirius Vector), which followed as a requirement of the experiment design. Although the predicted ratings show that the Sirius Vector is indeed a suitable simulator for the experiment, a large improvement in motion cueing quality could have been obtained by opting for the Sapphire and Ruby Space simulators, given that the cost and noise metrics were of lower importance compared to the fatigue study. Furthermore, given that the differences between the Vega Vector and Sirius Vector are marginal, the former may have been just as suitable for the experiment as the Sirius Vector. This may be an important consideration in the logistical distribution of experiments, for example if another experiment would have to be performed on the Sirius Vector at the same time slot. Furthermore, to increase the comparability in the experiment itself, it was chosen to use the same tuning parameters across all three scenarios, restricting the freedom of parameters between the scenarios. Nevertheless, in all three simulations, the actual tuning "T2" comes reasonably close (within one rating point) to the potential tuning.

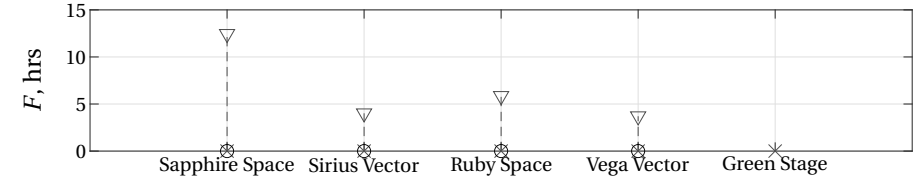
In this case, measured overall post-hoc ratings were in fact obtained after the final experiment. These are included in Figures 9.5 to 9.7 with the "*" symbol at the Sirius Vector results. To correct for the relative differences between experiments, as described in Chapter 6 and Cleij [2020], an MTP of 1.57 was used. The ratings thus vary from those



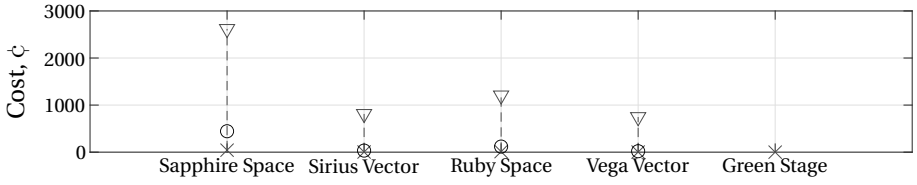
(a) Incongruence ratings



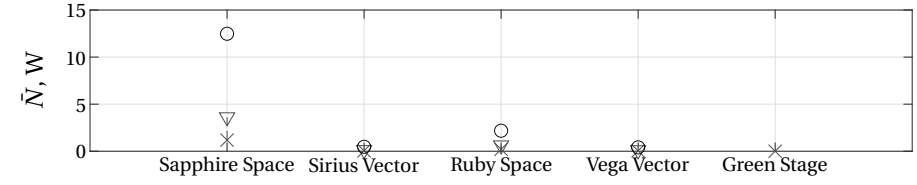
(b) Energy consumption



(c) Tuning effort



(d) Cost



(e) Noise power

Figure 9.4: The calculated metrics of the tuning configurations for the five simulators (Experiment A: Fatigue in Automated Driving).

shown in Figure 9.3. These ratings are close (within one rating point) to the predicted ratings, further validating the rating prediction method.

9.4 Discussion

9.4.1 Metric Evaluation

In all experiments, the Sapphire Space showed to be the simulator with the highest potential quality of the motion. This is not surprising, considering its large workspace in three separate motion subsystems. The additional quality, cost, and tuning metrics provided in this chapter show to have a strong effect on the motion cueing quality as well. While providing the relatively high-quality motion of the Sapphire Space, this comes with drawbacks in all other four considered metrics (energy consumption, tuning effort, cost, and noise power). This shows that, depending on the importance of these metrics in an experiment, the Sapphire Space may not always the best-suited simulator. Similarly, the same can be said about the Oracle MCA considered in Experiment A, although this MCA has the additional benefit of having little to no required tuning effort.

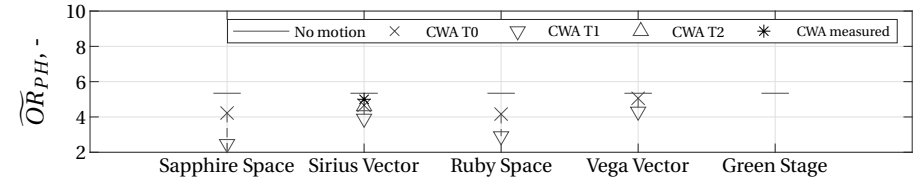
The absolute metrics will depend on the exact motion system and MCA characteristics, which likely differ between driving simulator institutes. Apart from the identification of the most important metrics itself, future work may investigate how the metric models can be made more accurate. Furthermore, once more accurate metrics are established, it would be useful to identify the effect of changing tuning parameters on the various metrics, i.e., through a dedicated sensitivity analysis.

A large part of this dissertation focused on the prediction of subjective motion incongruence ratings, which also formed a large part of the present analysis. The no-motion benchmark used in this chapter was based on the rating models of this dissertation, by calculating the predicted ratings based on having no simulator motion. To verify this approach, it would be more accurate to actually measure the continuous ratings in a simulator where no motion is applied. These ratings of a complete lack of motion are likely to be independent on the motion system itself, as the sense of motion is then obtained from the visual system only. Using this approach, tuning parameters that may lead to a motion cueing with *worse* ratings than no motion could be identified more accurately.

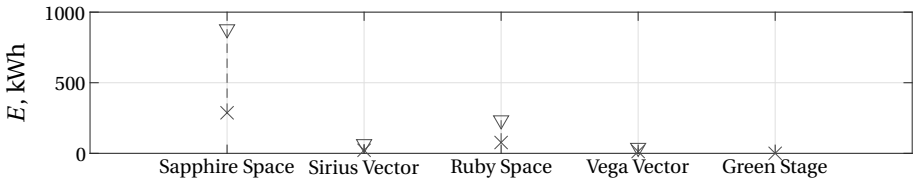
9.4.2 Potential Estimation

The potential estimation of the tuning provides a benchmark of how much the motion cueing can still be improved. This is a highly useful, and more fair, method of comparing motion cueing methods. For example, an MCA and simulator combination with a large potential, but badly tuned parameters, would then not automatically be neglected as a poor option. Once the method is applied more frequently and many baseline, potential, and actual tuning sets are collected, is nevertheless recommended to perform an additional on whether the current potential cueing estimation method is able to accurately predict the actual tuning set, or whether adaptations to the method are required.

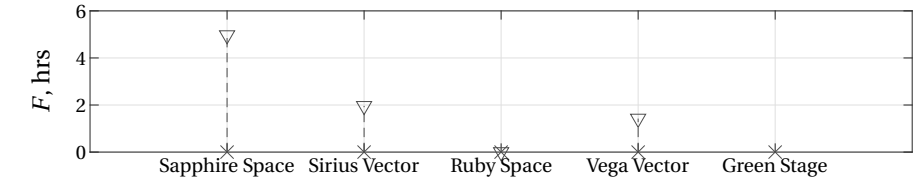
To further improve the potential analysis, future work should also consider the architecture of the MCA and simulator. In the method presented in this chapter, a full non-linear optimization method was applied to estimate the potential. Some MCAs, such as the CWA, may not be able to fully exploit the non-linear kinematics of the motion system.



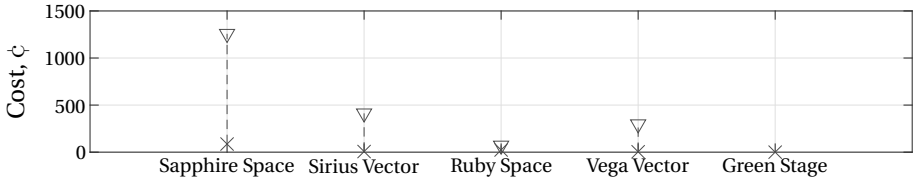
(a) Incongruence ratings



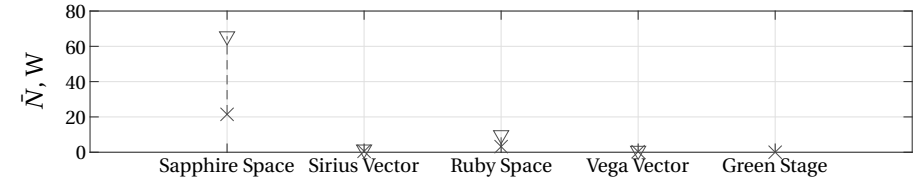
(b) Energy consumption



(c) Tuning effort

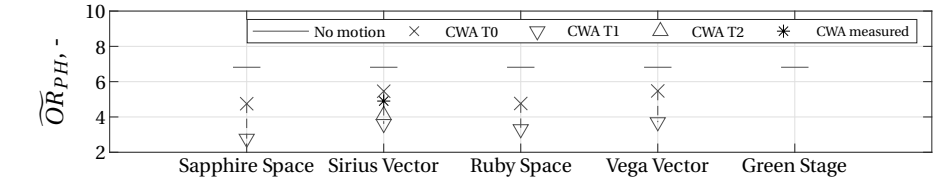


(d) Cost

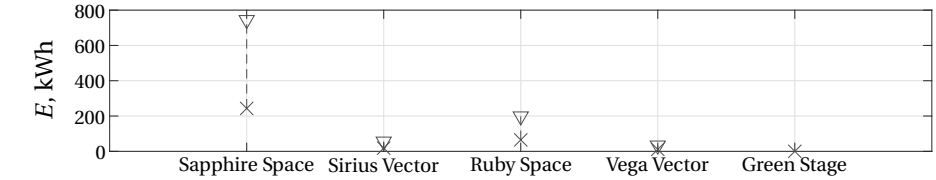


(e) Noise power

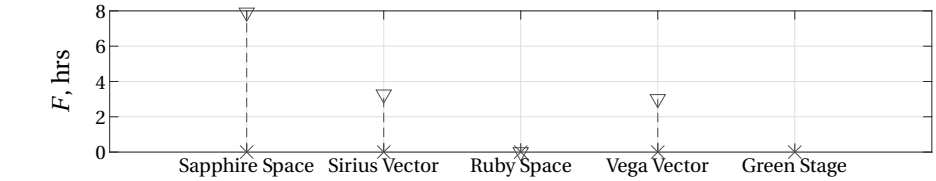
Figure 9.5: The calculated metrics of the tuning configurations for the five simulators (Experiment B: Motion System Comparison Study, urban)



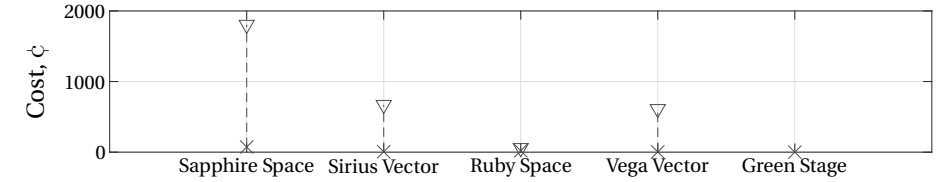
(a) Incongruence ratings



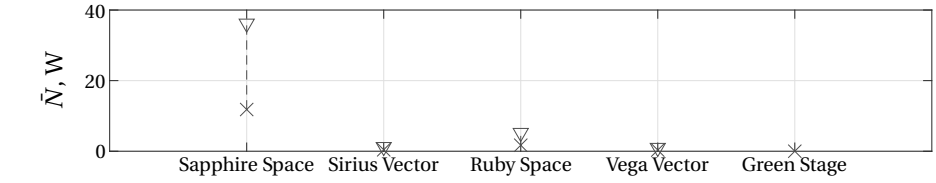
(b) Energy consumption



(c) Tuning effort

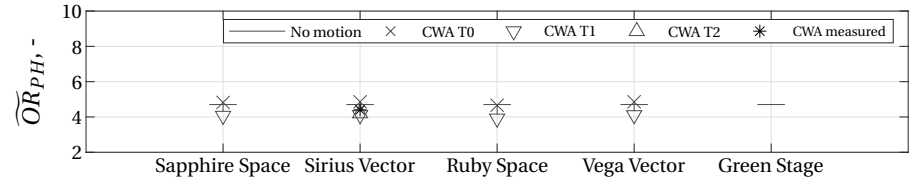


(d) Cost

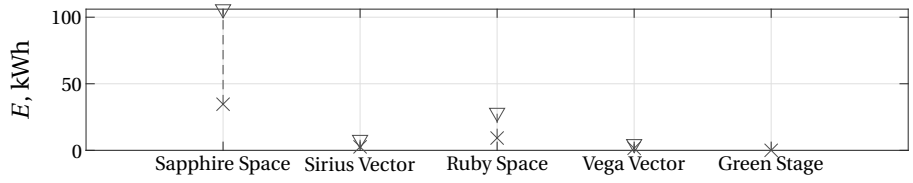


(e) Noise power

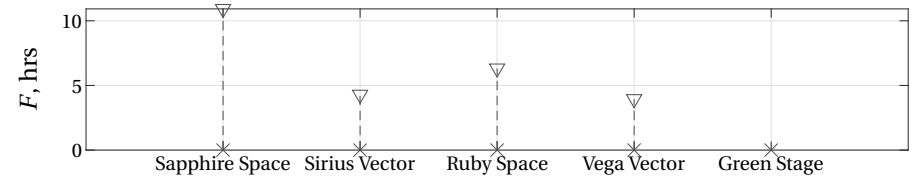
Figure 9.6: The calculated metrics of the tuning configurations for the five simulators (Experiment B: Motion System Comparison Study, rural)



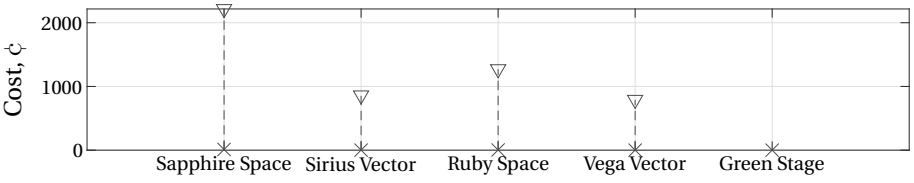
(a) Incongruence ratings



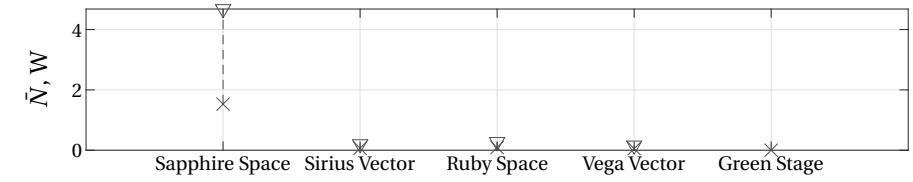
(b) Energy consumption



(c) Tuning effort



(d) Cost



(e) Noise power

Figure 9.7: The calculated metrics of the tuning configurations for the five simulators (Experiment B: Motion System Comparison Study, highway)

In that case, the predicted potential describes a state that the motion cueing method can never reach. In other approaches, such as MPC, the potential may even be calculated exactly by optimizing the motion beforehand, which often requires little tuning, as the optimization itself takes care of determining the proper motion output. Furthermore, for the potential method here presented for the CWA, more advanced methods can be applied, as the current method only used an adaptation of the MCA gains, rather than a full optimization of all available tuning parameters.

9.4.3 Motion Cueing Method Selection

How the various metrics may be traded off with each respect to each other is not yet performed in an objective manner in the current analysis. Here, the recommendation is to consider the various motion cueing methods in consultation with the various stakeholders in the experiment. Attempts to objectify this process, for example by weighing the various metrics in a cost function and selecting the method with the lowest cost function value, will likely always remain highly subjective, however.

For future work, apart from improving and/or standardizing this selection process of the method, one should look into validating the method as much as possible. For example, as also presented in Experiment B, overall post-hoc ratings can be obtained from the participants to confirm the validity of the selected method. Furthermore, the method can be systematically improved by applying it and collecting the various data. For example, the more driving data are available through the variety of analyzed experiments, the more representatively driving data can be used in the motion cueing selection for a new experiment. Similarly, if a motion cueing method is decided upon and actually tuned, this tuning set can serve as the baseline tuning for the next experiments, as shown in Figure 2.1.

9.5 Conclusion

This chapter described the analysis and trade-off for the MCA and simulator selection (the "motion cueing method") for future driving simulation experiments with motion cueing. A first attempt at providing measurable and quantitative metrics of motion cueing quality, including all relevant quality, cost, and tuning metrics was made for two real driving simulator experiment cases. This analysis further includes a straightforward method for predicting the available motion cueing *potential* of a certain method, rather than having to solely rely on existing sub-optimal baseline tuning sets in the comparison. The results showed that these additional metrics, next to the quality of the motion itself, can greatly affect the choice of MCA and simulator. Better motion cueing can thus be obtained when other metrics are taken into account, as well as by considering the potential that each motion cueing method can reach. This provides a fairer and more effective method to properly select an MCA and simulator, ensuring that future driving simulator experiments can truly use the best available motion cueing method.

10

Conclusions and Recommendations

DRIVING simulators operate at the forefront of automotive research. Properly designed motion cueing improves the validity of driving simulator experiments, and is thus crucial in guiding research and development. Determining how the motion cueing is best laid out for each separate experiment, by selecting the best-suited simulator, MCA, and tuning parameters, is a complex and multi-faceted design problem. Key difficulties are the (objective) identification of the most important quality metrics, the difficulty in measuring and predicting subjective ratings of the perceived motion cueing quality, as well as combining and efficiently weighing these aspects in an objective evaluation process. Therefore, the overall goal of this dissertation, as stated in Chapter 1 is:

Overall Research Goal

To develop a validated method to **objectively** assess and improve the **total potential** quality of motion cueing for **closed-loop** driving simulator experiments.

To accomplish this overall goal, research was divided into three parts, each with its own goal, corresponding with the three dissertation parts. The main findings, conclusions, and recommendations for future work are discussed in this final chapter.

10.1 Part I - Fundamentals of Motion Cueing Quality

Part I described several fundamental principles that are used in this dissertation. The goal of Part I was:

Research Goal 1

To develop a framework to evaluate the total motion cueing quality of an experiment, explicitly including its relevant quality, cost, and tuning metrics.

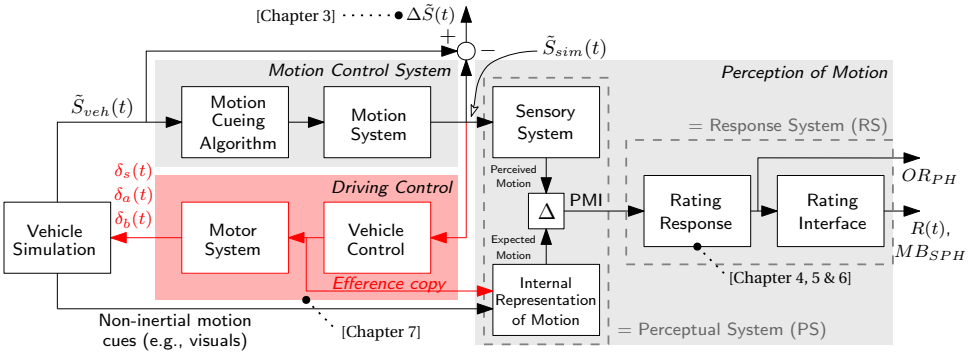


Figure 10.1: A description of the motion cueing rating process in closed-loop driving simulations used throughout the dissertation.

10.1.1 A Framework for Motion Cueing Quality

As a means to this research goal, **Chapter 2** described a framework to describe the motion cueing selection and tuning process that takes place prior to an experiment. Chapter 2 then proposed to identify and include quality, cost, and tuning metrics as part of the "total" quality, through an expert survey performed by the most important stakeholder groups for a typical driving simulator experiment. The survey revealed that apart from the quality of the motion itself, the financial cost, simulator sickness, ease of use, tuning effort, complexity, and stability together define the total motion cueing quality for driving simulator experiments. Including these metrics results in a more effective comparison and trade-off between different MCA and simulator options (the *motion cueing method*). Whereas this framework aimed to accomplish Research Goal 1, future work arose through the need for more objective trade-off methods of the various metrics.

In a trade-off between various MCAs and motion systems, a more objective comparison can be made if an estimation of their full motion cueing potential is considered. This in turn requires predictions of the motion cueing quality, as hypothetical "potential" motion cueing methods cannot be tested in a simulator yet. Whereas a method to provide this estimation thus needs to be developed, a crucial element of the framework is the ability to *predict* subjective ratings. The rating process of motion cueing is shown in Figure 10.1, which has been the main recurring figure throughout the dissertation, as it forms the basis for understanding and predicting subjective Motion Incongruence Ratings (MIRs). Chapter 2 therefore also gave an outline on how ratings on the quality of the motion can be predicted through objective metrics (such as through rating models), which forms the basis of Part II. Important open questions showed to be how humans perceive and rate motion cueing mismatches in a variety of driving scenarios (Chapters 4, 5, and 6), as well as the investigation on possible differences between open-loop and closed-loop driving (Chapter 7).

10.1.2 Kinematics of Motion Systems

Before subjective rating models can be developed, the objective mismatches that occur between the vehicle reference and simulator motion must be calculated correctly, taking

into account the full motion system kinematics. An important element of the framework is thus the derivation of the mismatch signals $\Delta\tilde{S}(t)$ (see Figure 10.1) for different simulator motion systems. **Chapter 3** first derived the kinematic relations for a six DoF hexapod simulator. The chapter then investigated the relations for the nine DoF Sapphire Space (yaw-drive on top) and Ruby Space (yaw-drive at the bottom) simulators at BMW, as these were the two motion systems used in the experiments of Part II.

The chapter first showed that the specific forces and rotational rates relations strongly depend on the order of the motion subsystems. A correct implementation of the kinematic relations is therefore key for correctly expressing the mismatches between vehicle and simulator motion. Second, this difference showed to also strongly affect the complexity of the motion control problem, as having a yaw-drive located on a hexapod's upper platform greatly complicates the rotational rate relations. The third finding of Chapter 3 was that for large-excitation yaw-drives (> 20°), the offset between the Motion Reference Point (MRP) (the point where the motion applies) and the CRP (the position where the driver is located) must be corrected for in the motion control to avoid erroneous simulator motion. Combined, the derivations of the kinematic relations of the Sapphire Space and Ruby Space, as well as the relations for the MRP corrections are highly relevant due to their fundamental kinematic differences. These relations also laid the foundation for the incongruence models discussed in Part II, in which the mismatches in the motion channels are analyzed and used for modelling (Figure 10.1).

10.2 Part II - Predicting Motion Cueing Quality

Part II aimed to fulfill Research Goal 2, which was stated as:

> **Research Goal 2**
>
> To objectively and systematically predict subjective motion incongruence ratings of closed-loop driving simulator experiments in urban, rural, and highway driving scenarios.

For this purpose, Chapters 4, 5, and 6 investigated whether the acquired rating data were reliable and predictable for realistic urban, highway, and rural simulations, respectively, by analyzing and comparing the rating response in these simulations (Figure 10.1). The rating models were developed based on objective mismatches derived in Chapter 3. An investigation of the applicability of the open-loop rating methods for closed-loop driving was presented in Chapter 7. Finally, Chapter 8 represented an extension of Research Goal 2 by applying the same methods to simulator sickness prediction.

10.2.1 Reliability

Predicting subjective MIRs requires a model of the subjective ratings. Developing a rating model requires reliable rating data to be fitted on. **Chapter 4** described an experiment in which participants rated the quality of the motion using the continuous rating method developed by Cleij et al. [2018], but then applied in a realistic urban simulation. A core finding of Chapter 4 was that the better the motion cueing is rated, the *less reliable*

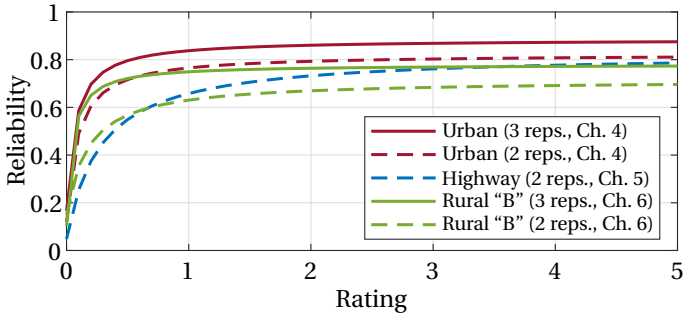


Figure 10.2: Reliability trends using two or three repetitions (reps.), showing a decreased reliability for more congruent motion, and a lower reliability for highway and rural simulations than for urban simulations.

these ratings are. Vice versa, motion that is rated worse leads to higher reliability. This implies that improving the quality of the motion simultaneously requires a greater effort in the validation of the improved quality.

Chapter 5 focused on a similar rating experiment on a highway scenario. The same simulator, experiment leader, experiment set-up, briefing, MCAs, and highly similar conditions were used; only the participant group and the scenario were different. This allowed for a unique direct comparison with the reliability estimates of Chapter 4. Similarly, **Chapter 6** directly compared three previously performed rating experiments in rural simulations [Ellensohn, 2020]. To compare the reliability between the different scenarios, the fits acquired on the reliability estimates of the continuous ratings acquired in the urban, highway and rural "B" experiments are shown in Figure 10.2.

Note that the urban and rural "B" experiments had three repetitions of each condition, whereas the highway experiment only had two repetitions. A lower number of repetitions automatically leads to a lower reliability of the data. To perform a fair comparison between the reliability trends of the scenarios, this must thus be taken into account. Therefore, the fits of the urban and rural "B" data sets were also evaluated for two repetitions only. This indeed leads to a decrease of the reliability compared to the fit based on three repetitions. Figure 10.2 furthermore shows that the highway and rural "B" ratings were less reliable than the urban ratings. An explanation is that in open-loop driving, upcoming maneuvers in rural and highway simulations (e.g., lane changes, accelerations) are more difficult to anticipate, and harder to rate than the more predictable maneuvers in urban simulations (e.g., corners, stopping in front of a traffic light).

The lower reliability for more congruent motion was visible in these three rating experiments, and to a lesser degree also in the rural "A" and "C" experiments analyzed in Chapter 6. As the ratings decrease (more congruent motion), their reliability also decreases. Understanding the associated reliability of obtained rating data is thus crucial in the evaluation of motion. It is advisable to always include an evaluation of reliability of the provided subjective ratings. The cost of improving motion cueing might thus not only be a higher required tuning effort in achieving this, but it will also become more difficult for drivers to confirm that the motion has indeed been improved.

Measuring a low reliability is not necessarily a sign of high motion cueing quality,

however, as low reliability could *also* be caused by a poor task understanding, in which a participant provides a rating that does not truly correspond to the sensed incongruences. Future work should also investigate whether test drivers use the continuous rating method in an actually continuous manner, or whether they only actively rate when incongruent motion appears. If the latter is the case, the continuous rating method may have an inherently time-varying reliability, where the points in the continuous rating recorded at congruent motion are less reliable than the points of incongruent motion.

10.2.2 Rating Relationships

The second key contribution of Part II is explicitly relating the continuous ratings to overall and maneuver-based ratings. The value of this contribution is that the overall and maneuver-based rating methods can be measured, and thus validated, in closed-loop driving experiments, whereas the continuous ratings cannot. Understanding these 'rating relationships' is thus necessary to be able to comply with Research Goal 2.

Chapter 4 showed that the overall rating of motion cueing in an urban drive strongly correlates with the worst motion occurrence *during* that drive. Furthermore, **Chapter 7** also found this to be true for maneuver-based ratings, where the highest rating (i.e., most incongruent) point in a maneuver was fully consistent with the participants' maneuver-based rating. Whereas the rating relationships provide a practical approach for estimating the overall and maneuver-based ratings based on the continuous ratings, a more fundamental perspective would be beneficial.

Especially the degree to which a maneuver is incongruent compared to the rest of the drive may affect the rating relationships. For example, a point of highly incongruent motion in an otherwise perfectly congruent drive may indeed strongly affect the overall rating. Vice versa, the most incongruent maneuver that occurs in an already mediocre motion cueing experience likely has a reduced impact on the overall rating.

10.2.3 Rating Models

The main focus of the chapters in Part II was to develop a predictive model of the subjective MIRs based on the objective mismatch signals derived in Chapter 3. For the urban rating data, Chapter 4, a linear rating model was shown to explain a high degree (90%) of the measured continuous rating variance. This model contains a low-pass filter response (with cut-off frequency ω_c) to the modeled PMI ($\widetilde{P}_m = K_{\widetilde{P}_m}|\Delta\widetilde{S}_m|$), weighted per motion channel m:

$$\hat{\widetilde{R}}(j\omega) = \left(\frac{\omega_c}{j\omega + \omega_c}\right) \sum_m K_{\widetilde{P}_m} \hat{\widetilde{P}}_m(j\omega). \tag{10.1}$$

In the urban driving experiment of Chapter 4, the subjective ratings of the motion cueing were dominated by the lateral and longitudinal specific force mismatches. Only a small and situational contribution of the yaw rate was found. This explanatory rating model was further validated in Chapter 7, which used different rating data acquired from a different participant group and simulator, thus showing the ability of the rating model to *predict* continuous ratings *across* experiments, an important step for Research Goal 2.

The succeeding Chapter 5 investigated how the acquired urban model transferred when used on highway scenarios. Furthermore, Chapter 6 investigated rural scenarios

and therefore aimed to complete the set of the three main scenario types (urban, highway, and rural). Rather than performing a new experiment, existing data sets from Ellensohn [2020] were used for the modeling. Here, the model structure of (10.1), with the response to the longitudinal and lateral specific force mismatches, was shown to successfully predict the continuous ratings for these scenarios as well. Although the balance in the contributions of the two channels were slightly different between the scenarios, this shows the explanatory validity of the rating model *across* three typical driving simulator scenario types. The linear model furthermore allows for a variety of applications, such as its use in linear optimization schemes. In that way, they can potentially be applied in MCAs to optimize for the *subjective experience*, rather than the purely objective cost function approaches in the state-of-the-art.

A central assumption in the work above has been that rating data acquired in open-loop experiments (participants being driven as passengers) are also representative for closed-loop driving (participants driving themselves). The experiment described in Chapter 7 explicitly verified the equivalence of ratings given by participants in an open-loop setting compared to them driving themselves, i.e., closed-loop. In the open-loop conditions, participants were presented with played-back recordings of their own closed-loop drives, allowing for the direct comparison between closed- and open-loop driving for identical motion cueing. Both maneuver-based and overall ratings were found to be equivalent in open- and closed-loop driving. This thus confirms that open-loop driving and the continuous rating method can be applied to make assessments on the motion cueing quality of closed-loop driving simulator experiments.

Combined, this shows that the model proposed in this dissertation can predict ratings in all three scenario types and is valid in closed-loop driving, thus fulfilling Research Goal 2. The end result is a transparent and relatively *simple* rating model that bypasses the need for complex visual-vestibular perception models (e.g., Markkula et al. [2019]) and works across MCAs, simulators, and participant groups. This allows for the model to be used *independently* of the motion cueing architecture. Clear benefits arise in saving on-site testing time and costs, as well as in determining the potential that a motion cueing method can reach based on its objective mismatch signals.

10.2.4 Simulator Sickness

Low motion cueing quality can lead to simulator sickness in driving simulators. As discussed in Chapter 2, avoiding simulator sickness is of importance in most driving simulation experiments. **Chapter 8** therefore presented an experiment in which the primary focus was to quantify the link between motion incongruences and simulator sickness, i.e., a sidestep of Research Goal 2. Next to post-hoc ratings, the continuous rating method was, for the first time, applied to rate simulator sickness. The continuous simulator sickness ratings indicate that the occurrence and severity of simulator sickness are directly linked to mismatches in simulator motion during specific maneuvers, similar to motion incongruences.

Specific manipulations in the motion cueing – i.e., applying scaled, missing, and false cue motion – showed that the latter induce the most simulator sickness, followed by missing cues and then scaled motion. Furthermore, results showed that mismatches in longitudinal motion induce the most simulator sickness, followed by the lateral motion,

and then the yaw rate. Similar research for highway and rural simulations, as performed in this dissertation for the MIRs, is strongly recommended. Furthermore, a next step in the measurement of continuous and post-hoc simulator sickness ratings would be to evaluate their reliability similar to the MIR data presented in this Chapters 4-7. This would, however, require several repetitions of the same condition. This experiment design was considered infeasible in the present experiment, where each of the twelve conditions was tested only once due to the already high number of conditions. This trade-off between the number of conditions and the amount of repetitions is, in fact, highly representative for the trade-off required for *any* driving simulator experiment where subjective ratings are involved. A final step in this line of research would be predicting the continuous Motion Illness Symptoms Classification (MISC) ratings, similar to the MIR models, which was not performed in Chapter 8 due to time constraints. Nevertheless, this is a crucial next step to enable evaluating the potential simulator sickness in the motion cueing selection of a future experiment.

10.3 Part III - Improving Motion Cueing Quality

Research Goal 3

To develop a methodology to select the potentially best-suited motion cueing method (simulator and MCA) of a driving simulator experiment to improve its motion cueing.

Chapter 9 provided an example assessment of how the metrics proposed in Chapter 2 can be used to predict the motion quality that a given MCA and simulator combination can *potentially* reach. This potential was estimated using a direct non-linear optimization of the simulator motion's scaling with respect to a baseline tuning, defining a limit in the quality that the motion cueing method may reach once the tuning would be performed. As shown by applying this methodology to two real driving simulator use cases, this potential estimation thus provides a fairer and more effective trade-off in selecting between different candidate MCA and simulator combinations, which was shown to have a large impact on the motion cueing method choice. Although the chapter provided several objective metrics used in the trade-off, the main goal was to provide a framework for a more accurate, multi-faceted, and complete trade-off in the selection of motion cueing algorithms and simulators, improving the quality of future driving simulation experiments.

10.4 Conclusions

Based on the previous discussion, the main conclusions drawn in this dissertation are:

- Next to the achieved quality of a simulator's motion, a mix of quality, cost, and tuning metrics determine the total quality of a motion cueing method (simulator and MCA) and should thus guide the method's selection process.

- Including the potential quality that each motion cueing method can reach leads to a fairer and more effective comparison between the methods.
- The objective mismatches between vehicle and simulator motion that are required for the prediction of motion incongruence ratings strongly depend on the kinematic configuration of the motion system.
- Subjective motion incongruence ratings become less reliable when the quality of the simulator motion increases.
- Continuous ratings in urban, highway, and rural driving scenarios can be accurately predicted through computing a first-order low-pass filtered response to a weighted sum of the objective lateral and longitudinal specific force mismatches.
- Overall and maneuver-based ratings that summarize a test driver's perceived incongruence across (part of) a simulated scenario strongly correlate with the most incongruent point.
- Subjective motion incongruence ratings in open- and closed-loop driving are equivalent, validating the continuous rating method and open-loop rating models for predicting closed-loop driving performance. Motion cueing can thus be systematically improved by employing open-loop driving during development and testing phases.

Finally, this implies that the *Overall Research Goal* of this dissertation is achieved:

- The motion cueing selection and design of future closed-loop driving simulator experiments can be assessed using the workspace potential that the motion cueing quality can reach, described by the objective simulator mismatches. This includes accurate predictions on subjective ratings of the motion using an open-loop rating model, and experiment-specific quality, cost, and tuning metrics. This results in more objective assessments and an improved selection process for the available motion cueing methods.

10.5 Recommendations

Five key recommendations are identified that, if followed, can further enhance the quality of future driving simulator experiments, as well as form the basis for future work.

10.5.1 Improvements Quality Evaluations

The findings of Part I discussed that the comparison of motion cueing methods is more objective when considering their full potential, as well as quality, cost, and tuning metrics. In the design of future driving simulator experiments, it is thus recommended to identify the metrics that can determine the quality of the motion cueing, as well as of the wishes of the stakeholders involved in the experiment.

There are, furthermore, a variety of directions in which future work can improve the selection method proposed in this dissertation. This includes a more objective, detailed and measurable definition of the considered metrics, such as energy consumption; in

this dissertation simplified calculations were used to highlight the metrics applicability, rather than providing truly accurate estimates. This was done to limit the complexity of the dissertation and to keep the application as general as possible. The exact metric definition, as well as their relative importance, will likely differ per institution, such that a "one-size-fits-all" approach is unlikely to exist. For this purpose, the survey may be used by a wider audience. Similarly, an improved and more objective trade-off framework, including a formal weighting *between* the different metrics, is a crucial next step.

Even if only the quality of the motion itself would be considered, such as in evaluation studies of novel MCAs (e.g., [Cleij et al., 2019; Ellensohn et al., 2020]), it is highly recommended to evaluate the motion cueing quality in terms of the *potential* that each MCA can reach for a fair comparison. The estimates of the potential cueing themselves can also be further improved. For example, automatic tuning approaches can be developed, rather than the presented workspace extrapolation method. Even if estimating this potential is not possible, at least the awareness that near optimally-tuned MCAs are used in a comparison would already be beneficial.

10.5.2 Ensuring the Reliability of Subjective Ratings

This dissertation has shown that subjective MIRs become less reliable when motion congruence increases. Great care should therefore be taken when using subjective ratings, especially for cases where highly congruent motion is achieved.

The reliability of subjective incongruence ratings can be increased by increasing the incongruence of the motion or by increasing the number of repetitions. If the reliability of the whole data set (i.e., all subjects) is considered, reliability can of course always be further improved by increasing the number of subjects. The findings also highlight that while a rating or evaluation may always be recorded, it does not necessarily have to be reliable, especially when the sample size is small. In the tuning phase of an experiment, in which only a handful of test drivers typically evaluate the motion, it is thus paramount to estimate or at least be aware of the reliability of the rating data. A tuning process based on formal rating scales would thus be a large benefit, as current tuning approaches often involve the test drivers giving informal comments whenever they want, rather than presenting quantitative ratings that have a specific range, timing and focus.

Generally, it is recommended to always perform *identical* repetitions of a single test drive to increase the number of repetitions and boost reliability. This is especially recommended in experiments where the motion cueing is expected to be highly congruent, for example on BMW's Sapphire Space. The reliability estimates described in Part II showed that three repetitions were generally sufficient to result in reliable data. This furthermore allows for the calculation of McDonald's Omega, which requires at least three data points. This reliability metric is a more accurate estimation of the true reliability and produced higher reliability estimates in all three driving simulation scenarios (urban, highway, and rural) tested in this dissertation.

10.5.3 Benefits of Open-Loop Driving

So far, the need for open-loop driving in rating experiments has been described as a *restriction* necessary to allow for more "invasive" rating methods, such as the continuous rating method. In some experiments, such as expert evaluations of driving dynamics,

closed-loop driving is a necessity to be able to give relevant ratings. In other experiments, such as for the design of motion cueing, open-loop driving may in fact also offer advantages due to its inherent benefits that can aid the evaluation process of motion cueing. As the explicit comparison of ratings of open-loop and closed-loop driving in Chapter 7 showed the ratings to be equivalent, open-loop driving can thus be effectively utilized as a useful tool in the design and verification of motion cueing methods.

For example, open-loop driving allows for a more systematic testing of motion cueing, as *repeatability* increases: it is possible to replay specific recordings several times, even to different test drivers, such that the *same* motion cueing is produced every time, which is impossible in closed-loop driving. This has benefits in terms of the reliability of the obtained rating data, as discussed earlier. Furthermore, open-loop driving also has benefits in terms of *controllability*, as specific situations can be purposefully tested. These benefits can thus accelerate the tuning process and, in turn, reduce the MCA tuning effort.

10.5.4 Toward Human-Centered Motion Cueing

Objective metrics of motion cueing quality are generally used in the current state-of-the-art design of MCAs, such as in the cost function of an MPC, whereas subjective ratings methods are used to evaluate the final designed motion. Novel MCAs can directly benefit from the findings on subjective ratings, however, by using a more human-centered approach in the MCA design and implementation.

The high resolution of the continuous rating data, as well as of the maneuver-based rating method, has given detailed insights in how humans evaluate motion. This knowledge can be more directly implemented in MCAs. For example, the continuous rating models can be implemented as a dominant cost function contribution in optimization-based algorithms. Furthermore, the balance in the weighting of the various motion channels can be altered to match the relative importance of the different motion channels [Cleij et al., 2020] as described in Chapters 4, 5, and 6. Current methods often apply weights of $[1,\ 1,\ 1]^\top$ for the specific forces and $[10,\ 10,\ 10]^\top$ for the rotational rates. These are based on the typical standard deviations of the motion signals [Katliar et al., 2015] and aim to make the translational and rotational cueing error contributions of roughly the same order of magnitude despite unit differences, without necessarily leading to the optimal motion cueing quality for drivers.

Furthermore, the insight that the most incongruent point in the motion greatly determines the overall impression of the participants can be leveraged as well. For example, the tuning process of the motion cueing can be accelerated when always first considering the worst segment of a drive. By weighing the maneuvers that would receive higher ratings more heavily, rather than also optimizing already good motion cueing, the overall impression of motion cueing can be improved.

10.5.5 Flight Simulation Applications

Whereas this dissertation solely focused on driving simulation, several approaches can be applied directly to flight simulation. In general, flight simulation has a different focus due to its higher importance of behavioral fidelity [Pool, 2012] and use for pilot training, in which subjective ratings are less suitable. Nevertheless, subjective ratings are widely

adopted [Mallery, 1987; Perfect et al., 2014; Pool, 2012; Wong et al., 2012]. The continuous rating method has already found an application in flight simulation, such as in aircraft upset [Bakker et al., 2025] and helicopter [Miletović, 2020] simulations, to evaluate different MCA types. Similar to the methods presented in this dissertation, both the reliability and predictability of such provided ratings can be further investigated to better understand subjective rating methods in flight simulation.

Appendices

A

Survey

The appendix contains the questions as presented to the stakeholders (in German) as described in Chapter 2. The questions and possible answers translated from German to English are indicated in italic.

A.1 Possible Answers

Stimme überhaupt nicht zu ☐ (1)	Stimme nicht zu ☐ (2)	Stimme eher nicht zu ☐ (3)	Teils / teils ☐ (4)	Stimme eher zu ☐ (5)	Stimme zu ☐ (6)	Stimme voll und ganz zu ☐ (7)	Ich verstehe die Frage nicht ☐ (-)
Fully disagree	*Disagree*	*Slightly disagree*	*Neutral*	*Slightly agree*	*Agree*	*Fully agree*	*I don't understand the question*

Table A.1: Possible answers, with the score values corresponding to Figure 2.2.

A.2 Survey Questions

A.2.1 Immersion

Während der Simulation kann das Gefühl entstehen, in der virtuellen Welt da zu sein (Immersion). Bitte beantworten anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
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 die Bewegung im Simulator so realistisch ist, wie möglich.
the motion in the simulator is as realistic as possible.

Q2 man in die virtuelle Welt reingezogen wird.
one is immersed into the virtual world.

Q3 die Simulatorbewegung sich nicht unerwartet verhält.
the simulator motion does not behave unexpectedly.

Q4 nie in die Grenzen des Bewegungssystems gefahren wird.
the simulator is never driven into its limits.

Q5 die Simulation einer Fahrt nie abstürzt.
the simulation of a drive never crashes.

A.2.2 Simulatorkrankheit (*Simulator sickness*)

Während einer Fahrt kann Simulatorkrankheit auftreten. Manchmal muss dadurch der Versuch abgebrochen werden. Bitte bewerten Sie anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
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:

Q6 schwere Simulatorkrankheitsymptome (Übelkeit, übergeben) so viel wie möglich vermieden werden.
heavy simulator sickness symptoms (nausea, throwing up) are avoided as much as possible.

Q7 die gesamte Ausfallsrate des Experiments so niedrig wie möglich ist.
the overall drop-out rate of the experiment is as low as possible.

Q8 leichte Simulatorkrankheitssymptome (Kopfschmerzen, Schwindel, Schwitzen) so viel wie möglich vermieden werden.
light simulator sickness symptoms (headache, dizziness, sweating) are avoided as much as possible.

Q9 der Versuch bis zum Ende gefahren werden kann (keine Ausfälle).
the experiment can be driven until the end (no drop-outs).

Q10 nach dem Versuch Probanden noch in der Lage sind, zu arbeiten.
participants are still able to work after the experiment.

Q11 die aufgezeichneten Daten immer vollständig sind.
the recorded data are always complete.

A.2.3 Verhalten des Fahrers (*Behavior of the driver*)

Bitte bewerten Sie anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
Using the following questions, please rate whether it is important to you that:

Q12 der gleiche Fahrstil in Querrichtung (Lenken) erzeugt wird, wie in einem Realfahrzeug.
the same driving style in the lateral direction (steering) is induced as in the real vehicle.

Q13 der gleiche Fahrstil in Längsrichtung erzeugt wird (Pedalerie und/oder Schalten), wie in einem Realfahrzeug.
the same driving style in the longitudinal direction (pedals and/or shifting gears) is induced as in the real vehicle.

Q14 das gleiche Verhalten mit bestimmten sekundären Komponenten (z.B. Navigation oder Radio) erzeugt wird, wie in einem Realfahrzeug.
the same behavior with certain secondary components (e.g., navigation or radio) is induced as in the real vehicle.

Q15 das gleiche Verhalten mit dem Umgebungsverkehr erzeugt wird, wie in einem echten Fahrzeug.
the same behavior with the surrounding traffic is induced as in the real vehicle.

A.2.4 Sicherheit des Fahrers (*Driver safety*)

Bitte bewerten Sie anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
Using the following questions, please rate whether it is important to you that:

Q16 man sich im Simulator immer sicher fühlt.
one always feels safe in the simulator.

Q17 der Gesundheit des Fahrers nicht geschadet wird.
the health of the participant is not harmed.

A.2.5 Hardware

Bitte bewerten Sie anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
Using the following questions, please rate whether it is important to you that:

Q18 während des Versuchs so viel wie möglich Strom gespart wird.
as much electricity as possible is saved during the experiment.

Q19 dem Simulator nicht geschadet wird.
the simulator is not damaged.

Q20 wegen des virtuellen Versuchs Treibstoff gespart wird.
fuel is saved by performing the experiment virtually.

Q21 die Wartung des Simulators minimiert wird.
simulator maintenance is minimized.

Q22 der Simulator während der Fahrt wenig Lärm erzeugt.
the simulator generates little noise while moving.

Q23 der Simulator sich nicht unnötig bewegt.
the simulator does not move unnecessarily.

A.2.6 Kosten (*Costs*)

Bitte bewerten Sie anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
Using the following questions, please rate whether it is important to you that:

Q24 nicht zu viel Aufwand investiert wird, den Versuch vorzubereiten.
not too much effort is invested in preparing the experiment.

Q25 die Kosten für das Unternehmen so gering wie möglich sind.
the costs for the organization are as low as possible.

A.2.7 Parametrierung des MCA (*Parametrization of the MCA*)

Bitte bewerten Sie anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
Using the following questions, please rate whether it is important to you that:

Q26 es möglichst wenig Parameter gibt.
there are as few tuneable parameters as possible.

Q27 die Parameter physisch interpretierbar sind.
the tuneable parameters can be interpreted physically.

Q28 der MCA und die dazugehörigen Parameter sich linear verhalten.
the MCA and the accompanying parameters behave linearly.

Q29 es für jede Achse nur einen Parameter gibt.
there is only one parameter for each axis.

Q30 man nicht immer aufsuchen muss, was bestimmte Parameter bedeuten.
one doesn't constantly have to search for the meaning of certain parameters.

Q31 es in jedem Freiheitsgrad des Simulators so wenig Parameter wie möglich gibt.
there are as few parameters as possible in each DoF of the simulator.

Q32 der Stand eines vorherigen Versuches benutzt wird.
the tuning of a previously performed experiment is used.

Q33 das Tuning ohne Simulator ausgetestet werden kann.
the tuning can be tested without a simulator.

A.2.8 Modelstruktur des MCA (*Model structure of the MCA*)

Bitte bewerten Sie anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
Using the following questions, please rate whether it is important to you that:

Q34 der MCA kein Black-Boxsystem ist.
the MCA is not a black-box system.

Q35 ein MCA benutzt wird, der immer konsistent den gleichen Output zu einem bestimmten Input ergibt.
an MCA is used, that consistently gives the same output for a given input.

Q36 man genau sehen kann, was unter der Haube des MCA passiert.
one can clearly see what happens under the hood of the MCA.

Q37 ein MCA benutzt wird, der modular tauschbar ist mit anderen MCAs.
an MCA is used that is modularly interchangeable with other MCAs.

A.2.9 Bedienung des MCA (*Operating the MCA*)

Bitte bewerten Sie anhand der nachfolgenden Fragen, ob es für Sie wichtig ist, dass:
Using the following questions, please rate whether it is important to you that:

Q38 ein stabiler MCA benutzt wird.
a stable MCA is used.

Q39 der MCA leicht zu bedienen ist.
the MCA is easy to use.

Q40 keine unklaren Fehler in der Simulationsbedienungsoberfläche erzeugt werden.
no unclear errors in the simulation control interface are produced.

Q41 alle Optionen in der Simulationsbedienungsoberfläche selbstverständlich sind.
all options in the simulation control interface are self-explanatory.

B

Briefings

The following briefing (in German) was used for the urban data collection experiment, as described in Chapter 4. The corresponding translations to English (not visible for the participants) are indicated in italic. In the other continuous rating experiments, similar briefings were used, although adapted to represent the respective experiments.

Herzlich willkommen zum Driver-in-Motion-Fahrsimulator in Garching! Heute gibt es für Sie eine einzigartige Möglichkeit eine virtuelle Umgebung zu erfahren in einem Bewegungssimulator und damit unsere Forschungsthemen zu unterstützen. In ungefähr einer Stunde und 30 Minuten werden Sie erfahren, wie die Bewegung eines Simulators auf verschiedene Weise gestaltet werden kann. Viel Spaß!

Welcome to the Driver-in-Motion Driving Simulator in Garching! Today you will get a unique opportunity to experience a virtual environment in a motion simulator and support our research and development. In roughly one hour and 45 minutes you will experience how the motion of a simulator can be generated in various ways. Have fun!

Beschreibung der Aufgabe / *Description of Assignment*

In diesem Versuch werden Sie mehrmals kurze Strecken in der Stadt fahren. Die Systeme für Hochautomatisiertes Fahren sind ständig eingeschaltet. Das bedeutet, dass Sie nur Passagier sind. Das Fahrzeug wird eine vorbestimmte Strecke fahren.

Während der Fahrt wird der Simulator sich bewegen, um eine realitätsnahe Bewegung darzustellen. Weil der Bewegungsraum des Simulators kleiner ist als bei einem echten Fahrzeug, passt die Bewegung des Simulators bei bestimmten Manövern nicht genau zu dem, was Sie in der Realität erfahren würden.

Ihre Aufgabe ist es, zu bewerten wie gut diese Bewegung passt zu dem, was Sie von einem echten Fahrzeug erwarten würden. Da Sie nicht selbst fahren, können Sie sich völlig auf die Bewegung konzentrieren.

Es geht nur darum wie gut die Bewegung ist, also **nicht** um die Qualität der Visualisierung, **nicht** um die Qualität des hochautomatisierten Fahrens, den Fahrstil oder etwas anderes!

In this study you will drive multiple short runs through a city. The autonomous driving systems are continuously running. This means you are a passenger. The vehicle will drive a pre-programmed route. This route is indicated by arrows.

During the runs the simulator will move around to provide a realistic movement. As the motion space of the simulator is smaller than that of a real vehicle, the movement of the simulator might, in some cases, not exactly fit to what you would experience in a real vehicle.

It is your task to rate how good this movement fits to that what you would expect from a real vehicle. As you will not be driving yourself, you can fully focus on the movement.

*This task is only about the quality of the movement, so **not** about the quality of the visuals, **not** about the quality of the autonomous driving, the driving style, or anything else!*

Bewertungsmethodik / *Rating method*

Während der Fahrt können Sie mit dem iDrive-Controller auf der mittleren Konsole (Figure B.1) kontinuierlich eine Bewertung zu der Qualität der Bewegung geben. Die Bewertung ist auf einer Skala von 0 bis 10. Eine 0 bedeutet, dass nach Ihrer Meinung die Bewegung sich in diesem Moment genauso anfühlt wie in einem echten Fahrzeug. Eine 10 bedeutet eine sehr schlechte oder unrealistische Bewegung des Simulators.

During the drive you can use the iDrive-Controller on the center console (Figure B.1) to give a continuous rating of the movement quality. The rating is on a scale from 0 to 10. A 0 means, that according to you, the movement at this point exactly feels like what you would feel in a real vehicle. A 10 means an extremely bad or unrealistic movement of the simulator.

Figure B.1: Der iDrive-Controller / *The iDrive-Controller*

Als Beispiel: Wenn das Auto sich überhaupt nicht bewegt, und der Simulator sich auch nicht, ist was Sie im Simulator erfahren genau das, was Sie in einem echten Auto erleben würden. Damit wird erwartet, dass Sie hier eine Bewertung von ’0’ angeben.

Beispiele von “schlechter” Simulatorbewegung sind zu starke oder zu schwache Bewegung, Bewegung in die falsche Richtung oder fehlende Bewegung. Hier wird eine Bewertung erwartet, die auf jeden Fall nicht ’0’ ist.

Die von Ihnen ausgegebene Bewertung ist auf dem Bildschirm sichtbar und wird mit einem Balken angezeigt (Figure B.2).

As an example: When the car is not moving at all, and neither is the simulator, what you are experiencing in the simulator is the same as what you would experience in real car. Therefore, it is expected you give a rating of '0' here.

Examples of "bad" simulator movement are too strong or weak movement, movement in the wrong direction or missing movement. Here, a rating is expected that is at least not '0'.

The rating you are currently giving is visible on the screen and is displayed with a bar (Figure B.2).

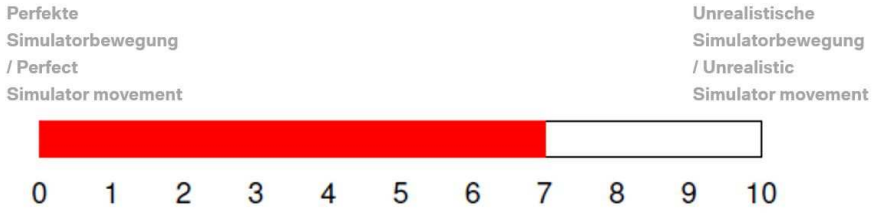


Figure B.2: Der Bewertungsbalken, in diesem Fall mit einer ziemlich schlechten Simulatorbewegung. / *The rating bar, in this case representing a quite bad simulator movement.*

Die Bewertung ist vor allem subjektiv. Das heißt, es gibt kein Richtig oder Falsch.

Nach jeder Fahrt werden Sie nach einem Wert als Gesamtbewertung für die Fahrt gefragt, auf der gleichen Skala wie bei der kontinuierlichen Bewertung.

The rating is above all subjective. That means, there is no right or wrong.

After each run you will also be asked for a single score representing the movement quality of the whole run, on the same scale as the continuous rating.

Ihr Wohlbefinden und Bewegungskrankheit / *Your well-being and motion sickness*

Für Ihre Sicherheit und Wohlbefinden wird Ihr Zustand immer überwacht und mit einem Fragebogen (Table B.1) verfolgt. Nach jeder Fahrt wird gefragt wie Sie sich fühlen und welche Nummer am besten zutrifft.

Wenn es dazu kommt, dass Sie sich während einer Fahrt unwohl oder übel fühlen, informieren Sie bitte sofort den Versuchsleiter. Ihr Wohlbefinden hat immer die höchste Priorität. Der Versuch kann jederzeit gestoppt werden.

For your safety and well-being, your condition is monitored using a questionnaire (Tab. B.1). After each run you will be asked how you are feeling and which value best represents your current condition.

If you are feeling uncomfortable or nauseous, please inform the experiment leader immediately. Your well-being always has the highest priority. The experiment can be stopped at all times.

Table B.1: Die Skala für Ihr Wohlbefinden. *The rating for your well-being.*

Symptom / *Symptom*		**Wert / *Score***
Keine Probleme / *No problems*		0
Etwas unwohl (keine eindeutigen Symptome) / *Slight discomfort but no specific symptoms*		1
Unwohlsein aber keine Übelkeit:	sehr gering / *vague*	2
(z.B. Gähnen, Müdigkeit, kalt/heiß, schwitzen,	gering / *some*	3
verschwommene Sicht, Kopfschmerzen,	mittel / *medium*	4
Speichelfluss, Reizung im Magen/Hals, aufstoßen) / *Dizziness, warm, headache, stomach awareness, sweating, etc.*	stark / *severe*	5
Übelkeit /	gering / *some*	6
Nausea	mittel / *medium*	7
	stark / *severe*	8
	würgen / *retching*	9

Maßnahmen bezüglich der COVID-19-Pandemie / *Measures due to the COVID-19 pandemic*

Um Ihre Gesundheit gegen COVID-19 zu schützen, gibt es zusätzliche Vorschriften während des Versuchs. Sie sind verpflichtet eine Maske zu tragen. Im Simulator dürfen Sie die Maske runterschieben, aber nehmen Sie bitte die Maske nicht ab! Die Simulatorkabine wird immer vor und nach dem Versuch desinfiziert.

To protect your health against COVID-19, there are additional measures during the experiment. It is mandatory to wear a face mask. Inside the simulator you can move your mask down, but please do not take it off completely! The simulator cabin is disinfected before and after the experiment.

Probefahrt / *Test run*

Bevor mit dem tatsächlichen Versuch angefangen wird, gibt es die Möglichkeit eine Probefahrt zu machen, um sich an den Simulator, die Umgebung und die Aufgabe zu gewöhnen.

Before the actual experiment starts, you will be given the opportunity to do a practice run to get acquainted to the simulator, the environment and the task.

Zeitplanung / *Planning*

Der komplette Versuch dauert ungefähr eine Stunde und 30 Minuten. Insgesamt gibt es neun Fahrten. Nach jeder Fahrt geben Sie Ihre Werte für den allgemeinen Eindruck und für Ihr Wohlbefinden. Nach drei Fahrten gibt es eine Pause.

The whole experiment lasts roughly one hour and 45 minutes. In total there will be six/nine runs. After each run you give your value for the overall impression and a value for your well-being. After each set of three runs there will be a short break.

Datensicherheit und Ihre Rechte / *Data security and your rights*

Rechte und Vorschriften bezüglich Datensicherheit sind beschrieben in der „Informierte Datenschutzrechtliche Einwilligungserklärung“.

Ihre Teilnahme ist freiwillig. Sie dürfen jederzeit die Studie verlassen, auch ohne Angabe von Gründen.

Rights and regulations regarding data security are described in the "Declaration of Informed Consent".

Your participation is voluntary. You can leave the experiment at any time, also without specifying a reason.

Fragebogen zu Ihnen als Fahrer / *Questionnaire regarding you as a driver*

Was ist Ihr Alter?
Was ist Ihr Geschlecht?
Seit wann besitzen Sie Ihren Fuhrerschein? [Jahr]
Wie häufig fahren Sie mit dem Auto? [km pro Jahr]
Hatten Sie schon mal Erfahrung mit Fahrsimulation? [J/N]

What is your age?
What is your gender?
Since when do you possess your driver's license? [Year]
How often do you drive a car? [km per year]
Do you have experience with driving simulation? [Y/N]

C

Incongruence Ratings in Dynamic Driving

Abstract

This appendix describes a continuous rating experiment performed to investigate highly dynamic driving on a race track scenario. Nine expert drivers were subjected to two open-loop conditions (i.e., they were being driven), comparing a classical washout algorithm and an optimization-based algorithm with full knowledge of the future states. Participants rated the motion in these conditions continuously and using an overall rating. In a third condition, using only the classical washout algorithm, they drove themselves (closed-loop), in which a novel maneuver-based rating method was employed. Whereas all three ratings methods provided reliable results, the maneuver-based rating method is not useful for high-performance driving due to the high induced workload. Generally, the results show a successful application of both a high-performance simulator and motion cueing for highly dynamic driving, as well as their validation through novel rating methods.

C.1 Introduction

HIGHLY dynamic driving forms an important part of BMW's driving simulation use-cases. In real-life, this testing is performed on dedicated test tracks, which allow for near-limit testing of the vehicle. Such testing is only performed by experienced and specially trained drivers who can operate the vehicle at its limit. So far, the chapters in this dissertation focused on the motion cueing for "customer studies". This refers to studies with everyday drivers in mind, such as to evaluate a novel navigation system.

For highly dynamic driving, the focus is often more on the handling characteristics of vehicles [Salisbury and Limebeer, 2017]. In terms of the motion cueing, the stronger dynamics induce larger accelerations [Salisbury and Limebeer, 2014] compared to regular driving, which are more difficult to reproduce by the simulator. A lack of accurate motion cueing can negatively affect driving performance [Schwarzhuber et al., 2020]. Therefore, specifically designed cueing strategies exist [Salisbury and Limebeer, 2014; Salisbury and Limebeer, 2017; Schwarzhuber et al., 2021] that differ from strategies used in motion cueing for customer studies. This is done by leveraging the known geometry of the test track, or by implementations of side slip reproduction [Bruschetta et al., 2018b].

Similar as for motion cueing for customer studies (e.g., Chapter 4), accurate evaluation tools of motion cueing for highly dynamic driving would be beneficial. These would reduce the required on-site testing, saving time and money. Motion cueing for highly dynamic driving, however, which requires high platform accelerations and large workspaces, has not yet been tested on BMW's motion systems. It is furthermore currently unknown whether the continuous rating method introduced in Cleij et al. [2018] and Chapter 4 can be used in highly dynamic driving simulations. For example, the intensity of the motion in such simulations might be too demanding to also, at the same time, evaluate the motion and perform a rating task, even in open-loop driving.

This appendix describes a highly dynamic driving study on a race track in which expert drivers evaluated the motion in three conditions: open-loop "Oracle" motion cueing [Ellensohn et al., 2019c], open-loop CWA, and closed-loop CWA. The goal of the experiment was to assess the viability of using open-loop driving and the continuous rating method for highly dynamic use-cases. In the open-loop conditions, participants were driven around using a pre-recorded drive on the circuit. In these conditions, the continuous rating method using a rating knob was employed [Cleij et al., 2018]. Furthermore, the viability of a maneuver-based rating method was investigated in the closed-loop condition, in which participants could drive themselves. Finally, the highly demanding use-case induced strong platform motion and was therefore used as a validation in terms of the motion reproduction capabilities of BMW's largest driving simulator.

This appendix is structured as follows. The applied methods are discussed in Section C.2. The experiment set-up is discussed Section C.3, followed by the results in Section C.4 and the discussion in Section C.5. The appendix is concluded in Section C.6.

C.2 Methods

In the experiment, the participants' task was to rate the physical motion of the simulation by evaluating the Perceived Motion Incongruence (PMI). The PMI, however, is not measurable, as it is internal to the human operator. Therefore, a Motion Incongruence

Rating (MIR) was asked from the drivers, ranging from values of 0 (congruent motion) to 10 (highly incongruent motion), with steps of 1, as proposed in Cleij et al. [2018]. Three types of rating methods were applied, similar as in Chapter 7.

C.2.1 Overall Post-Hoc Ratings

Overall ratings were extracted at the end of the drive (post-hoc) and aim to represent the overall impression of the whole drive. These ratings are straightforward in their use and their recordings OR_{PH} provide a direct answer as to which of various motion cueing conditions is preferred, as only a single value is obtained per drive. However, they do not provide information on *where* in the scenario the incongruences occurred. Overall ratings were recorded in both the open- and closed-loop drives.

C.2.2 Maneuver-Based Ratings

Ratings were also obtained for distinct maneuvers [Ellensohn et al., 2020]. These were directly asked after each maneuver, i.e., Section-Wise Post-Hoc (SPH) (see Chapter 2). Maneuver-based ratings can be acquired either by verbal communication between the participant and the experiment leader or through a rating interface. However, due to the demanding requirements of operating the vehicle at high speed around the race track, operating a rating interface as a secondary task might jeopardize safety, as it requires temporary steering with only one hand. For this experiment, it was therefore chosen to obtain the maneuver-based ratings verbally. These ratings were only collected in the closed-loop driving condition.

C.2.3 Continuous Ratings

In the continuous rating task, drivers continuously rated the quality of the motion using a rating interface [Cleij et al., 2018], reflecting their current PMI at each point in time. The resulting rating signals are of high temporal resolution, which is especially beneficial for modeling approaches [Cleij et al., 2018; Ellensohn et al., 2019c; Kolff et al., 2022; Kolff et al., 2024b]. As explained in Chapter 4, the high workload while performing this task does not allow for simultaneously driving the vehicle. Therefore, this rating method can only be employed while “driving” open-loop, i.e., using a playback. The continuous rating task has been established in customer studies (Part II), but has never been tested in a highly dynamic simulation. Especially the requirement on open-loop driving might be problematic, as it takes away the expectation of motion based on the drivers' control actions (See Chapter 7), an aspect that many expert drivers use in the evaluation of motion in highly dynamic and near-limit driving.

C.3 Experiment Set-up

C.3.1 Scenario

The scenario considered in the experiment was the virtual equivalent of Circuit de Miramas. It consists of a single lap of 3.3 km. The simulated version (Figure C.1) is of equal dimensions. Each simulation consisted of the full first lap and the beginning of the second lap. For the maneuver-based rating, the track was divided into twelve maneuvers, shown

to the participants by green patches on the road surface. The maneuvers are shown in Table C.1, including the most dominant motion channels. These are also shown in Figure C.1. As the beginning of the first lap was repeated, maneuvers M11 and M12 (in lap 2) corresponded to the maneuvers M1 and M2 (in lap 1), respectively. The total length of the drive was 3.9 km.

Table C.1: Overview of the maneuvers to be rated using the maneuver-based rating method.

Maneuver	Description	Dominant channel
M1	Initial acceleration	Longitudinal
M2	Slalom	Lateral & Yaw
M3	Corner right	Lateral & Yaw
M4	Corner right	Lateral & Yaw
M5	Acceleration	Longitudinal
M6	Braking	Longitudinal
M7	Corner left	Lateral & Yaw
M8	Straight	None
M9	Corner right	Lateral & Yaw
M10	Straight	None
M11	Bump	Vertical
M12	Braking	Longitudinal

C.3.2 Apparatus

The experiment was performed on BMW's Sapphire Space simulator (Figure C.2a). The motion system has nine DoFs, consisting of three subsystems: a 19.14 m×15.70 m xy-drive, a 1.15 m stroke hexapod, and a 360° yaw-drive. The simulator is fully enclosed by a dome, in which a BMW 5 series (G30) vehicle mockup was placed. The steering wheel rotated corresponding to the simulated drive.

The iDrive navigation knob on the center console (Figure C.2b) was used as the rating interface by the drivers to provide the continuous rating $R(t)$. Inside the dome, visuals were rendered using Unreal Engine (visible in Figure C.2b) and displayed using 12 Norxe P1 projectors, resulting in a highly realistic full 360° projection in the dome. The current rating value in the form of a "rating bar" was visible in the central field of view in a type of head-up display. The size and color of this rating bar changed from rating 0 (short, white) to rating 10 (long, red). The complete rating method was thus the same as the experiments described in Part II of this dissertation. The velocity of the vehicle was shown on the tachometer on the dashboard and in a head-up display alike screen projection.

C.3.3 Independent Variables

The experiment had two independent variables: the driving method (levels: open-loop and closed-loop) and the MCA (levels: Oracle and CWA). The tuning parameters of the CWA in the closed- and open-loop conditions were the same. The Oracle motion cueing used an optimization-based scheme using perfect prediction capabilities (only possible in open-loop driving) [Ellensohn et al., 2019c]. Both the implementations of the Oracle

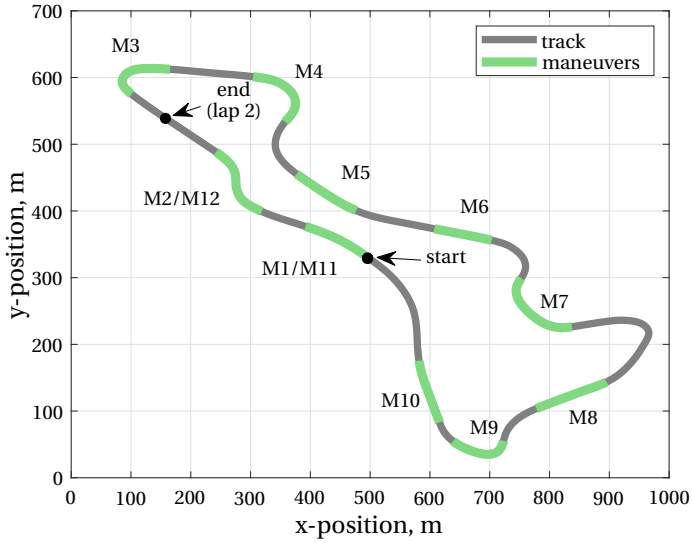


Figure C.1: Overview of the track (grey) and the maneuvers to be rated in the CWA-CL condition (green overlay).

Table C.2: Experimental conditions with the two independent variables (motion cueing and driving method), including the rating methods used in each condition.

Condition	ORC	CWA-OL	CWA-CL
Motion cueing	Oracle	CWA	CWA
Driving method	open-loop	open-loop	closed-loop
Rating method	$R(t)$, OR_{PH}	$R(t)$, OR_{PH}	MB_{SPH}, OR_{PH}

and CWA were similar to those tested in Chapter 4. As Oracle motion cueing is only possible in open-loop simulations, three experiment conditions were tested. These are shown in Table C.2. The maneuvers to be rated in the closed-loop CWA condition were chosen as the maneuvers of the drive where either much longitudinal or lateral motion occurs. For M8 and M10 the opposite was true, as in these maneuvers little motion was present. These should thus be rated with low ratings (i.e., congruent motion), as the mismatches in these maneuvers are small, see Figure C.3. These maneuvers thus serve as a check for the responsiveness of the participants and the validity of the rating method.

Similarly, for the closed-loop drives, the mismatches are shown in Figure C.4. Because the closed-loop drives are inherently different due to differences in velocity, the covered distance is a better measure to compare the motion between the drives. Therefore, the along-track distance to a point n is calculated [Kolff et al., 2019] as:

$$a_n = \sum_{i=1}^{n} \sqrt{\Delta x_i^2 + \Delta y_i^2}, \tag{C.1}$$

where i represents the elements up until the point n. Δx_i and Δy_i, respectively, rep-

(a) The Sapphire Space.

(b) A participant using the rating knob.

Figure C.2: The experiment set-up.

resent the increase in x- and y-position at point i with respect to the previous point. Figures C.4c and C.4f clearly show the deterministic nature of the f_y and ω_z channels, respectively, as drivers have to follow the geometry of the track. In contrast, the longitudinal specific force channel f_x is less deterministic (and hence more spread-out), as the acceleration and deceleration points can be chosen more freely by the drivers.

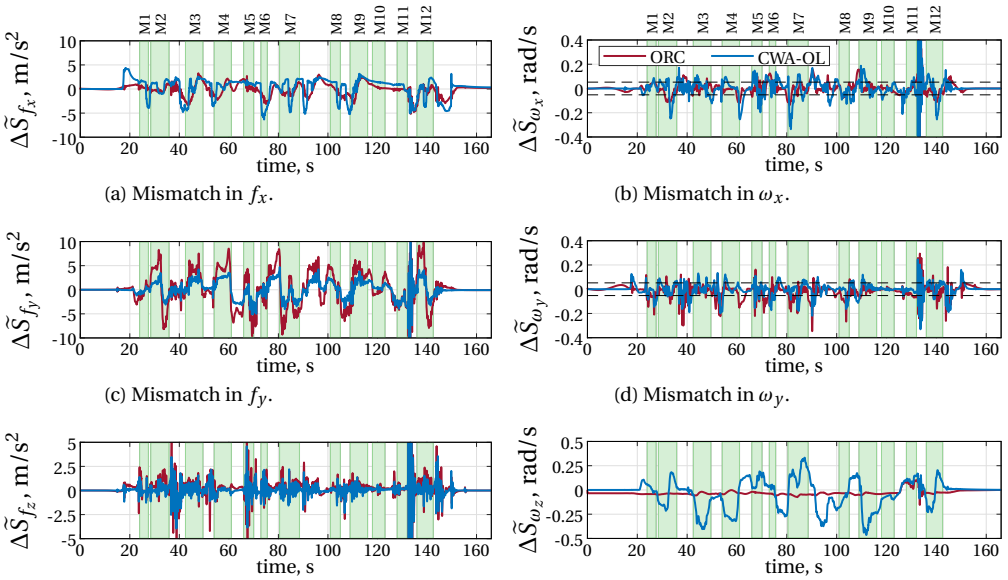
C.3.4 Participants and Procedures

Nine participants took part in the experiment. All of them were in the possession of a European car driver's license B for at least five years ($M = 20.6$ yrs, $SD = 9.1$ yrs). All of them were expert drivers with extensive experience in vehicle testing and had BMW driving academy training on advanced vehicle handling and prototype driving. The average yearly driven distance for the participants was $M = 27,100$ km ($SD = 14,410$ km). The average age was $M = 37.7$ yrs ($SD = 9.2$ yrs). All of them had previous experience in driving simulators. Five participants had driving experience on the real-life Circuit de Miramas track. All participants provided informed consent and the experiment was approved following BMW's internal ethics review procedures.

The experiment consisted of ten drives, which were grouped into three blocks. In the first block, three open-loop drives were driven, being either ORC-CWA-ORC or CWA-ORC-CWA, depending on the participant. In the second block, the four closed-loop drives were performed. In the third block, the opposite of the first block was performed (CWA-ORC-CWA or ORC-CWA-ORC). After each block a five minute break was taken.

A training phase was conducted before the actual experiment phase. Here, participants were driven one lap of the open-loop CWA condition to get acquainted to the rating set-up. Furthermore, at least two closed-loop laps were driven to get accustomed to the track and the maneuver-based rating method. This phase was extended if participants required more training to stay on the track or to understand the rating task. After this, the measurement phase started.

For the closed-loop drives, drivers were instructed to drive dynamically, but not on the absolute limits of the vehicle, to ensure they stayed on the track. For the maneuver-based ratings (closed-loop condition only), drivers were asked to give their rating of each



(a) Mismatch in f_x.
(b) Mismatch in ω_x.
(c) Mismatch in f_y.
(d) Mismatch in ω_y.
(e) Mismatch in f_z.
(f) Mismatch in ω_z.

Figure C.3: Mismatches of the open-loop conditions. Green areas are the maneuvers; dashed lines in C.3b and C.3d are the perceptual thresholds (±3 deg/s [Reymond and Kemeny, 2000]).

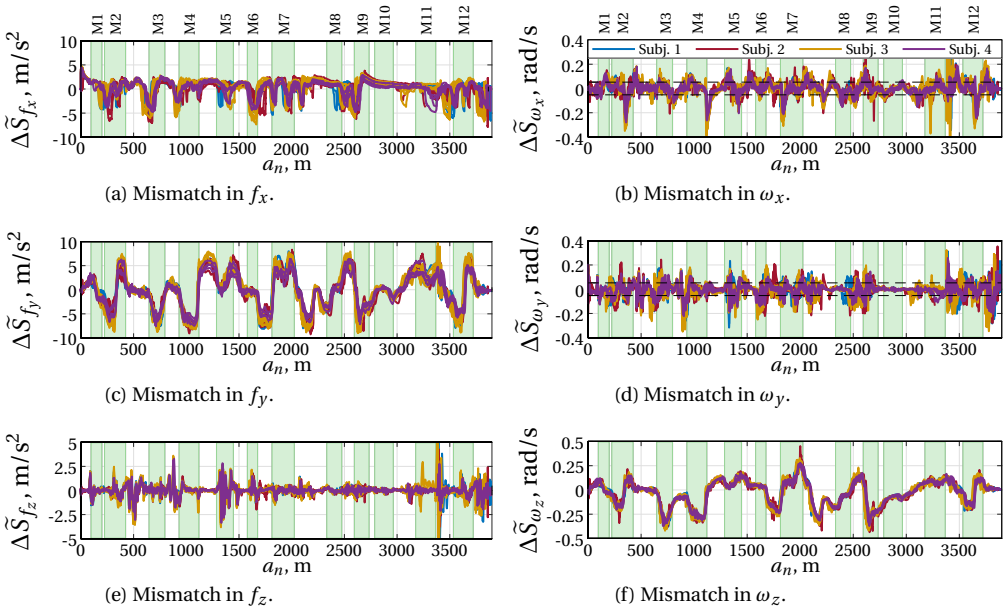
maneuver verbally. The selected maneuvers were spaced such to give drivers enough time to give their rating while still being able to focus on the driving task.

C.4 Results

The experiment had, with ten participants, a relatively small sample size. Due to technical issues caused by the novelty of the system, two participants could not finish the experiment. Furthermore, four participants suffered from simulator sickness that occurred during the experiment. This leaves only four complete data sets for analysis that are considered as subjects. For this reason, the differences between the conditions are only analyzed descriptively and are not subjected to any significance analyses.

C.4.1 Continuous Ratings

The obtained continuous rating signals obtained in the open-loop driving conditions (ORC and CWA-OL) are shown in Figure C.5 (left). The continuous lines represent the mean over all subjects and repetitions, with the standard deviation shown as a shaded area. Note that there are two indications that subjects are able to rate the motion in this dynamic simulation. First, the ratings increase (hence, the motion was evaluated to get worse) at the points where the mismatches occur (i.e., the maneuvers). Second, in those maneuvers (mostly M8 and M10) in which there was little motion (and therefore no mismatches), subjects indeed decrease the rating (most notably at M10). This indicates the responsiveness of the subjects to the mismatches. The individual rating data are included in Appendix D.5.



(a) Mismatch in f_x.

(b) Mismatch in ω_x.

(c) Mismatch in f_y.

(d) Mismatch in ω_y.

(e) Mismatch in f_z.

(f) Mismatch in ω_z.

Figure C.4: Mismatches in the CWA-CL condition as function of the along-track distance. Each colored line represents a single drive. Green areas are the maneuvers; dashed lines in C.4b and C.4d are the perceptual thresholds (±3 deg/s [Reymond and Kemeny, 2000]).

Generally, the Oracle motion cueing (ORC) was rated better (i.e., lower ratings) than CWA-OL. An exception is just before maneuver M8. Participants' comments revealed that some here perceived an unexpected drift (i.e., erroneous yaw motion). This could be due to too aggressive pre-positioning, required to allow for sufficient workspace in the yaw direction for maneuver M9 soon after the 180° turn between M7 and M8.

Figure C.6 shows the reliabilities of the continuous rating data acquired per subject. The reliability is estimated using Cronbach's Alpha and McDonald's Omega, as in Chapter 4. The value $\bar{R}_p$ represents the average continuous rating per participant. The figures also contain the regression fit of α (grey lines) and Ω (black lines) of the form $r = a - 1/(b\bar{R}_p + c)$ acquired on the urban rating data of Chapter 4. For CWA-OL, with the exception of Subject 1 ($\Omega = 0.40$), the reliabilities are high (Subject 2: $\Omega = 0.76$, Subject 3: $\Omega = 0.92$, and Subject 4: $\Omega = 0.95$). For Subjects 1 and 2, in the ORC condition, ($\Omega = 0.45$ and $\Omega = 0.52$, respectively) the reliabilities are lower than Subjects 3 and 4 ($\Omega = 0.85$ and $\Omega = 0.92$, respectively). Subjects 1 and 2 are below the predicted reliabilities using the fit from Chapter 4. A key finding of Chapter 4 was that the reliability decreases for lower ratings (i.e., for a more positive subjective rating of the motion). Although this finding is reproduced here, the reliabilities of Subject 2 and especially Subject 1 are thus lower than predicted, indicating a lower task understanding. As discussed in Chapter 4, and reproduced here, McDonald's Omega also provides a benefit mainly for subjects for whom Cronbach's Alpha is low. Although only a limited number of subjects is available in the present study, these findings thus generally confirm those of Chapter 4.

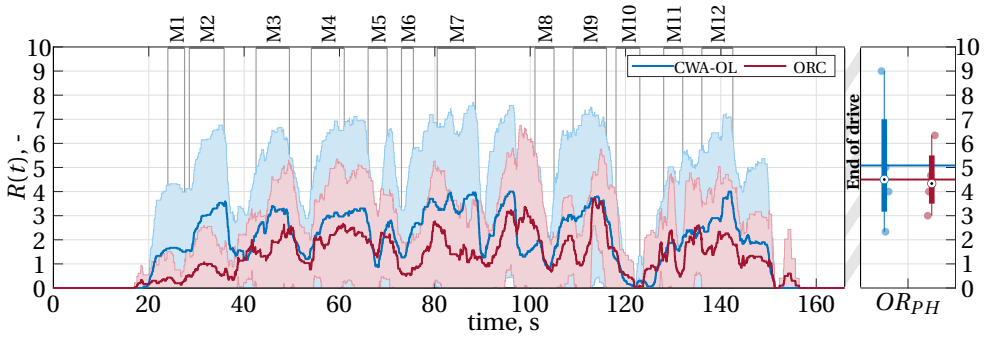
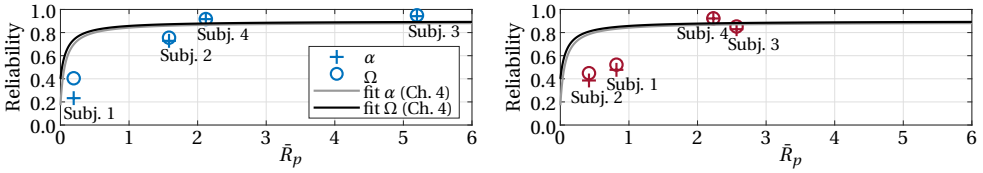


Figure C.5: Left: The averaged MIRs per MCA with the standard deviation displayed as shaded areas. Right: Box plots of the overall ratings; their means are indicated by horizontal lines.



(a) Reliability for condition CWA-OL.

(b) Reliability for condition ORC.

Figure C.6: Reliability for the two experiment conditions of the four subjects.

C.4.2 Overall Ratings

The right side of Figure C.5 shows the overall ratings of the ORC and CWA-OL conditions. The box plots indicate the spread over the recordings. The horizontal bars indicate the means (for ORC: 4.5, for CWA-OL: 5.1), such that ORC is again rated better than CWA-OL. One outlier in the ORC condition (visible in Figure C.5) lies at a rating of 9, which has, considering the limited number of subjects, a strong effect on the distribution of the overall rating. As discussed in Chapter 4, reliability of the overall ratings can only be calculated for the whole group: for ORC: $\Omega = 0.92$ and for CWA-OL: $\Omega = 0.99$.

C.4.3 Maneuver-based Ratings

The maneuver-based ratings of the *closed*-loop condition CWA-CL are shown in the left part of Figure C.7 (left) and displayed as box plots. Similar to what was seen in the continuous ratings, maneuvers M8 and M10 (which did not contain a high degree of motion, and therefore also little mismatches) were indeed also given low ratings. This indicates the subjects' understanding of and responsiveness to the mismatch rating task.

The right part in Figure C.7 displays the overall ratings in the CWA-CL condition, averaged for each participant. Note that the distribution of these ratings is similar to the distribution of the open-loop CWA condition shown in Figure C.5. This indicates that the open-loop and closed-loop were rated similarly. Reliability cannot be calculated here, because all closed-loop runs are inherently different.

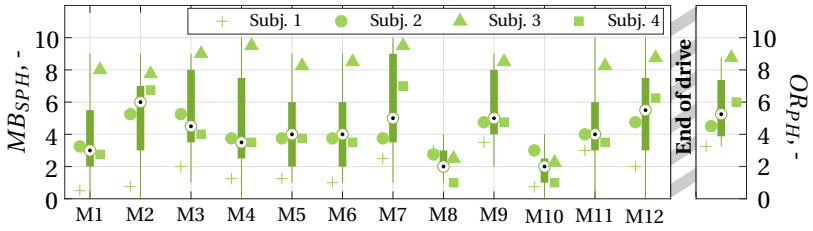


Figure C.7: The maneuver-based (left) and overall (right) ratings given by the subjects in the CWA-CL condition, averaged over the four repetitions.

C.5 Discussion

C.5.1 Motion System Validation

Part of the motivation for the experiment was to serve as a pilot study to test both the motion cueing, as well as the capabilities of the simulator's motion system, under highly dynamic conditions. Although the high degree of motion induced some technical difficulties that caused two drop-outs, generally, the motion system showed to be well capable of reproducing the commanded motion by reaching the required very high accelerations up to 1 g. Furthermore, both MCAs, primarily designed for use in customer studies, generated outputs that were rated as realistic using the presented rating methods.

C.5.2 Rating Measurements

A second consideration was whether the rating methods used for customer studies (Part II) are feasible for highly dynamic studies with expert drivers. The reliability estimates for both the continuous and overall ratings showed this indeed to be the case. Significance analyses were not performed due to the small group of expert drivers available, in combination with technical issues and drop-outs due to simulator sickness.

Considering the measurements of the continuous ratings, the Oracle condition was preferred over the CWA. This is supported by the overall post-hoc ratings. However, the difference between the ratings (both continuous and overall) is smaller than might be expected, considering the large mismatches in the CWA condition (See Figure C.3). One explanation for this is that the Oracle revealed some (inaccurate) aspects in the motion cueing that were not present in the CWA condition. Such inaccuracies could for example come from the vehicle model. For the CWA condition such aspects might not be noticed (due to the filtering), and are therefore also not rated. Another explanation, strengthened by participants' comments, is that the Oracle condition was sometimes perceived as "too strong", confirming findings of Berthoz et al. [2013]. This might especially be an issue for open-loop driving, in which the participants cannot actively control the motion through the vehicle control, given the intensity of the motion.

C.5.3 Lessons for Chapter 7

C.5.3.1 Comparison of Closed-Loop Drives

Allowing participants to drive themselves inherently results in different road positions and velocities, such that recorded signals cannot be compared in the time domain. Sec-

tion C.3.3 proposed to use the along-track distance of the vehicle instead, making signal comparison between the drives possible. This, however, only corrects for differences in velocity, and not in differences in road position. Each lap of the circuit will have slightly different along-track distances. For a less deterministic driving scenario, such as urban driving (where drivers for example have more freedom in the lane they choose to drive on), a different method must be applied. In Chapter 7 it was therefore decided to apply a "nearest-neighbor" approach, which is a more precise, but also more complex approach.

C.5.3.2 Comparison Between Closed- and Open-loop Drives

From the experiment, the explicit research question whether open- and closed-loop driving are rated equivalently has arisen. This comparison cannot be made based on the present experiment, as the rating methods used in the open- and closed-loop conditions were different. The only rating method with overlap was the overall rating method. These ratings were indeed similar (between the open- and closed-loop CWA condition), but form a too limited perspective on equivalence of the rating task and can, again, not be analyzed for significant equivalence. Furthermore, in the present study, there were differences in the drives: in the closed-loop condition participants evaluated their own drives, whereas in the open-loop condition they evaluated a pre-recorded drive. The ratings of these conditions cannot be compared without the bias of varying motions experienced between the drives. This formed a prime motivation of letting drivers experience their own drives both closed- and open-loop in the experiment of Chapter 7.

C.5.3.3 Maneuver-based Rating Extraction

The acquisition of maneuver-based ratings showed several challenges. Maneuvers need to be separated enough from each other to allow the participant to rate without any added workload, and to not let impressions on the maneuvers overlap. Through feedback from the participants, it was also noted that they found the verbal extraction of the maneuver-based ratings confusing. The interactions that followed after each maneuver-based rating were too long and sometimes came close to the next maneuver to be rated.

Using a rating knob (as in Chapter 4) for the maneuver-based ratings would require briefly taking one hand of the steering wheel for each maneuver. In highly dynamic driving, this jeopardizes safety as both hands are necessary to steer the vehicle. The maneuver-based rating method is thus generally not suitable for highly dynamic simulations. For customer studies, recording maneuver-based ratings using a rating knob was shown to be of high value in Chapter 7. By automatically logging the values through a knob, the risk of the experiment leader writing down the wrong value is also reduced. Interactions about the ratings between the participant and the experiment leader, such as clarifications on the rating and rating scale, can still be performed after each drive.

C.6 Conclusion

This appendix described a driving simulator experiment, in which expert drivers evaluated the motion of highly dynamic driving on a race track simulation. In the experiment, the drivers experienced both an optimization-based Oracle (open-loop) and a CWA MCA (both closed- and open-loop). The combination of the large, nine-DoF Sapphire Space simulator in combination with the two MCAs successfully provided a highly dynamic

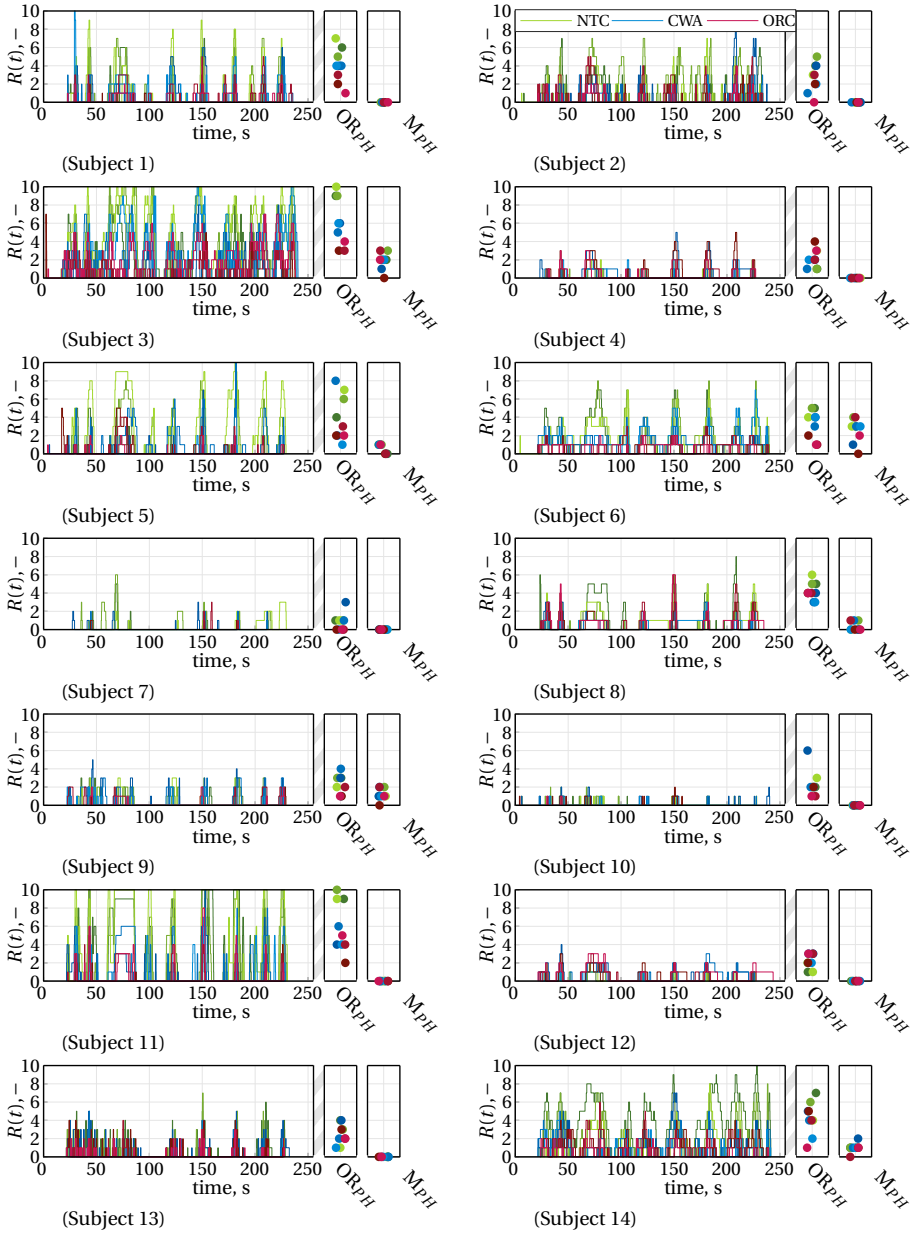
motion cueing experience. Both the continuous and overall rating methods provided reliable results. While maneuver-based ratings provide additional and seemingly reliable insights in the motion cueing of closed-loop driving, using a rating interface is recommended for future experiments. For highly dynamic simulations, the maneuver-based rating method is likely less suitable due to constraints in the workload of drivers. Furthermore, explicit studies investigating the equivalence of closed- and open-loop driving are necessary as conclusions based on the present experiment could not be made. The findings show the validity of the rating methods in highly dynamic driving and provide the basis for a further understanding of ratings for future closed-loop driving simulations.

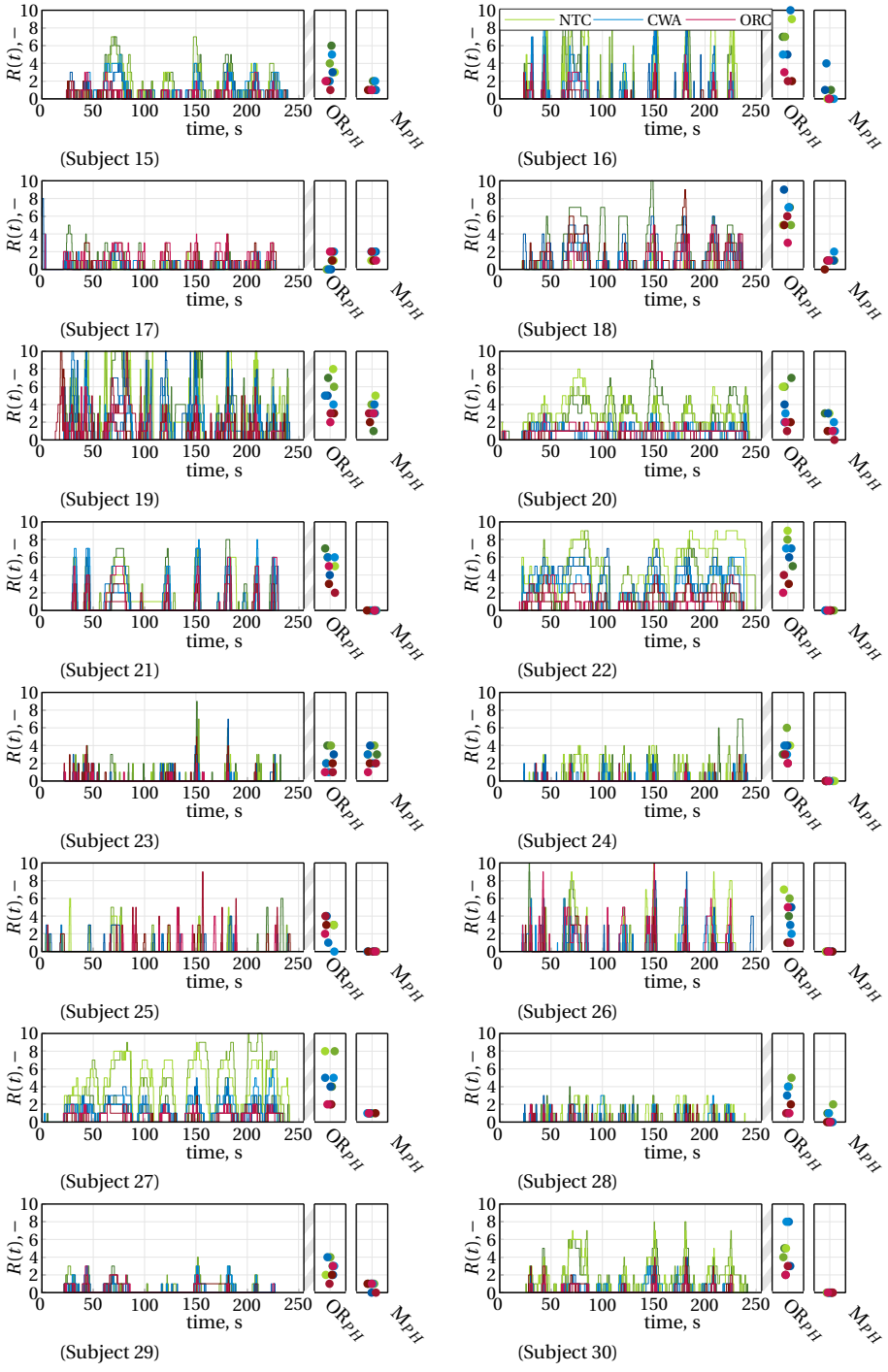
D

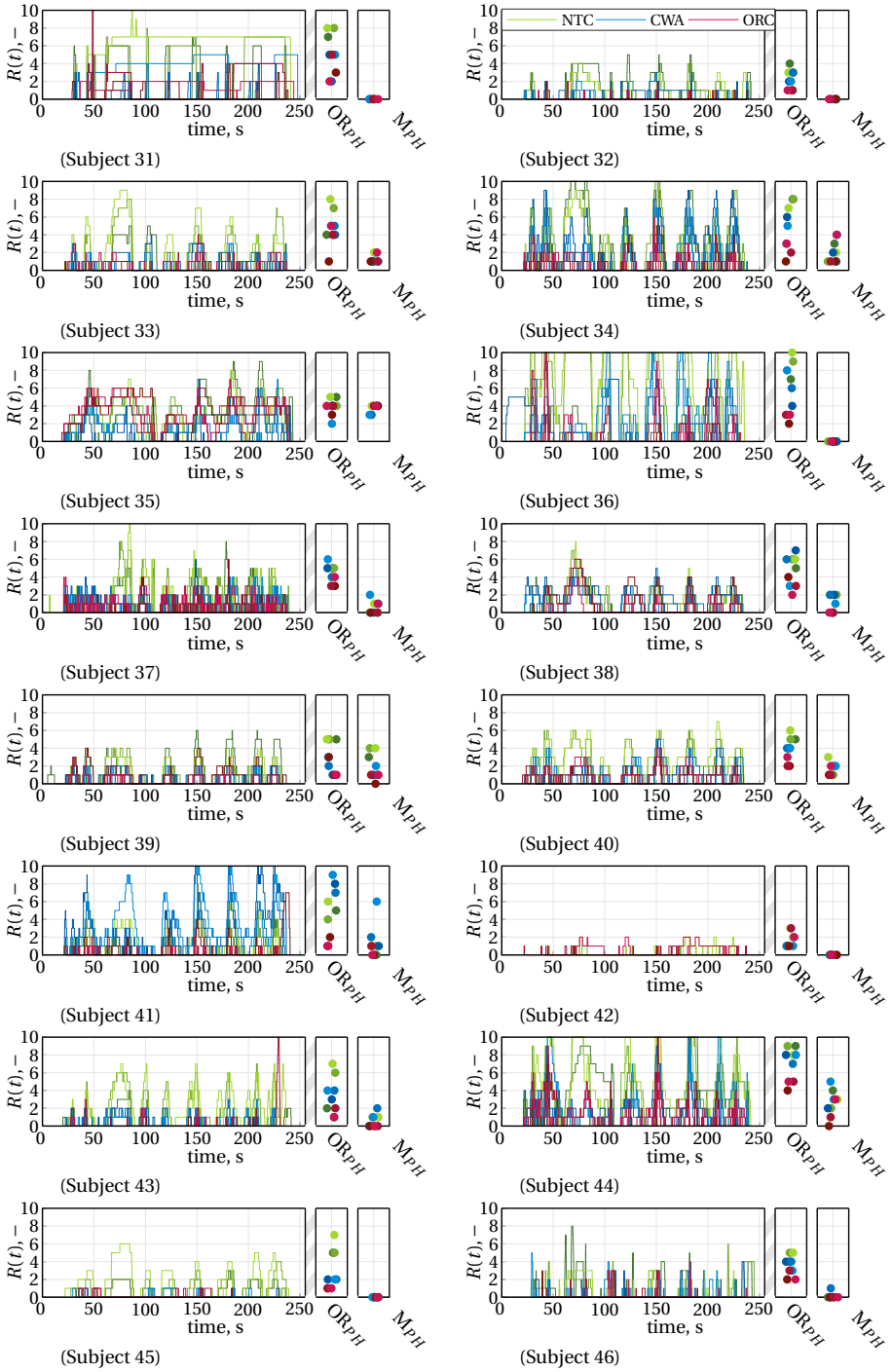
Overview of Individual Ratings

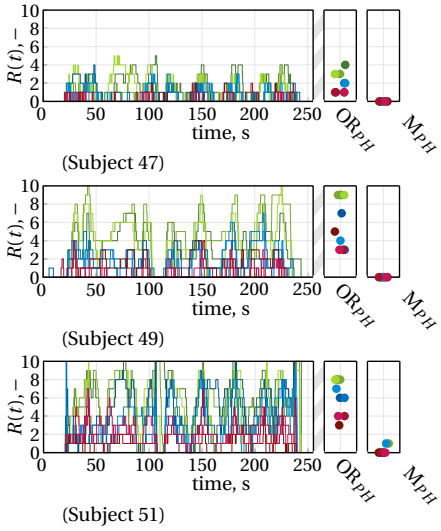
This Appendix contains the individual ratings signals as recorded in the new experiments described in this dissertation. As in the chapters of Part II, the left side of each figure shows the continuous rating signals, whereas the right side shows the overall MIR and MISC ratings. Varying shades of a single color (corresponding to a condition) represent multiple repetitions of that condition.

D.1 Chapter 4 - Incongruences in Urban Simulations

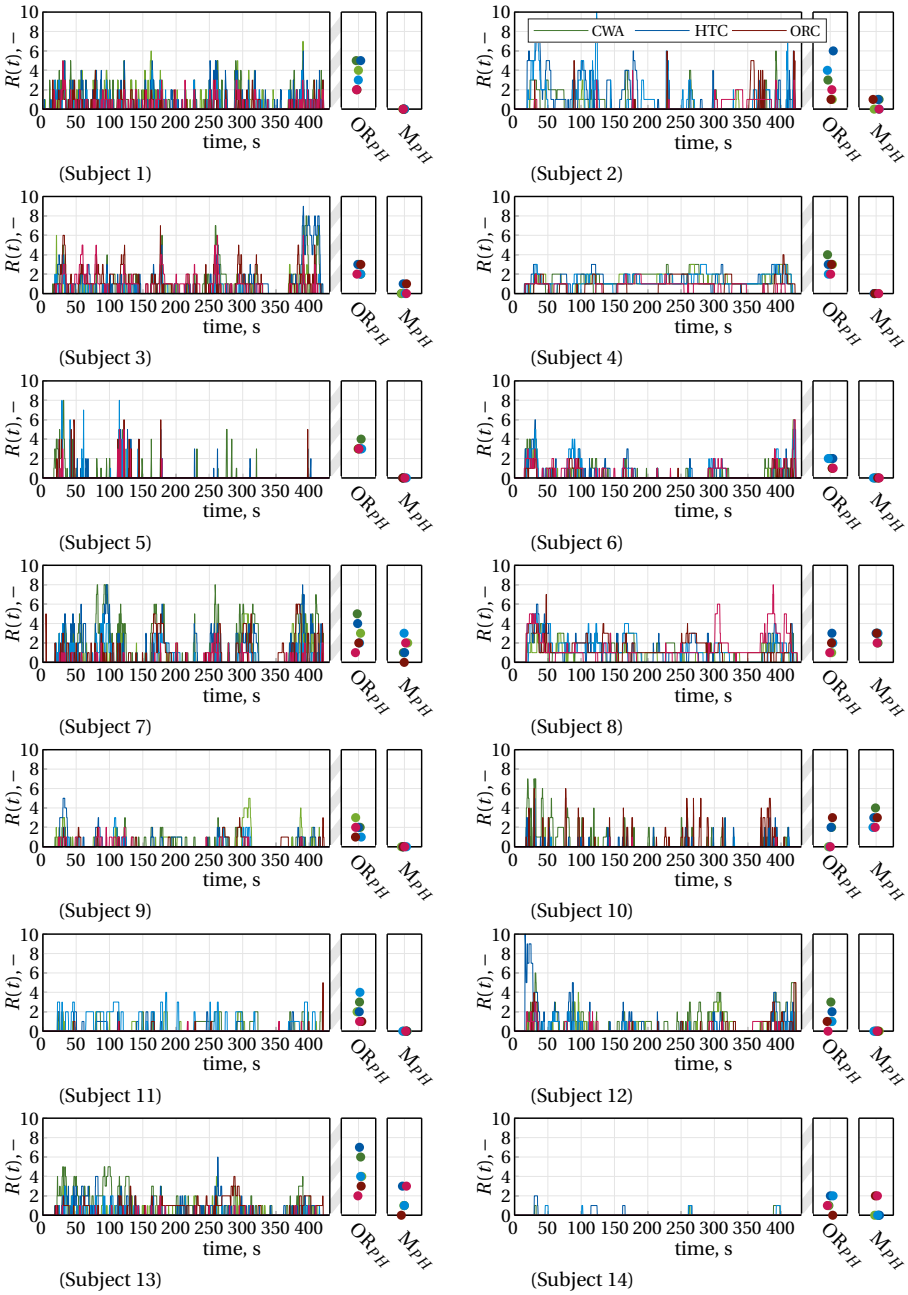


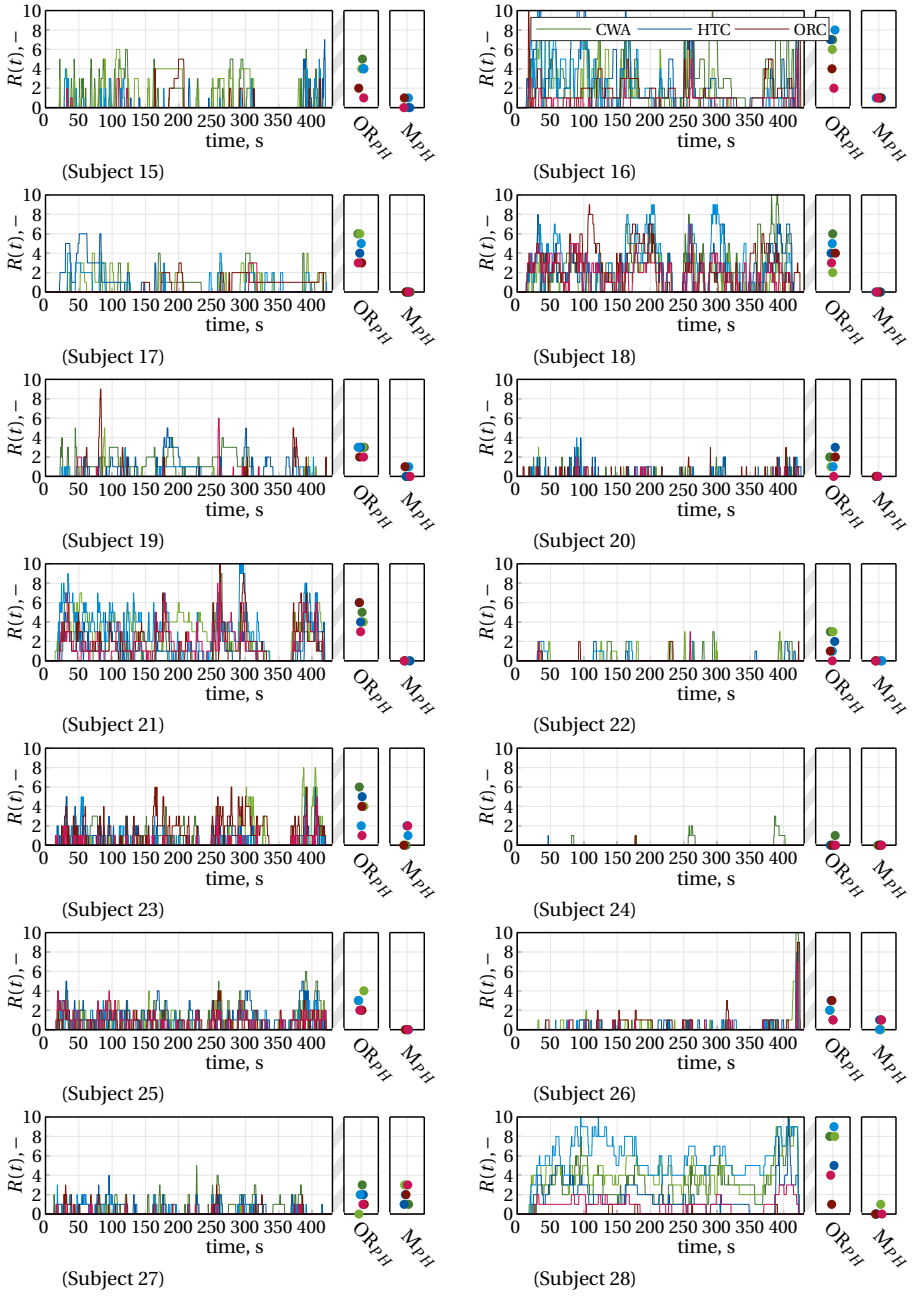

NTC
CWA
ORC
R(t), –
time, s
ORPH
MPH
(Subject 15)
(Subject 16)
(Subject 17)
(Subject 18)
(Subject 19)
(Subject 20)
(Subject 21)
(Subject 22)
(Subject 23)
(Subject 24)
(Subject 25)
(Subject 26)
(Subject 27)
(Subject 28)
(Subject 29)
(Subject 30)


NTC
CWA
ORC
R(t), –
time, s
ORPH
MPH
(Subject 31)
(Subject 32)
(Subject 33)
(Subject 34)
(Subject 35)
(Subject 36)
(Subject 37)
(Subject 38)
(Subject 39)
(Subject 40)
(Subject 41)
(Subject 42)
(Subject 43)
(Subject 44)
(Subject 45)
(Subject 46)


R(t), –
time, s
ORPH
MPH
(Subject 47)
(Subject 48)
(Subject 49)
(Subject 50)
(Subject 51)
NTC
CWA
ORC

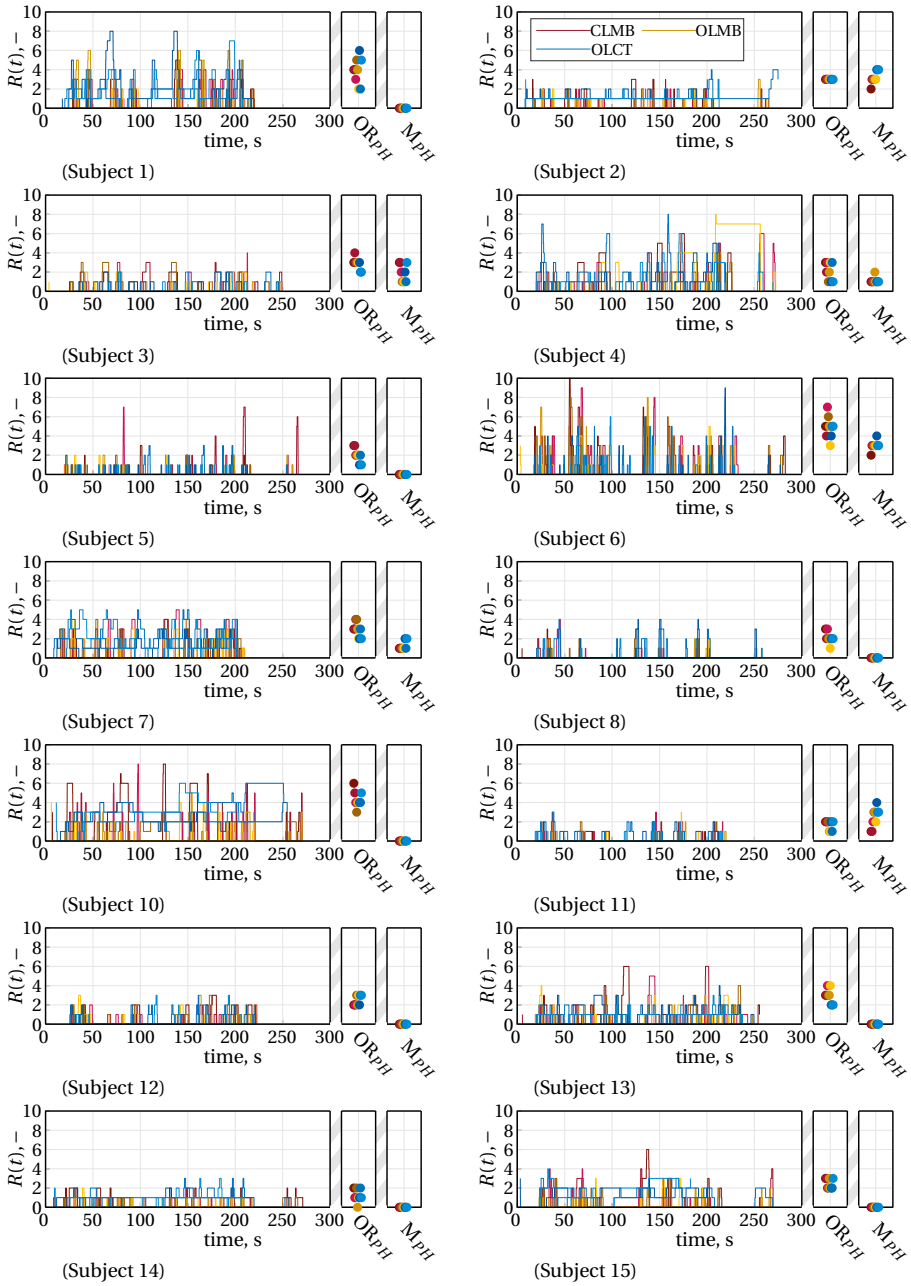
D.2 Chapter 5 - Incongruences in Highway Simulations

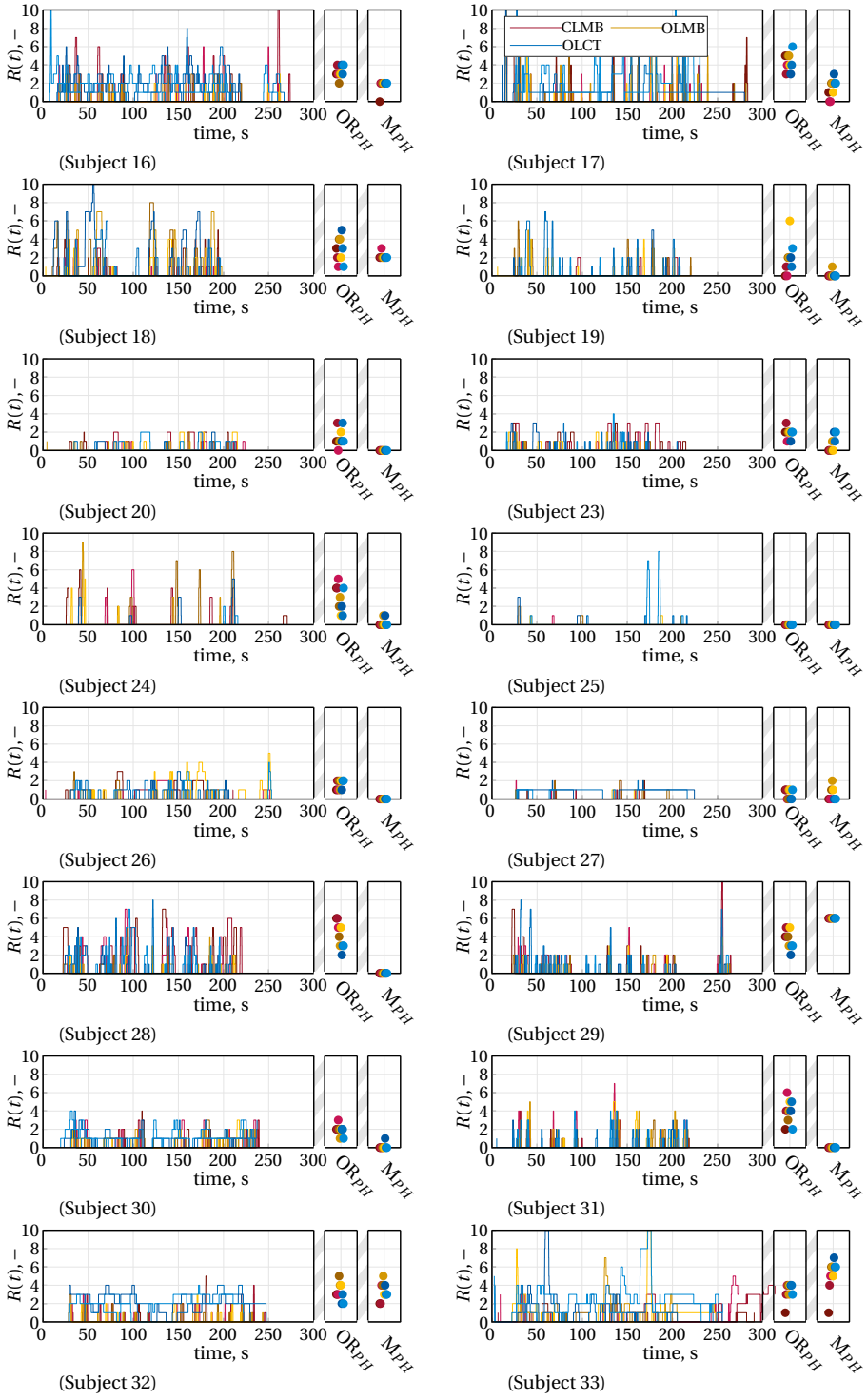


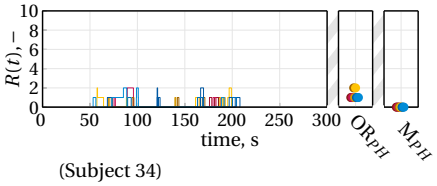

R(t), –
time, s
CWA
HTC
ORC
OR_{PH}
M_{PH}
(Subject 15)
(Subject 16)
(Subject 17)
(Subject 18)
(Subject 19)
(Subject 20)
(Subject 21)
(Subject 22)
(Subject 23)
(Subject 24)
(Subject 25)
(Subject 26)
(Subject 27)
(Subject 28)

D

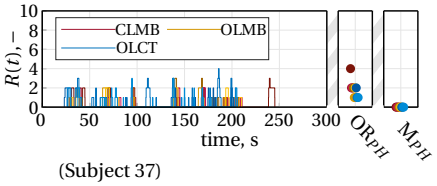
D.3 Chapter 7 - Incongruences in Closed-loop Driving



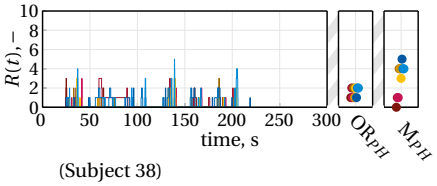

R(t), –
time, s
ORPH
MPH
CLMB
OLMB
OLCT
(Subject 16)
(Subject 17)
(Subject 18)
(Subject 19)
(Subject 20)
(Subject 23)
(Subject 24)
(Subject 25)
(Subject 26)
(Subject 27)
(Subject 28)
(Subject 29)
(Subject 30)
(Subject 31)
(Subject 32)
(Subject 33)



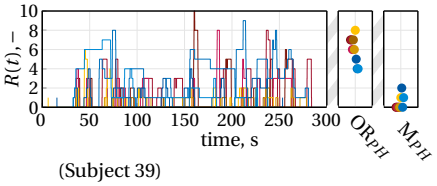
(Subject 34)



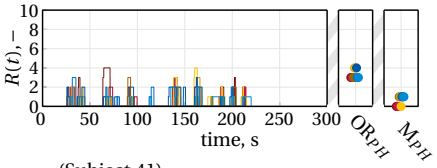
(Subject 37)



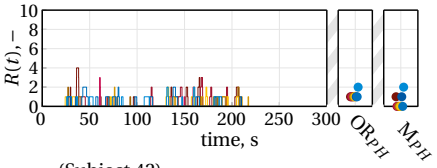
(Subject 38)



(Subject 39)

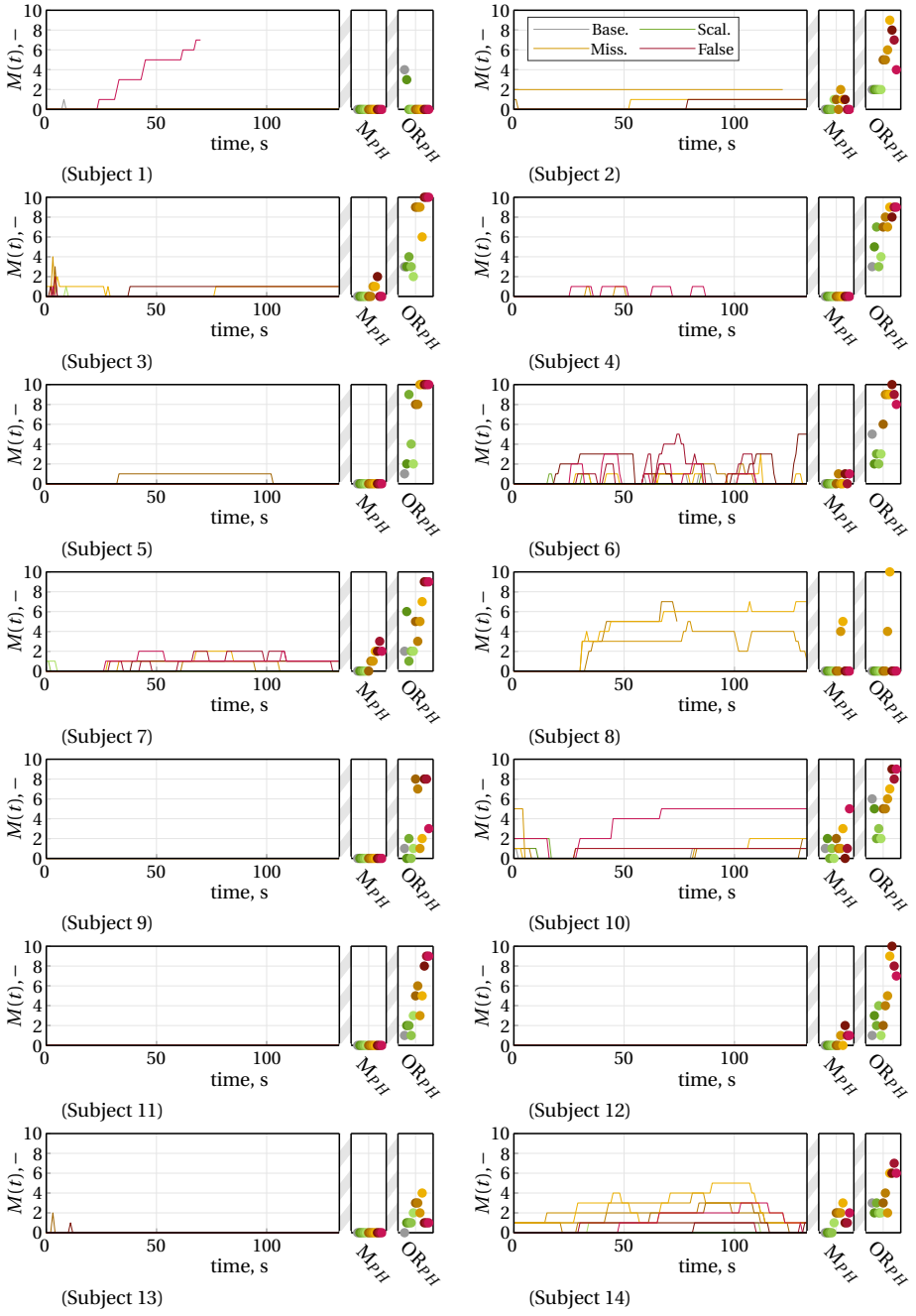


(Subject 41)

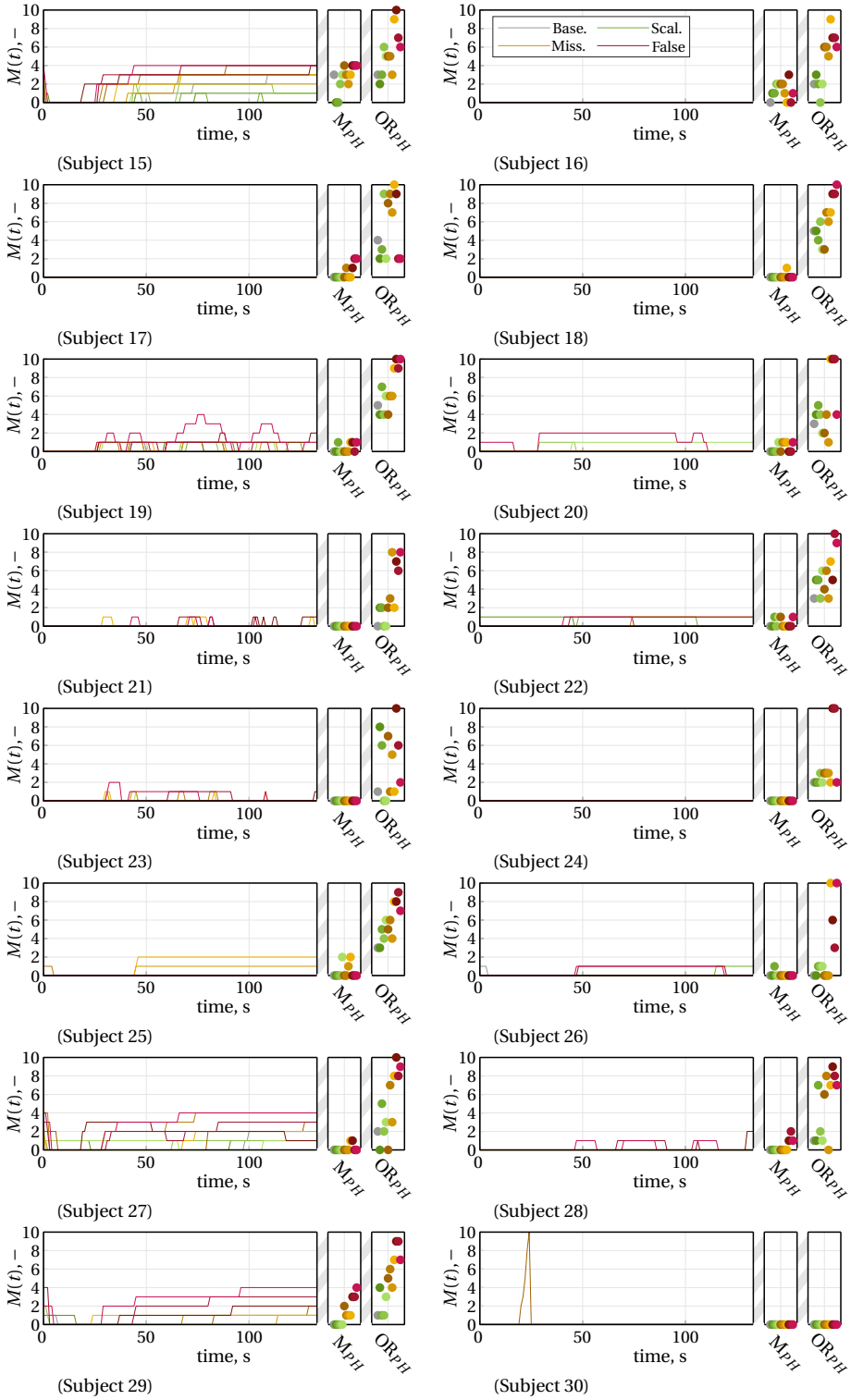


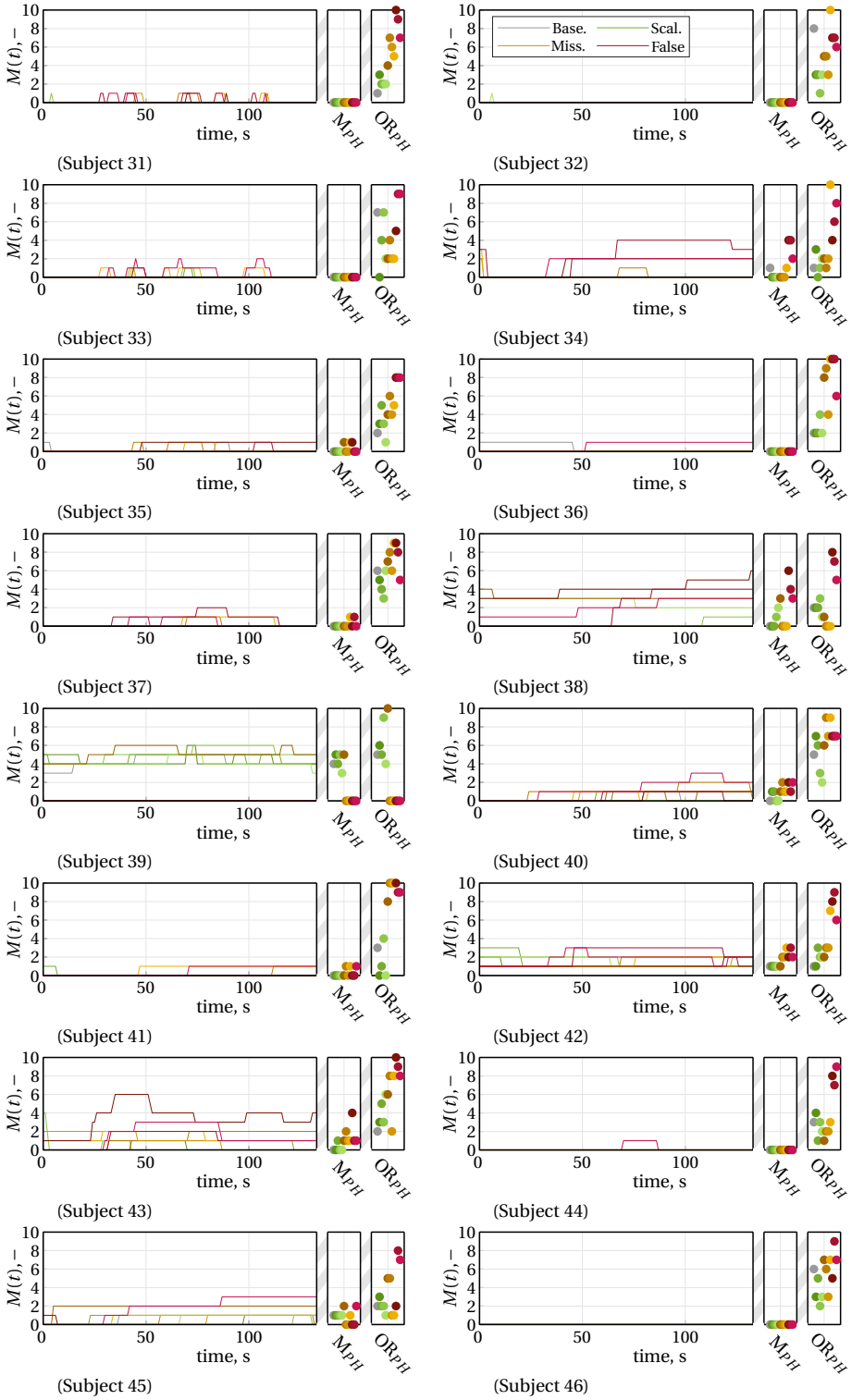
(Subject 42)

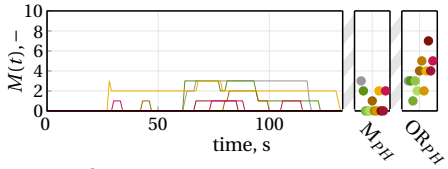
D.4 Chapter 8 - Incongruences and Simulator Sickness



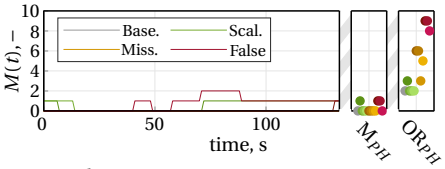
D


M(t), –
time, s
MPH
ORPH
Base.
Miss.
Scal.
False
(Subject 15)
(Subject 16)
(Subject 17)
(Subject 18)
(Subject 19)
(Subject 20)
(Subject 21)
(Subject 22)
(Subject 23)
(Subject 24)
(Subject 25)
(Subject 26)
(Subject 27)
(Subject 28)
(Subject 29)
(Subject 30)

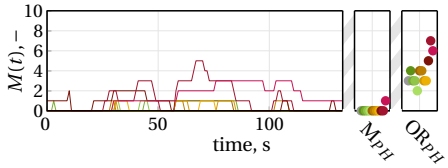

M(t), –
time, s
MPH
ORPH
Base.
Scal.
Miss.
False
(Subject 31)
(Subject 32)
(Subject 33)
(Subject 34)
(Subject 35)
(Subject 36)
(Subject 37)
(Subject 38)
(Subject 39)
(Subject 40)
(Subject 41)
(Subject 42)
(Subject 43)
(Subject 44)
(Subject 45)
(Subject 46)



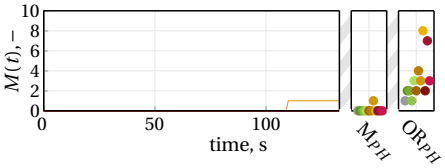
(Subject 47)



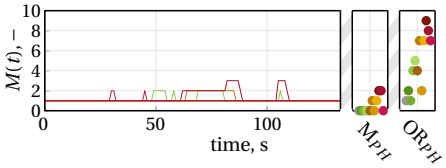
(Subject 48)



(Subject 49)

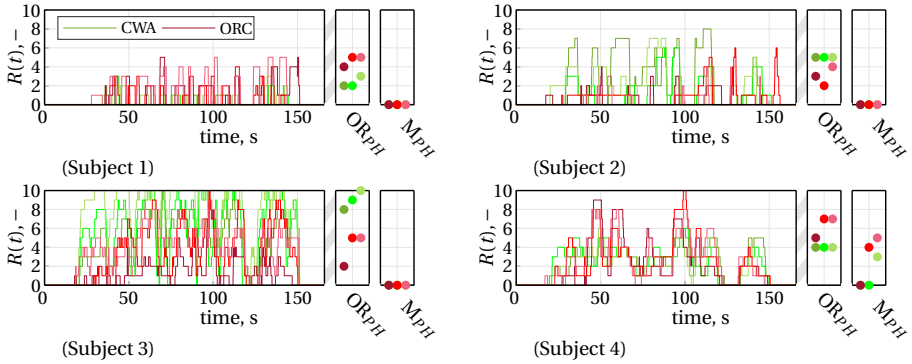


(Subject 50)



(Subject 51)

D.5 Appendix C - Incongruence Ratings in Dynamic Driving



Bibliography

Allen, R., Rosenthal, T., and Parseghian, Z (1995). Low Cost Driving Simulation for Research, Training and Screening Applications. In: *International Congress & Exposition*. SAE Technical Paper. No: 950171. doi: 10.4271/950171.

Almallah, M., Hussain, Q., Reinolsmann, N., and Alhajyaseen, W. K. (2021). Driving simulation sickness and the sense of presence: Correlation and contributing factors. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 78, pp. 180–193. doi: 10.1016/j.trf.2021.02.005.

Aminzadeh, M., Mahmoodi, A., and Sabzehparvar, M. (2012). Optimal Motion-Cueing Algorithm Using Motion System Kinematics. In: *European Journal of Control* 18 (4), pp. 363–375. doi: 10.3166/ejc.18.363-375.

Ariel, D. and Sivan, R. (1984). False Cue Reduction in Moving Flight Simulators. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-14 (4), pp. 665–671. doi: 10.1109/TSMC.1984.6313342.

Asadi, H., Lim, C. P., Mohamed, S., Nahavandi, D., and Nahavandi, S. (2019). Increasing Motion Fidelity in Driving Simulators Using a Fuzzy-Based Washout Filter. In: *IEEE Transactions on Intelligent Vehicles* 4 (2), pp. 298–308. doi: 10.1109/TIV.2019.2904388.

Asadi, H., Mohamed, S., Nelson, K., Nahavandi, S., and Rah, D. (2015). Human Perception-Based Washout Filtering Using Genetic Algorithm. In: *Proceedings of the International Conference on Neural Information Processing (ICONIP)*. Istanbul, Turkey, pp. 401–411. doi: 10.1007/978-3-319-26535-3_46.

Aykent, B., Paillot, D., Merienne, F., Fang, Z, and Kemeny, A. (2011). Study of the Influence of Different Washout Algorithms on Simulator Sickness for a Driving Simulation Task. In: *Proceedings of the ASME 2011 World Conference on Innovative Virtual Reality*. WINVR2011-5545. Milan, Italy, pp. 331–341. doi: 10.1115/WINVR2011-5545.

Bakker, R. M., Stroosma, O., Pool, D. M., van Paassen, M. M., and Mulder, M. (2025). Comparison of Classical and Optimization-Based Motion Cueing for Simulating Aircraft Upset Maneuvers. In: *AIAA SCITECH 2025 Forum*. doi: 10.2514/6.2025-0978.

Baldoni, F., Galante, F., Pernetti, M., Russo, M., Terzo, M., and Toscano, M. (2011). Tuning and objective performance evaluation of a driving simulator to investigate tyre behaviour in on-centre handling manoeuvres. In: *Vehicle System Dynamics* 49 (9), pp. 1423–1440. doi: 10.1080/00423114.2010.527353.

Baumann, G., Krantz, W., and Pitz, J. (2014). Evaluation of steering feel and vehicle handling in the Stuttgart Driving Simulator (Bewertung von Lenkgefühl und Fahrverhal-

ten im Stuttgarter Fahrsimulator). In: *Proceedings of the 5th International Munich Chassis Symposium 2014*, pp. 201–215. doi: 10.1007/978-3-658-05978-1_17.

Bazilinskyy, P., Kooijman, L., Dodou, D., and de Winter, J. C. (2020). Coupled simulator for research on the interaction between pedestrians and (automated) vehicles. In: *Proceedings of the Driving Simulation Conference 2020 Europe*. Antibes, France.

Beghi, A., Bruschetta, M., and Maran, F. (2012). A real time implementation of MPC based Motion Cueing strategy for driving simulators. In: *Proceedings of the 51st IEEE Conference on Decision and Control*. Maui, HI, pp. 6340–6345. doi: 10.1109/CDC.2012.6426119.

Bennett, P. (1995). Economic justification of robotic systems using graphical simulation as a tool. In: *6. American Nuclear Society meeting on robotics and remote systems*. Monterey, CA.

Berthoz, A., Bles, W., Bülthoff, H. H., Correia Grácio, B. J., Feenstra, P., Filliard, N., Hühne, R., Kemeny, A., Mayrhofer, M., Mulder, M., Nusseck, H.-G., Pretto, P., Reymond, G., Schlüsselberger, R., Schwandter, J., Teufel, H. J., Vailleau, B., van Paassen, M. M., Vidal, M., and Wentink, M. (2013). Motion scaling for high-performance driving simulators. In: *IEEE Transactions on Human-Machine Systems* 43 (3), pp. 265–276. doi: 10.1109/TSMC.2013.2242885.

Biemelt, P., Böhm, S., Gausemeier, S., and Trächtler, A. (2021). Subjective Evaluation of Filter- and Optimization-Based Motion Cueing Algorithms for a Hybrid Kinematics Driving Simulator. In: *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Melbourne, Australia, pp. 1619–1626. doi: 10.1109/SMC52423.2021.9658974.

BMW Group (2018). *BMW Group builds new Driving Simulation Centre in Munich*. (Visited on 08/29/2023). url: https://www.press.bmwgroup.com/global/article/detail-/T0284380EN/bmw-group-builds-new-driving-simulation-centre-in-munich?-language=en.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. In: *Trends in Ecology & Evolution* 24 (3), pp. 127–135. doi: 10.1016/j.tree.2008.10.008.

Bolstad, W. M. and Curran, J. M. (2007). *Introduction to Bayesian Statistics*. 2nd ed. Hoboken N.J.: Wiley-Interscience.

Bos, J., de Vries, S., Emmerik, M., and Groen, E. (2010). The effect of internal and external fields of view on visually induced motion sickness. In: *Applied ergonomics* 41 (4), pp. 516–521. doi: 10.1016/j.apergo.2009.11.007.

Bos, J. E. (2011). Nuancing the relationship between motion sickness and postural stability. In: *Displays* 32 (4), pp. 189–193. doi: 10.1016/j.displa.2010.09.005.

Bos, J. E., MacKinnon, S. N., and Patterson, A. (2005). Motion Sickness Symptoms in a Ship Motion Simulator: Effects of Inside, Outside, and No View. In: *Aviation, Space, and Environmental Medicine* 76 (12), pp. 1111–1118.

Bosetti, P., Da Lio, M., and Saroldi, A. (2014). On the human control of vehicles: an experimental study of acceleration. In: *European Transport Research Review* 6 (2), pp. 157–170. doi: 10.1007/s12544-013-0120-2.

Brems, W., van Doornik, J., de Vries, E., and Wiedemann, J. (2015). Frequency response and latency analysis of a driving simulator for chassis development and vehicle dynamics evaluation. In: *Proceedings of the Driving Simulation Conference 2015 Europe*. Tübingen, Germany.

Bruck, L., Haycock, B., and Emadi, A. (2021). A Review of Driving Simulation Technology and Applications. In: *IEEE Open Journal of Vehicular Technology* 2, pp. 1–16. doi: 10.1109/OJVT.2020.3036582.

Bruschetta, M., Cenedese, C., Beghi, A., and Maran, F. (2018a). A Motion Cueing Algorithm With Look-Ahead and Driver Characterization: Application to Vertical Car Dynamics. In: *IEEE Transactions on Human-Machine Systems* 48 (1), pp. 6–16. doi: 10.1109/THMS.2017.2776207.

Bruschetta, M., de Winkel, K. N., Mion, E., Pretto, P., Beghi, A., and Bülthoff, H. H. (2021). Assessing the contribution of active somatosensory stimulation to self-acceleration perception in dynamic driving simulators. In: *PLoS One* 16 (11), pp. 1–18. doi: 10.1371/journal.pone.0259015.

Bruschetta, M., Maran, F., and Beghi, A. (2017). A Nonlinear, MPC-Based Motion Cueing Algorithm for a High-Performance, Nine-DOF Dynamic Simulator Platform. In: *IEEE Transactions on Control Systems Technology* 25 (2), pp. 686–694. doi: 10.1109/TCST.2016.2560120.

Bruschetta, M., Mendola, D. L., Beghi, A., and Minen, D. (2018b). An MPC based Motion Cueing Algorithm with side slip dynamics. In: *Proceedings of the Driving Simulation Conference 2018 Europe*. Antibes, France, pp. 55–59.

Burnham, K. P., Anderson, D. R., Burnham, K. P., and Anderson, D. R. (1998). *Practical use of the information-theoretic approach*. New York, NY: Springer. doi: 10.1007/978-1-4757-2917-7_3.

Buschmann, T. (2014). "Dynamics and Control of Redundant Robots". Habilitation Thesis. Technische Universität München.

Cai, B. and Wang, X. (2024). Estimation of the Smallest Acceptable Sample Size in Bilateral Approaches to Coefficient Estimation and Accuracy Prediction. In: *Transportation Research Record* 2678 (8), pp. 690–699.

Caird, J. K. and Horrey, W. J. (2011). *Handbook of driving simulation for engineering, medicine, and psychology*. Vol. 2. Boca Raton, FL: CRC Press.

Casas, S., Coma, I., Portalés, C., and Fernández, M (2016). Towards a simulation-based tuning of motion cueing algorithms. In: *Simulation Modelling Practice and Theory* 67, pp. 137–154. doi: 10.1016/j.simpat.2016.06.002.

Casas, S., Inmaculada, C., Vicente Riera, J., and Fernández, M. (2015). Motion-Cueing Algorithms: Characterization of Users' Perception. In: *Human Factors* 57 (1), pp. 144–162. doi: 10.1177/0018720814538281.

Casas-Yrurzum, S., Portalés-Ricart, C., Morillo-Tena, P., and Cruz-Neira, C. (2020). On the Objective Evaluation of Motion Cueing in Vehicle Simulations. In: *IEEE Transactions on Intelligent Transportation Systems* 99 (5), pp. 1–13. doi: 10.1109/TITS.2020.2978498.

Chai, C., Zeng, X., Alvarez, I., and Elli, M. S. (2020). Evaluation of Responsibility-Sensitive Safety (RSS) Model based on Human-in-the-loop Driving Simulation. In: *23rd IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Rhodes, Greece, pp. 1–7. doi: 10.1109/ITSC45102.2020.9294637.

Chanmas, G., Taveekitworachai, P., Paliyawan, P., Thawonmas, R., Thawonmas, R., Nukoolkit, C., and Dajpratham, P. (Jan. 2023). Driving scenarios and environmental settings in simulator-based driving assessment systems for stroke: a systematic review. In: *Topics in Stroke Rehabilitation* 30, pp. 872–880. doi: 10.1080/10749357.2023.2165273.

Chiew, Y. S., Abdul Jalil, M. K., and Hussein, M (2008). Kinematic Modeling of Driving Simulator Motion Platform. In: *Proceedings of the 2008 IEEE Conference on Innovative Technologies in Intelligent Systems and Industrial Applications*. Cyberjaya, Malaysia, pp. 30–34. doi: 10.1109/CITISIA.2008.4607330.

Cleij, D., Venrooij, J., Pretto, P., Katliar, M., Bülthoff, H. H., Steffen, D., Hoffmeyer, F., and Schöner, H.-P. (2019). Comparison between filter- and optimization-based motion cueing algorithms for driving simulation. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 61, pp. 53 –68. doi: 10.1016/j.trf.2017.04.005.

Cleij, D. (2020). "Measuring, modelling and minimizing perceived motion incongruence". PhD Dissertation. Delft University of Technology. doi: 10.4233/uuid:45fd3f70-2ba6-43fa-a2c4-018967bfdc88.

Cleij, D., Pool, D. M., Mulder, M., and Bülthoff, H. H. (2020). Optimizing an Optimization-Based MCA using Perceived Motion Incongruence Models. In: *Proceedings of the Driving Simulation Conference Europe*. Antibes, France, pp. 53–60.

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. In: *IEEE Transactions on Human-Machine Systems* 48 (1), pp. 17–29. doi: 10.1109/THMS.2017.2717884.

Cobb, S. V., Nichols, S., Ramsey, A., and Wilson, J. R. (1999). Virtual reality-induced symptoms and effects (VRISE). In: *Presence: Teleoperators & Virtual Environments* 8 (2), pp. 169–186. doi: 10.1162/105474699566152.

Conrad, B., Douvillier, J. G., and Schmidt, S. F. (1973). Washout Circuit Design for Multi-Degrees-Of-Freedom Moving Base Simulators. In: *Proceedings of the AIAA Visual and Motion Simulation Conference*. AIAA-1973-929. Palo Alto, CA. doi: 10.2514/6.1973-929.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. In: *Psychometrika* 16, 297–334.

Dagdelen, M., Reymond, G., Kemeny, A., Bordier, M., and Maïzi, N. (2004). MPC Based Motion Cueing Algorithm: Development and Application to the ULTIMATE Driving Simulator. In: *Proceedings of the Driving Simulation Conference 2004 Europe*. Paris, France, pp. 221–233.

Dagdelen, M., Reymond, G., Kemeny, A., Bordier, M., and Maïzi, N. (2009). Model-based predictive motion cueing strategy for vehicle driving simulators. In: *Control Engineering Practice* 17 (9), pp. 995–1003. doi: 10.1016/j.conengprac.2009.03.002.

de Winkel, K., Irmak, T., Kotian, V., Pool, D. M., and Happee, R. (2022). Relating individual motion sickness levels to subjective discomfort ratings. In: *Experimental Brain Research* 240, 1231–1240. doi: 10.1007/s00221-022-06334-6.

de Winter, J., van Leeuwen, P., and Happee, R. (2012). Advantages and Disadvantages of Driving Simulators: A Discussion. In: *Proceedings of Measuring Behavior 2012*. Utrecht, The Netherlands.

Demir, M., McNeese, N., Gorman, J., Cooke, N., Myers, C., and Grimm, D. (2021). Exploration of Teammate Trust and Interaction Dynamics in Human-Autonomy Teaming. In: *IEEE Transactions on Human-Machine Systems* 51, pp. 696–705. doi: 10.1109/THMS.2021.3115058.

Diebel, J. (2006). *Representing Attitude: Euler Angles, Unit Quaternions, and Rotation Vectors*.

Diels, C., Dugenetet, P., Brietzke, A., and Pham Xuan, R. (2023). Design strategies to alleviate motion sickness in rear seat passengers – a test track study. In: *Proceedings of the IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. Bilbao, Spain, pp. 5254–5258. doi: 10.1109/ITSC57777.2023.10421968.

Ellensohn, F., Hristakiev, D., Schwienbacher, M., Venrooij, J., and Rixen, D. (2019a). Evaluation of an Optimization Based Motion Cueing Algorithm Suitable for Online Application. In: *Proceedings of the Driving Simulation Conference 2019 Europe*. Strasbourg, France, pp. 93–100.

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

Ellensohn, F., Schwienbacher, M., Venrooij, J., and Rixen, D. (2019b). Motion Cueing Algorithm for a 9 DoF Driving Simulator: MPC with Linearized Actuator Constraints. In: *SAE International Journal of Connected and Automated Vehicles* 2 (3), pp. 145–155. doi: 10.4271/12-02-03-0010.

Ellensohn, F., Spannagl, M., Agabekov, S., Venrooij, J., Schwienbacher, M., and Rixen, D. (2020). A hybrid motion cueing algorithm. In: *Control Engineering Practice* 97, p. 104342. doi: 10.1016/j.conengprac.2020.104342.

Ellensohn, F., Venrooij, J., Schwienbacher, M., and Rixen, D. (2019c). Experimental evaluation of an optimization-based motion cueing algorithm. In: *Transportation Research Part F Traffic Psychology and Behaviour* 62, pp. 115–125. doi: 10.1016/j.trf.2018.12.004.

Eppink, J., Kolff, M., Venrooij, J., Pool, D. M., and Mulder, M. (2023). Probabilistic Prediction of Longitudinal Driving Behaviour for Driving Simulator Pre-Positioning. In: *Proceedings of the Driving Simulation Conference 2023 Europe*. Antibes, France, pp. 119–126.

Ezzati Amini, R., Al Haddad, C., Batabyal, D., Gkena, I., De Vos, B., Cuenen, A., Brijs, T., and Antoniou, C. (2023). Driver distraction and in-vehicle interventions: A driving simulator study on visual attention and driving performance. In: *Accident Analysis & Prevention* 191, p. 107195. doi: 10.1016/j.aap.2023.107195.

Fang, Z., Tsushima, M., Kitahara, E., Machida, N., Wautier, D., and Kemeny, A. (2017). Motion cueing algorithm for high performance driving simulator using yaw table. In: *IFAC-PapersOnLine* 50 (1), pp. 15965–15970. doi: 110.1016/j.ifacol.2017.08.1750.

Fang, Z. and Kemeny, A. (2012). Explicit MPC motion cueing algorithm for real-time driving simulator. In: *Proceedings of The 7th International Power Electronics and Motion Control Conference*. Harbin, China, pp. 874–878. doi: 10.1109/IPEMC.2012.6258965.

Fang, Z. and Kemeny, A. (2016). An efficient Model Predictive Control-based motion cueing algorithm for the driving simulator. In: *SIMULATION* 92 (11), pp. 1025–1033. doi: 10.1177/0037549716667835.

Fang, Z., Wautier, D., and Kemeny, A. (2022). FFT based optimal MCA for AD/ADAS driving tests. In: *Proceedings of the Driving Simulation Conference 2022 Europe*. Strasbourg, France, 119–126.

Field, Z., Miles, J., and Field, A. (2012). *Discovering statistics using R*. Thousand Oaks, CA: Sage Publications, pp. 1–992.

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

Fischer, M. and Werneke, J. (2008). The new time-variant motion cueing algorithm for the DLR dynamic driving simulator. In: *Proceedings of the Driving Simulation Conference 2015 Europe*. Monaco, France.

Fredrickson, B. L. and Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. In: *Journal of Personality and Social Psychology* 65 (1), pp. 45–55. doi: 10.1037//0022-3514.65.1.45.

Freeman, J. S., Watson, G., Papelis, Y. E., Lin, T. C., Tayyab, A., Romano, R. A., and Kuhl, J. G. (1995). *The Iowa Driving Simulator: An Implementation and Application Overview*. SAE Technical Paper. No: 950174.

Garrett, N. J. I. and Best, M. C. (2010). Driving simulator motion cueing algorithms - a survey of the state of the art. In: *Proceedings of the 10th International Symposium on Advanced Vehicle Control (AVEC)*. Loughborough, United Kingdom, pp. 183–188.

Garrett, N. J. I. and Best, M. C. (2013). Model predictive driving simulator motion cueing algorithm with actuator-based constraints. In: *Vehicle System Dynamics* 51 (8), pp. 1151–1172. doi: 10.1080/00423114.2013.783219.

Ghafarian, M., Watson, M., Mohajer, N., Nahavandi, D., Kebria, P., and Mohamed, S. (2023). A Review of Dynamic Vehicular Motion Simulators: Systems and Algorithms. In: *IEEE Access* 11, pp. 36331–36348. doi: 10.1109/ACCESS.2023.3265999.

Gough, V. E. (1962). Universal tyre test machine. In: *Proceedings of the FISITA 9th International Technical Congress*. London, United Kingdom, pp. 117–137.

Grant, P. R. and Reid, L. D. (1997). Motion Washout Filter Tuning: Rules and Requirements. In: *Journal of Aircraft* 34 (2), pp. 145–151. doi: 10.2514/2.2158.

Greig, G. L. (1987). "Masking of motion cues by random motion: comparison of human performance with a signal detection model". PhD Dissertation. University of Toronto.

Griffin, M. and Howarth, H. (2000). *Motion Sickness History Questionnaire*. Tech. rep. ISVR Technical Report No. 283. University of Southampton, Institute of Sound and Vibration Research.

Grottoli, M., Cleij, D., Pretto, P., Lemmens, Y., Happee, R., and Bülthoff, H. H. (2019). Objective evaluation of prediction strategies for optimization-based motion cueing. In: *Simulation* 95 (8), pp. 707–724. doi: 10.1177/0037549718815972.

Himmels, C., Rock, T., Venrooij, J., and Riener, A. (2022a). Simulator Fidelity Influences the Sense of Presence in Driving Simulators. In: *Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Seoul, Republic of Korea, pp. 53–57. doi: 10.1145/3544999.3552526.

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

Hlavačka, F., Mergner, T., and Schweigart, G. (1992). Interaction of vestibular and proprioceptive inputs for human self-motion perception. In: *Neuroscience Letters* 138, pp. 161–164. doi: 10.1016/0304-3940(92)90496-t.

Hogema, J., Wentink, M., and Bertollini, G. (2012). Effects of Yaw Motion on Driving Behaviour, Comfort and Realism. In: *Proceedings of the Driving Simulation Conference 2012 Europe*. Paris, France, pp. 149–158.

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

Houck, J., Telban, R., and Cardullo, F. (2005). *Motion Cueing Algorithm Development: Human-Centered Linear and Nonlinear Approaches*. Tech. rep. CR-2005-213747. State University of New York.

Huang, A. and Chihsiuh, C. (2003). A low-cost driving simulator for full vehicle dynamics simulation. In: *IEEE Transactions on Vehicular Technology* 52 (1), pp. 162–172. doi: 10.1109/TVT.2002.807157.

Huang, Y., Pool, D. M., Stroosma, O., Chu, Q. P., and Mulder, M. (2016a). A Review of Control Schemes for Hydraulic Stewart Platform Flight Simulator Motion Systems. In: *AIAA Modeling and Simulation Technologies Conference*, p. 1436. doi: 10.2514/6.2016-1436.

Huang, Y., Pool, D. M., Stroosma, O., Chu, Q. P., and Mulder, M. (2016b). Modeling and Simulation of Hydraulic Hexapod Flight Simulator Motion Systems. In: *AIAA Modeling and Simulation Technologies Conference*, p. 1437. doi: 10.2514/6.2016-1437.

Igoshina, E., Russo, F. A., Shewaga, R., Haycock, B., and Keshavarz, B. (2022). The relationship between simulator sickness and driving performance in a high-fidelity simulator. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 89, pp. 478–487. doi: 10.1016/j.trf.2022.07.015.

Irmak, T., Kotian, V., Happee, R., de Winkel, K., and Pool, D. M. (2022). Amplitude and Temporal Dynamics of Motion Sickness. In: *Frontiers in Systems Neuroscience* 16. doi: 10.3389/fnsys.2022.866503.

Irmak, T., Pool, D. M., de Winkel, K., and Happee, R. (2023). Validating models of sensory conflict and perception for motion sickness prediction. In: *Biological Cybernetics* 117, pp. 185–209. doi: 10.1007/s00422-023-00959-8.

Irmak, T., Pool, D. M., and Happee, R. (2021). Objective and subjective responses to motion sickness: the group and the individual. In: *Experimental Brain Research* 239 (2), pp. 515–531. doi: 10.1007/s00221-020-05986-6.

Jamson, A. H. J. (2010). "Motion cueing in driving simulators for research applications". PhD Dissertation. University of Leeds.

Jansson, J., Sandin, J., Augusto, B., Fischer, M., Blissing, B., and Källgren, L. (2014). Design and performance of the VTI Sim IV. In: *Proceedings of the Driving Simulation Conference 2014 Europe*. Paris, France, pp. 128–138.

JASP Team (2023). *JASP (Version 0.17)[Computer software]*. doi: https://jasp-stats.org/.

Jeffreys, H. (1961). *Theory of Probability*. 3rd ed. Oxford, United Kingdom: Oxford University Press.

Jeong, C., Kim, B., Yu, S., Suh, D., Kim, M., and Suh, M. (2013). In-vehicle display HMI safety evaluation using a driving simulator. In: *International Journal of Automotive Technology* 14 (6), 987–992. doi: 10.1007/s12239-013-0108-x.

Johnson, D. M. (2007). *Simulator sickness research summary*. Tech. rep. TR-HFM-121-Part-II. US Army Research Institute for the Behavioral and Social Science Ft. Rucker.

Kappe, B. and Van Winsum, W (2002). A cost-effective driving simulator. In: *Proceedings of the Driving Simulation Conference 2003 North America*. Paris, France, pp. 123–133.

Katliar, M., de Winkel, K., Venrooij, J., Pretto, P., and Bülthoff, H. (2015). Impact of MPC Prediction Horizon on Motion Cueing Fidelity. In: *Proceedings of the Driving Simulation Conference 2015 Europe*. Tübingen, Germany.

Katliar, M. (2020). "Optimal control of motion simulators". PhD Dissertation. Albert-Ludwigs-Universität Freiburg.

Kelter, R. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: a non-technical introduction to Bayesian inference with JASP. In: *BMC Medical Research Methodology* 20. doi: 10.1186/s12874-020-00980-6.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. In: *The International Journal of Aviation Psychology* 3 (3), pp. 203–220. doi: 10.1207/s15327108ijap0303_3.

Keshavarz, B., Ramkhalawansingh, R., Haycock, B., Shahab, S., and Campos, J. (2018). Comparing simulator sickness in younger and older adults during simulated driving under different multisensory conditions. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 54, pp. 47–62. doi: 10.1016/j.trf.2018.01.007.

Khusro, Y., Zheng, Y., Grottoli, M., and Shyrokau, B. (2020). MPC-Based Motion-Cueing Algorithm for a 6-DOF Driving Simulator with Actuator Constraints. In: *Vehicles* 2 (4), pp. 625–647. doi: 10.3390/vehicles2040036.

Klüver, M., Herrigel, C., Preuß, S., Schöner, H.-P., and Hecht, H. (2015). Comparing the incidence of simulator sickness in five different driving simulators. In: *Proceedings of the Driving Simulation Conference 2015 Europe*. Tübingen, Germany, 87–94.

Kolff, M., van der El, K., Pool, D. M., van Paassen, M. M., and Mulder, M. (2019). Approximating Road Geometry with Multisine Signals for Driver Identification. In: *Fourteenth IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, September 16-19*. Tallinn, Estonia, pp. 341–346. doi: 10.1016/j.ifacol.2019.12.082.

Kolff, M., Venrooij, J., Arcidiacono, E., Pool, D. M., and Mulder, M. (2024a). Predicting Motion Incongruence Ratings in Closed- and Open-Loop Urban Driving Simulation. In: *IEEE Transactions on Intelligent Transportation Systems* 26 (1), pp. 517–528. doi: 10.1109/TITS.2024.3503496.

Kolff, M., Venrooij, J., Schwienbacher, M., Pool, D. M., and Mulder, M. (2021). Quality Comparison of Motion Cueing Algorithms for Urban Driving Simulations. In: *Proceedings of the Driving Simulation Conference 2021 Europe*. Munich, Germany, 141–148.

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: *Proceedings of the Driving Simulation Conference 2022 Europe*. Strasbourg, France, 111–118.

Kolff, M., Venrooij, J., Schwienbacher, M., Pool, D. M., and Mulder, M. (2023). The Importance of Kinematic Configurations for Motion Control of Driving Simulators. In: *IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. Bilbao, Spain, pp. 1000–1006. doi: 10.1109/ITSC57777.2023.10422127.

Kolff, M., Venrooij, J., Schwienbacher, M., Pool, D. M., and Mulder, M. (2024b). Reliability and Models of Subjective Motion Incongruence Ratings in Urban Driving Simulations. In: *IEEE Transactions on Human-Machine Systems* 54 (6), pp. 634–645. doi: 10.1109/THMS.2024.3450831.

Kotian, V., Irmak, T., Pool, D. M., and Happee, R. (2024). The role of vision in sensory integration models for predicting motion perception and sickness. In: *Experimental Brain Research*. doi: 10.1007/s00221-023-06747-x.

Kuiper, O. X., Bos, J. E., Schmidt, E. A., Diels, C., and Wolter, S. (2020). Knowing What's Coming: Unpredictable Motion Causes More Motion Sickness. In: *Human Factors* 62 (8), pp. 1339–1348. doi: 10.1177/0018720819876139.

Kusachov, A., Bruzelius, F., and Xie, X (2015). The Importance of Yaw Motion Feedback in Driving Simulators. In: *Proceedings of the 24th Symposium of the International Association for Vehicle System Dynamics*. Graz, Austria, pp. 769–778.

Lakens, D. (2017). Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. In: *Social Psychological and Personality Science* 8, pp. 1–8. doi: 10.1177/1948550617697177.

Lakens, D., Scheel, A. M., and Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. In: *Advances in Methods and Practices in Psychological Science* 1 (2), pp. 259–269. doi: 10.1177/2515245918770963.

Lakerveld, P., Damveld, H., Pool, D., van der El, K., van Paassen, M., and Mulder, M. (2016). The Effects of Yaw and Sway Motion Cues in Curve Driving Simulation. In: *IFAC-PapersOnLine* 49 (19), pp. 500–505. doi: 10.1016/j.ifacol.2016.10.640.

Lamprecht, A., Steffen, D., Haecker, J., and Graichen, K. (2019). Comparison between a Filter- and an MPC-based MCA in an Offline Simulator Study. In: *Proceedings of the Driving Simulation Conference 2019 Europe*. Strasbourg, France.

Lamprecht, A., Steffen, D., Nagel, K., Haecker, J., and Graichen, K. (2021). Online Model Predictive Motion Cueing With Real-Time Driver Prediction. In: *IEEE*

Transactions on Intelligent Transportation Systems 23 (8), pp. 12414–12428. doi: 10.1109/TITS.2021.3114003.

Linde, M, Tendeiro, J., Selker, R, Wagenmakers, E., and van Ravenzwaaij, D. (2023). Decisions about equivalence: A comparison of TOST, HDI-ROPE, and the Bayes factor. In: *Psychol Mthds* 28 (3), pp. 740–755. doi: 10.1037/met0000402.

Lindner, J., Keler, A., Grigoropoulos, G., Malcolm, P., Denk, F., Brunner, P., and Bogenberger, K. (2022). A coupled driving simulator to investigate the interaction between bicycles and automated vehicles. In: *25th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1335–1341. doi: 10.1109/ITSC55140.2022.9922400.

Liu, K., Fitzgerald, J., and Lewis, F. (1993). Kinematic analysis of a Stewart platform manipulator. In: *IEEE Transactions on Industrial Electronics* 40 (2), pp. 282–293. doi: 10.1109/41.222651.

Mallery, C. J. (1987). The Effect of Experience on Subjective Ratings for Aircraft and Simulator Workload during IFR Flight. In: *Proceedings of the Human Factors Society Annual Meeting* 31 (7), pp. 838–841. doi: 10.1177/154193128703100734.

Markkula, G., Romano, R., Waldram, R., Giles, O., Mole, C., and Wilkie, R. (2019). Modelling visual-vestibular integration and behavioural adaptation in the driving simulator. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 66, pp. 310–323. doi: 10.1016/j.trf.2019.07.018.

McDonald, R. P. (2000). Test Theory: A Unified Treatment. In: *Journal of the American Statistical Association* 95 (451), pp. 1012–1013. doi: 10.2307/2669496.

McRuer, D. T. and Jex, H. R. (1967). A Review of Quasi-Linear Pilot Models. In: *IEEE Transactions on Human Factors in Electronics* HFE-8 (3), pp. 231–249. doi: 10.1109/THFE.1967.234304.

Meike, D and Ribickis, L (2011). Energy efficient use of robotics in the automobile industry. In: *15th International Conference on Advanced Robotics (ICAR)*. Tallinn, Estonia, pp. 507–511. doi: 10.1109/ICAR.2011.6088567.

Miletović, I. (2020). "Motion Cueing Fidelity in Rotorcraft Flight Simulation". PhD Dissertation. Delft University of Technology.

Mohajer, N., Abdi, H., Nelson, K., and Nahavandi, S. (2015). Vehicle motion simulators, a key step towards road vehicle dynamics improvement. In: *Vehicle System Dynamics* 53 (8), pp. 1204–1226. doi: 10.1080/00423114.2015.1039551.

Mourant, R. R., Rengarajan, P., Cox, D., Lin, Y., and Jaeger, B. K. (2007). The Effect of Driving Environments on Simulator Sickness. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 1232–1236. doi: 10.1177/154193120705101838.

Mourant, R. R. and Thattacherry, T. R. (2000). Simulator sickness in a virtual environments driving simulator. In: *Proceedings of the human factors and ergonomics society annual meeting*. Los Angeles, CA, pp. 534–537. doi: 10.1177/154193120004400513.

Mulder, M., Pool, D. M., Abbink, D. A., Boer, E. R., Zaal, P. M. T., Drop, F. M., van der El, K., and van Paassen, M. M. (2018). Manual Control Cybernetics: State-of-the-Art and Current Trends. In: *IEEE Transactions on Human-Machine Systems* 48 (5), pp. 468–485. doi: 10.1109/THMS.2017.2761342.

Mulder, M., Pool, D. M., van der El, K., and van Paassen, M. (2022). Neuroscience Perspectives on Adaptive Manual Control with Pursuit Displays. In: *IFAC-PapersOnLine* 55 (29), pp. 160–165. doi: 10.1016/j.ifacol.2022.10.249.

Munir, S., Hovd, M., Fang, Z., Olaru, S., and Kemeny, A. (2017). Complexity reduction in motion cueing algorithm for the ULTIMATE driving simulator. In: *IFAC-PapersOnLine* 50 (1), pp. 10729–10734. doi: 10.1016/j.ifacol.2017.08.2256.

Murdock Jr, B. B. (1962). Serial Position Effect of Free Recall. In: *Journal of Experimental Psychology* 64 (5), 482 – 488. doi: 10.1037/h0045106.

Nahon, M. A. and Reid, L. D. (1990). Simulator Motion-Drive Algorithms: A Designer's Perspective. In: *Journal of Guidance, Control, and Dynamics* 13 (2), pp. 356–362. doi: 10.2514/3.20557.

Naseri, A. and Grant, P. (2005). An Improved Adaptive Motion Drive Algorithm. In: *AIAA Modeling and Simulation Technologies Conference and Exhibit*. San Francisco, CA, pp. 1–9. doi: 10.2514/6.2005-6500.

Nehaoua, L., Amouri, A., and Arioui, H. (2005). Classic and Adaptive Washout Comparison for a Low Cost Driving Simulator. In: *Proceedings of the 2005 Mediterranean Conference on Control and Automation*. Limassol, Cyprus, pp. 586–591. doi: 10.1109/.2005.1467080.

Nesti, A., Nooij, S., Losert, M., Bülthoff, H. H., and Pretto, P. (2016). Roll rate perceptual thresholds in active and passive curve driving simulation. In: *SIMULATION* 92 (5), pp. 417–426. doi: 10.1177/0037549716637135.

Nguyen Van, Q. H. and Ito, S. (2016). On the Implementation of Steering Controls for Evaluation of Automated Driving Systems in Driving Simulator. In: *Proceedings of the Driving Simulation Conference 2016 Europe*. Paris, France, pp. 165–168.

Nogueira, D. W. P., Bacurau, R. M., Lima, D. A., Silva, F. O., and Neto, A. M. (2021). In Motion Low-Cost IMU-to-Vehicle Alignment for Intelligent Vehicle Applications Using Kalman Filter. In: *24th IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Indianapolis, IN, pp. 92–97. doi: 10.1109/ITSC48978.2021.9564862.

Olivari, M., Pretto, P., Venrooij, J., and Bülthoff, H. (2019). Defining the kinematic requirements for a theoretical driving simulator. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 61, pp. 5–15. doi: 10.1016/j.trf.2017.11.006.

Ordoñez, L. M., Solarte, J. E., Krut, S., and Vivas, A. (2017). Simulation and energy consumption analysis of a parallel robot. In: *IEEE 3rd Colombian Conference on Automatic Control*. Cartagena de Indias, Colombia, pp. 1–6. doi: 10.1109/CCAC.2017.8276417.

Parduzi, A. (2021). "Bewertung der Validität von Fahrsimulatoren anhand vibroakustischer Fahrzeugschwingungen". PhD Dissertation. Technical University of Berlin.

Parrish, R. V., Dieudonne, J. E., Bowles, R. L., and Martin Jr., D. J. (1975). Coordinated Adaptive Washout for Motion Simulators. In: *Journal of Aircraft* 12 (1), pp. 44–50. doi: 10.2514/3.59800.

Perfect, P., Timson, E., White, M. D., Padfield, G. D., Erdos, R., and Gubbels, A. W. (2014). A rating scale for the subjective assessment of simulation fidelity. In: *The Aeronautical Journal* 118 (1206), 953–974. doi: 10.1017/S0001924000009635.

Pham, D.-A. and Nguyen, D.-T. (2022). Auto-Tuning parameters of motion cueing algorithms for high performance driving simulator based on Kuka Robocoaster. In: *Science Progress* 105 (2). doi: 10.1177/00368504221104333.

Pinheiro, J. and Bates, D. (2000). "Linear Mixed-Effects Models: Basic Concepts and Examples". In: *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer New York, pp. 3–56. doi: 10.1007/0-387-22747-4_1.

Pool, D. M. (2012). "Objective Evaluation of Flight Simulator Motion Cueing Fidelity Through a Cybernetic Approach". PhD Dissertation. Delft University of Technology, Faculty of Aerospace Engineering.

Pouliot, N. A., Gosselin, C. M., and Nahon, M. A. (1998). Motion Simulation Capabilities of Three-Degree-of-Freedom Flight Simulators. In: *Journal of Aircraft* 35 (1), pp. 9–17. doi: 10.2514/2.2283.

Qaisi, I. A. and Trächtler, A. (2012). Human in the loop: Optimal control of driving simulators and new motion quality criterion. In: *IEEE International Conference on Systems, Man, and Cybernetics*. Seoul, Korea, pp. 2235–2240. doi: 10.1109/ICSMC.2012.6378073.

Qazani, M. R. C., Asadi, H., and Nahavandi, S. (2020). A Motion Cueing Algorithm Based on Model Predictive Control Using Terminal Conditions in Urban Driving Scenario. In: *IEEE Systems Journal* 15 (1), pp. 445–453. doi: 10.1109/JSYST.2020.2994154.

Qazani, M. R. C., Asadi, H., Najdovski, Z., Alsanwy, S., Zakarya, M., Alam, F., Ouakad, H. M., Lim, C. P., and Nahavandi, S. (2024). High-fidelity learning-based motion cueing algorithm by bypassing worst-case scenario-based tuning technique. In: *Cognitive Robotics* 4, pp. 116–127. doi: 10.1016/j.cogr.2024.07.001.

Qazani, M., Asadi, H., Zhang, L., Tabarsinezhad, F., Mohamed, S., Lim, C., and Nahavandi, S. (2022). A New Prepositioning Technique of a Motion Simulator Platform Using Nonlinear Model Predictive Control and Recurrent Neural Network.

In: *IEEE Transactions on Intelligent Transportation Systems* 22 (13), pp. 23268 –23277. doi: 10.1109/TITS.2022.3195964.

Qiu, Z., Mcgill, M., Pöhlmann, K. M. T., and Brewster, S. A. (2023). Manipulating the Orientation of Planar 2D Content in VR as an Implicit Visual Cue for Mitigating Passenger Motion Sickness. In: *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 1–10. doi: 10.1145/3580585.3607157.

Reason, J. T. and Brand, J. J. (1975). *Motion sickness*. Cambridge, MA: Academic press.

Reason, J. (1978). Motion Sickness Adaptation: A Neural Mismatch Model. In: *Journal of the Royal Society of Medicine* 71 (11), pp. 819–829. doi: 10.1177/014107687807101109.

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.

Rengifo, C., Chardonnet, J.-R., Mohellebi, H., and Kemeny, A. (2021). Impact of Human-Centered Vestibular System Model for Motion Control in a Driving Simulator. In: *IEEE Transactions on Human-Machine Systems* 51 (5), pp. 411–420. doi: 10.1109/THMS.2021.3102506.

Rengifo, C., Chardonnet, J.-R., Mohellebi, H., Paillot, D., and Kemeny, A. (2019). Feasibility Analysis For Constrained Model Predictive Control Based Motion Cueing Algorithm. In: *2019 International Conference on Robotics and Automation (ICRA)*. Montréal, Canada, pp. 2076–2082. doi: 10.1109/ICRA.2019.8794129.

Reuten, A., Nooij, S., Bos, J., and Smeets, J. (2021). How feelings of unpleasantness develop during the progression of motion sickness symptoms. In: *Experimental Brain Research* 239 (12), pp. 3615–3624. doi: 10.1007/s00221-021-06226-1.

Reymond, G. and Kemeny, A. (2000). Motion Cueing in the Renault Driving Simulator. In: *Vehicle System Dynamics* 34 (4), pp. 249–259. doi: 10.1076/vesd.34.4.249.2059.

Rolnick, A. and Lubow, R. E. (1991). Why is the driver rarely motion sick? The role of controllability in motion sickness. In: *Ergonomics* 34 (7), pp. 867–879.

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. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 65 (1), pp. 34–45. doi: 10.1016/j.trf.2019.07.001.

Rosenthal, R. and Jacobson, L (1968). Pygmalion in the classroom. In: *Urban Rev* 3, pp. 16–20. doi: 10.1007/BF02322211.

Rossi, R., Gastaldi, M., Biondi, F., Orsini, F., De Cet, G., and Mulatti, C. (2020). A Driving Simulator Study Exploring the Effect of Different Mental Models on ADAS System Effectiveness. In: *Lecture Notes in Computer Science Augmented Reality, Virtual Reality, and Computer Graphics*. Lecce, Italy, pp. 102–113. doi: 10.1007/978-3-030-58465-8_7.

Salisbury, I. and Limebeer, D. (2014). Optimal Race Car Motion Cueing. In: *Proceedings of the 19th World Congress The International Federation of Automatic Control*. Cape Town, South Africa, pp. 7547–7552.

Salisbury, I. and Limebeer, D. (2017). Motion cueing in high-performance vehicle simulators. In: *Vehicle System Dynamics* 55 (6), pp. 775–801. doi: 10.1080/00423114.2017.1280172.

Salter, S., Diels, C., Herriotts, P., Kanarachos, S., and Thake, D. (2019). Motion sickness in automated vehicles with forward and rearward facing seating orientations. In: *Applied Ergonomics* 78, pp. 54–61. doi: 10.1016/j.apergo.2019.02.001.

Savalei, V. and Reise, S. P. (2019). Don't forget the model in your model-based reliability coefficients: A reply to McNeish (2018). In: *Collabra: Psychology* 5 (1), p. 36. doi: 10.1525/collabra.247.

Schmidt, J., Braunagel, C., Stolzmann, W., and Karrer-Gauß, K. (2016). Driver drowsiness and behavior detection in prolonged conditionally automated drives. In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. Gothenburg, Sweden, pp. 400–405. doi: 10.1109/IVS.2016.7535417.

Schmidt, S. F. and Conrad, B. (1970). *Motion Drive Signals for Piloted Flight Simulators*. Tech. rep. NASA CR-1601. National Aeronautics and Space Administration, Ames Research Center.

Schrapel, M. and Vinel, A. (2024). A Highway Driving Simulator for Vehicle Platooning Applications. In: *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*. New York, NY, 536–539. doi: 10.1145/3701571.3703375.

Schwarz, C., Gates, T., and Papelis, Y. (2003). Motion characteristics of the national advanced driving simulator. In: *Proceedings of the Driving Simulation Conference 2003 North America*. Dearborn, MI.

Schwarzhuber, T., Graf, M., and Eichberger, A. (2021). Investigations on Model Predictive Control Objectives for Motion Cueing Algorithms in Motorsport Driving Simulators. In: *2021 IEEE Intelligent Vehicles Symposium (IV)*. Nagoya, Japan, pp. 49–54. doi: 10.1109/IV48863.2021.9575334.

Schwarzhuber, T., Von Schleinitz, J., Geiser, H., Graf, M., and Eichberger, A. (2020). Identification of Drivers' Controlled Stimuli in Nonlinear Vehicle Dynamics Driving Simulation. In: *Proceedings of the Driving Simulation Conference 2020 Europe*. Antibes, France, pp. 45–52.

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. In: *Psychometrika* 74 (1), pp. 107–120. doi: 10.1007/S11336-008-9101-0.

Sivak, M. (1996). The information that drivers use: is it indeed 90% visual? In: *Perception* 25 (9), pp. 1081–1089. doi: 10.1068/p251081.

Sivan, R., Ish-Shalom, J., and Huang, J.-K. (1982). An Optimal Control Approach to the Design of Moving Flight Simulators. In: *IEEE Trans. Syst., Man, Cybern.* 12 (6), pp. 818–827. doi: 10.1109/TSMC.1982.4308915.

Stahl, K., Abdulsamad, G., Leimbach, K.-D., and Vershinin, Y. A. (2014). State of the art and simulation of motion cueing algorithms for a six degree of freedom driving simulator. In: *17th IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Qingdao, China, pp. 537–541. doi: 10.1109/ITSC.2014.6957745.

Stassen, H. G., Johannsen, G., and Moray, N. (1990). Internal Representation, Internal Model, Human Performance and Mental Workload. In: *Automatica* 26 (4), pp. 811–820. doi: 10.1016/S1474-6670(17)53877-X.

Stewart, D. (1965). A Platform with Six Degrees of Freedom. In: *Proceedings of the Institution of Mechanical Engineers* 180 (1), pp. 371–386. doi: 10.1243/PIME_PROC_1965_180_029_02.

Stratulat, A., Roussarie, V., Vercher, J., and Bourdin, C. (2011). Does tilt/translation ratio affect perception of deceleration in driving simulators? In: *Journal of Vestibular Research: Equilibrium & Orientation* 21 (3), pp. 127–139. doi: 10.3233/VES-2011-0399.

Stroosma, O., van Paassen, M., and Mulder, M. (2003). Using the SIMONA Research Simulator for Human-machine Interaction Research. In: *AIAA Modeling and Simulation Technologies Conference and Exhibit*. Austin, Texas. doi: 10.2514/6.2003-5525.

Telban, R. J., Cardullo, F. M., and Kelly, L. C. (2005). *Motion Cueing Algorithm Development: New Motion Cueing Program Implementation and Tuning*. Tech. rep. NASA/CR-2005-213746. State University of New York.

Terken, J. M. B. and Pfleging, B. (2020). Toward Shared Control Between Automated Vehicles and Users. In: *Automotive Innovation* 3, pp. 53–61. doi: 10.1007/s42154-019-00087-9.

Trizano-Hermosilla, I. and Alvarado, J. M. (2016). Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. In: *Frontiers in Psychology* 7, p. 769. doi: 10.3389/fpsyg.2016.00769.

Turner, M. and Griffin, M. (2000). Motion sickness in public road transport: The effect of driver, route and vehicle. In: *Ergonomics* 42 (12), pp. 1646–1664. doi: 10.1080/001401399184730.

United Nations (2018). *World Urbanization Prospects: The 2018 Revision*. Tech. rep. ST/ESA/SER.A/420. Department of Economic and Social Affairs, Population Division. New York: United Nations.

Valente Pais, A. R., Pool, D. M., de Vroome, A. M., van Paassen, M. M., and Mulder, M. (2012). Pitch motion perception thresholds during passive and active tasks. In: *Journal of Guidance, Control, and Dynamics* 35 (3), pp. 904–918. doi: 10.2514/1.54987.

Valente Pais, A. R., Wentink, M., van Paassen, M. M., and Mulder, M. (2009). Comparison of Three Motion Cueing Algorithms for Curve Driving in an Urban Environment. In: *Presence: Teleoperators & Virtual Environments* 18 (3), pp. 200–221. doi: 10.1162/pres.18.3.200.

van der El, K., Padmos, S., Pool, D. M., van Paassen, M. M., and Mulder, M. (2018). Effects of Preview Time in Manual Tracking Tasks. In: *IEEE Transactions on Human-Machine Systems* 48 (5), pp. 486–495. doi: 10.1109/THMS.2018.2834871.

van der Ploeg, J. R., Cleij, D., Pool, D. M., Mulder, M., and Bülthoff, H. H. (2020). Sensitivity Analysis of an MPC-based Motion Cueing Algorithm for a Curve Driving Scenario. In: *Proceedings of the Driving Simulation Conference 2020 Europe*. Antibes, France, pp. 37–44.

Wagener, N., Russ, F., Beckmann, J., Legran, P., and Eckstein, L. (2023). ika's highly dynamic driving simulator. In: *Product Solution Book of the Driving Simulation Conference 2023 Europe*. Antibes, France, pp. 103–104.

Wang, X., Ye, X., Hurwitz, D., Wang, X., Schost, C., and Noyce, D. (2023). Effect of Motion Cues on Highway Driving Performance in Simulated Driving. In: *Transportation Research Record* 2677 (6), pp. 93–112. doi: 10.1177/03611981221144300.

Weiss, E. and Gerdes, J. C. (2023). High Speed Emulation in a Vehicle-in-the-Loop Driving Simulator. In: *IEEE Transactions on Intelligent Vehicles* 8 (2), pp. 1826–1836. doi: 10.1109/TIV.2022.3162549.

Wertheim, A., Bos, J., and Krul, A. (2001). *Predicting Motion Induced Vomiting from Subjective Misery (MISC) Ratings Obtained in 12 Experimental Studies*. Tech. rep. TNO-TM-01-A066. TNO Human Factors Research Institute.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., and Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? In: *Journal of animal ecology* 75 (5), pp. 1182–1189. doi: 10.1111/j.1365-2656.2006.01141.x.

Wijayaratna, K. P., Cunningham, M. L., Regan, M. A., Jian, S., Chand, S., and Dixit, V. V. (2019). Mobile phone conversation distraction: Understanding differences in impact between simulator and naturalistic driving studies. In: *Accident Analysis & Prevention* 129, pp. 108–118. doi: 10.1016/j.aap.2019.04.017.

Wong, L., Meyer, G., Timson, E., Perfect, P., and White, M. (2012). Objective and subjective evaluations of flight simulator fidelity. In: *Seeing and Perceiving* 25 (0), pp. 91 –91. doi: 10.1163/187847612X647108.

Xu, J, Min, J, and J, H. (2018). Real-time eye tracking for the assessment of driver fatigue. In: *Healthcare Technology Lettetters* 5 (2), pp. 54–58. doi: 10.1049/htl.2017.0020.

Xue, H., Ballo, F., Previati, G., and Gobbi, M. (2024). Eco-driving of Electric Vehicles - Objective and Subjective Evaluation of Passenger Comfort by a Dynamic Driving Simulator. In: *IEEE Transactions on Vehicular Technology*, pp. 1–12. doi: 10.1109/TVT.2024.3461166.

Zhan, W., Liu, C., Chan, C.-Y., and Tomizuka, M. (2016). A non-conservatively defensive strategy for urban autonomous driving. In: *Proceedings of the IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. Rio de Janeiro, Brazil, pp. 459–464. doi: 10.1109/ITSC.2016.7795595.

Zhao, L., Nybacka, M., Drugge, L., Rothhämel, M., Habibovic, A., and Hvitfeldt, H. (2024). The Influence of Motion-Cueing, Sound and Vibration Feedback on Driving Behavior and Experience—A Virtual Teleoperation Experiment. In: *IEEE Transactions on Intelligent Transportation Systems* 25 (8), pp. 9797–9809. doi: 10.1109/TITS.2024.3353465.

Zöller, I. M. (2015). "Analyse des Einflusses ausgewählter Gestaltungsparameter einer Fahrsimulation auf die Fahrerverhaltensvalidität". PhD Dissertation. Technical University of Berlin.

Acknowledgements

Intentionally left blank.

Curriculum Vitæ

Marius Johannes Claus KOLFF

27/06/1994 Born in Werkendam, The Netherlands.

Work Experience

since 2025 Post-Doctoral Researcher
Delft University of Technology
Section of Intelligent Vehicles

2023–2025 Data Scientist
BMW Group

2020–2023 PhD Candidate
BMW Group

2017 Intern
BMW Group

2015–2016 Chief Engineer
Eco-Runner Team Delft

Education

2020–2025 PhD, Aerospace Engineering
Delft University of Technology
Section of Control & Simulation

2016–2019 MSc, Aerospace Engineering
Delft University of Technology
Section of Control & Simulation

2012–2015 BSc, Aerospace Engineering
Delft University of Technology

List of Publications

Journal Publications

7. **M. Kolff**, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder (2025) - *Reliability and Models of Subjective Motion Incongruence Ratings in Rural Driving Simulations.* To be submitted.
6. **M. Kolff**, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder (2025) - *Reliability and Models of Subjective Motion Incongruence Ratings in Highway Driving Simulations.* To be submitted.
5. **M. Kolff**, K. van der El, D.M. Pool, and M. Mulder (2025) - *Modeling and Identification of Driver Steering and Gaze Behavior under Visual Occlusion.* To be submitted.
4. R. Jacumet, **M. Kolff**, J. Venrooij, S. Wagner, M. Schwienbacher, D. Wollherr, M. Leibold, D.M. Pool, and M. Mulder (2025) - *A Review on Optimization-based Motion Cueing Algorithms for Driving Simulation.* IEEE Transactions on Intelligent Transportation Systems, under review.
3. **M. Kolff**, C. Himmels, J. Venrooij, A. Parduzi, D.M. Pool, A. Riener, and M. Mulder (2025) - *Effect of Motion Mismatches on Ratings of Motion Incongruence and Simulator Sickness in Urban Driving Simulations.* Transportation Research Part F: Traffic Psychology and Behaviour, accepted.
2. **M. Kolff**, J. Venrooij, E. Arcidiacono, D.M. Pool, and M. Mulder (2025) - *Predicting Motion Incongruence Ratings in Closed- and Open-Loop Urban Driving Simulation.* IEEE Transactions on Intelligent Transportation Systems, 26(1): pp. 517-528. doi: 10.1109/TITS.2024.3503496.
1. **M. Kolff**, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder (2024) - *Reliability and Models of Subjective Motion Incongruence Ratings in Urban Driving Simulations.* IEEE Transactions on Human-Machine Systems, 54(6): pp. 634-645. doi: 10.1109/THMS.2024.3450831.

Conference Publications

9. **M. Kolff**, R. Jacumet, S. Wagner, D. Wollherr, and M. Leibold (2025) - *A Frequency-Aware Model Predictive Control Motion Cueing Algorithm.* In: Proceedings of the Driving Simulation Conference 2025 Europe, Stuttgart, France.
8. **M. Kolff**, R. Jacumet, J. Venrooij, S. Wagner, M. Schwienbacher, M. Peller, D.M. Pool, and M. Mulder (2024) - *Motion Cueing in BMW's Driving Simulation Center: Experiences Versus Common Knowledge.* In: Proceedings of the Driving Simulation Conference 2024 Europe, Strasbourg, France.
7. **M. Kolff**, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder (2023) - *The Importance of Kinematic Configurations for Motion Control of Driving Simulators.* In: Proceedings of the 26th IEEE International Conference on Intelligent Transportation Systems 2023, Bilbao, Spain. doi: 10.1109/ITSC57777.2023.10422127.
6. **M. Kolff**, J. Venrooij, D.M. Pool, and M. Mulder (2023) - *Driving Simulator Experiment Stakeholder Perspectives on Motion Cueing Algorithm Quality.* In: Proceedings of the Driving Simulation Conference 2023 Europe, Antibes, France.
5. J. Eppink, **M. Kolff**, J. Venrooij, D.M. Pool, and M. Mulder (2023) - *Probabilistic Prediction of Longitudinal Driving Behaviour for Driving Simulator Pre-Positioning.* In: Proceedings of the Driving Simulation Conference 2023 Europe, Antibes, France.
4. **M. Kolff**, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder (2022) - *Motion Cueing Quality Comparison of Driving Simulators using Oracle Motion Cueing.* In: Proceedings of the Driving Simulation Conference 2022 Europe, Strasbourg, France.
3. **M. Kolff**, J. Venrooij, M. Schwienbacher, D.M. Pool, and M. Mulder (2021) - *Quality Comparison of Motion Cueing Algorithms for Urban Driving Simulations.* In: Proceedings of the Driving Simulation Conference 2021 Europe, Munich, Germany.
2. **M. Kolff**, J. Venrooij, D.M. Pool, and M. Mulder (2020) - *Comparison of Motion Cueing Algorithms in a Virtual Test Environment.* In: Proceedings of the Driving Simulation Conference 2020 Europe, Antibes, France.
1. **M. Kolff**, K. van der El, D.M. Pool, M.M. van Paassen, and M. Mulder (2019) - *Approximating Road Geometry with Multisine Signals for Driver Identification.* In: Proceedings of the 14th IFAC Symposium on Analysis, Design, and Evaluation of Human Machine Systems HMS 2019, Tallinn, Estonia. doi: 10.1016/j.ifacol.2019.12.082.

Patents

1. M. Buchner, J. Wolters, F. Bade, and **M. Kolff** (2020) - *Method, Device, Computer Program and Computer Program Product for Detecting the Attentiveness of the Driver of a Vehicle.* DE102018214935A1