MSc thesis in Mechanical Engineering

# Validation of a Moving Base Driving Simulator for Motion Sickness Research

T. M. W. Talsma

Delft University of Technology

# Validation of a Moving Base Driving Simulator for Motion Sickness Research

by

T. M. W. Talsma

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on January 27, 2022


| | | |
|---|---|---|
| Student number: | 4365631 | |
| Project duration: | May 4, 2021 – January 27, 2022 | |
| Thesis committee: | Dr. K. N. De Winkel | Delft University of Technology, supervisor |
| | Ir. O. Hassanain | Cruden, supervisor |
| | Dr. ir. R. Happee | Delft University of Technology |
| | Dr. ir. D. Pool | Delft University of Technology |


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

TU Delft

## Abstract

Higher levels of automation in driving may allow drivers to engage in other activities, but may also increase the likelihood of Motion Sickness (MS). The exact causes of MS are not well understood, and various susceptibility factors (e.g. age, gender, ethnicity) can cause large individual differences. To better understand and predict MS, it is ideally studied in a safe and controlled environment, such as a driving simulator. However, the validity of driving simulator studies on MS as a proxy for on-road studies with real vehicles has not been properly evaluated. We conducted an experiment where the temporal aspects and symptom profiles of MS in a real-road driving scenario are compared to Simulator Sickness (SS) in a reproduction of this scenario in a motion-base driving simulator. A cohort of 25 participants was exposed to both the car and the simulator conditions. The scenarios consisted of sections of provocative (slaloming, stop-and-go) and normal driving. Sickening stimuli of the simulator were similar to the car accelerations in design ($r = 0.51$) but different in outcome ($r = 0.27$) as a result of motion cueing. MIsery Scale (MISC) scores on a 30 s interval, post-experiment Motion Sickness Assessment Questionnaire (MSAQ) scores, Galvanic Skin Response (GSR) and Electrogastrography (EGG) data were collected. Results showed significant correlations between the car and simulator conditions for 3 out of 4 MSAQ symptom categories ($0.48 < r < 0.73$, $p < 0.02$) and a relation ($r = 0.57$, $p = 0.004$) for individual sensitivity to sickness. Sickness onset times did not differ between the car and the simulator [$F(1,308) = 4.80$, $p = 0.029$], after Bonferroni corrections had been applied. Both MS and SS increased and decreased as a result of the driving style, with the effect being larger in the car condition, than in the simulator (for MISC [$F(1,248) = 19.15$, $p = 0.000$] and for GSR [$F(1,230) = 5.55$, $p = 0.019$]). Results from all four measures indicate that severity of sickness was higher in the car as compared to in the simulator. EEG responses did not fully show expected outcomes. However, the signal quality was limited and dedicated EGG equipment may yield different results. Because individual sensitivity and temporal aspects of SS and MS were similar between the car and simulator but different in magnitude, we conclude relative validity for the simulator. As the human vestibular system, a prominent contributor in causing sickness, is solely sensitive to accelerations, we attribute the difference in magnitude due to downscaling of the vehicle motion in the simulator. In order to obtain absolute validity, either extensive training or considerable technological advances may be necessary.



## Acknowledgements

I would firstly like to thank Ksander de Winkel who was always available on the other side of the virtual meeting for help, but also challenged me to find out a great deal by myself and I definitely learned a lot from that. And Riender Happee, who would always encourage me to dig deeper into the subject and get the most out of it. Thank you Tugrul, Jork and Varun, for troubleshooting whenever I was struggling with my experiment and equipment. The enthusiasm, optimism and faith in me of everyone in this Intelligent Vehicles group, was inspiring. Thank you Jelte Bos for our philosophical conversations, I enjoyed them and you definitely helped me understand the subject better. A special thanks to all the family, friends and other kind angels, who have taken the commitment and effort to participate in my experiments. This research would not have been as valuable without the data I collected from you. Especially to those that became fairly motion sick, I am very grateful for your contribution. It was a pleasure and an enrichment to be able to perform part of the experiment in the Cruden environment. This way I could learn from the expertise that is present at this company and the willingness to help, from everyone around there. Thank you Nico van Leeuwen for assisting me in "how to thesis" and in understanding the simulator. I would like to specifically thank my supervisor Omar Hassanain for always being available to think along, reflect, and explain mathematical and technical concepts with all the patience in the world. I am very grateful to have had your support and am still amazed at how you happen to know everything. Lastly, thank you Milan Izarin. This page is too short, to sum up how much you have helped, comforted and supported me throughout the last five years. I am mostly happy to write that this thesis report not only marks the end of my Mechanical Engineering Master, but also the end of my time as a student. I have enjoyed the ride.

# Contents


Nomenclature 1

1 Introduction 3
- 1.1 Problem statement . . . 3
- 1.2 Simulator validation studies . . . 4
- 1.3 Motion Sickness. . . . 4
- 1.4 Quantifying sickness . . . 6
- 1.5 Outline . . . 6

2 Theory and design 7
- 2.1 Simulator Sickness . . . 7
  - 2.1.1 Differentiation . . . 7
  - 2.1.2 Spectrum . . . 8
  - 2.1.3 Fidelity theory . . . 9
- 2.2 Research question . . . 10
- 2.3 Hypotheses . . . 11
- 2.4 Experimental design . . . 11
  - 2.4.1 Route . . . 12
  - 2.4.2 Motion stimuli . . . 12
  - 2.4.3 Task . . . 14
  - 2.4.4 Simulator design . . . 15

3 Methods 19
- 3.1 Participants . . . 19
  - 3.1.1 Ethics statement . . . 19
- 3.2 Apparatus . . . 19
  - 3.2.1 Car . . . 19
  - 3.2.2 Simulator . . . 20
- 3.3 Procedure . . . 21
- 3.4 Measures of sickness . . . 21
  - 3.4.1 MISC. . . . 21
  - 3.4.2 MSAQ . . . 22
  - 3.4.3 EGG . . . 23
  - 3.4.4 GSR . . . 24
- 3.5 Data analysis . . . 25

4 Results 27
- 4.1 Motion stimuli . . . 27
  - 4.1.1 Vestibular motion . . . 27
  - 4.1.2 Visual motion . . . 29

4.2 MISC . . . 30
4.2.1 Individually averaged MISC . . . 31
4.2.2 Time to sickness . . . 32
4.2.3 Mean MISC group response . . . 33
4.2.4 MISC rate . . . 34
4.3 MSAQ . . . 35
4.3.1 Overall MSAQ score . . . 35
4.3.2 Subscale MSAQ score . . . 36
4.4 GSR . . . 38
4.5 EGG . . . 40

5 Discussion 43
5.1 Are symptoms as a result of SS, different from classical MS?. . . . 43
5.2 Does a relationship exist between individual sensitivity to MS and SS? . . . 44
5.3 Are temporal aspects of sickness different in the car and in the simulator? . . . 44
5.3.1 Time to onset of sickness. . . . 44
5.3.2 Increase and decrease of sickness . . . 45
5.3.3 Amplitude of sickness . . . 45
5.4 Can we study MS in a driving simulator? . . . 46
5.5 Limitations . . . 47

6 Conclusion 49
6.1 Research acknowledgements . . . 49

A Fidelity and simulator sickness: An uncanny valley 61
A.1 Method . . . 62
A.2 Results . . . 63
A.3 Discussion . . . 64

B Misery Scale rating and Sudoku 67

C Motion Sickness Assessment Questionnaire 69

# Nomenclature

**Abbreviations**

| | |
|---|---|
| (A)SS | (Absolute) Simulator Sickness |
| ANOVA | Analysis of Variance |
| ART | Aligned Rank Transform |
| CNS | Central Nervous System |
| CS | Carsickness |
| DF | Dominant Frequency |
| DoF | Degree of Freedom |
| EDA | Electrodermal Activity |
| EGG | Electrogastrography |
| FoV | Field of View |
| GSR | Galvanic Skin Response |
| HMD | Head-Mounted Display |
| MCA | Motion Cueing Algorithm |
| MISC | Misery Scale |
| MPC | Model Predictive Control |
| MS | Motion Sickness |
| MSAQ | Motion Sickness Assessment Questionnaire |
| MSSQ | Motion Sickness Susceptibility Questionnaire |
| SCS | Simulated Carsickness |
| SI | Sickness Incidence |
| SIS | Simulator Induced Sickness |
| SSQ | Simulator Sickness Questionnaire |
| VIMS | Visually Induced Motion Sickness |
| VR | Virtual Reality |

**Symbols**

| | |
|---|---|
| $R_{XY}$ | Cross-correlation |
| $C_{xy}$ | Spectral coherence |
| RMS | Root Mean Square |
| std | Standard deviation |

# 1 Introduction

## 1.1 Problem statement

Autonomous vehicles are expected to minimize environmental impact, increase passenger comfort and enhance safety [1]. The introduction of autonomous functionality is divided into six incremental levels, SAE levels [2]. SAE level 0 refers to full manual control and the driver being completely in control of the vehicle whereas level 5 signifies full autonomy, changing the driver to a user. Level 1 systems, referring to systems that allow taking the feet of the pedals and level 2 systems, referring to systems that allow releasing the hands from the steering wheel, have already been introduced on the roads. Currently, automotive manufacturers are hoping to get vehicles with level 3 automation, defined as the possibility for the driver to also take his or her eyes off the road and engage in non-driving tasks, on the roads very soon [3].

A potential hindrance to the widespread adoption of automated vehicles of all levels is the occurrence of *Motion Sickness* (MS) [4]. Risk of occurrence of MS is not only proven to be higher even in low levels of automation, also its presence is more severe than when travelling in human-operated vehicles [5]. Passengers, or users in this case, have an increased risk of MS compared to drivers, as a result of the absence of control [6] and a restricted out-the-window view [7]. This is effect furthermore supported by the fact that users of future self-driving cars have largely indicated their preference to reading texts, working or watching on-board screens [8]. When being a passenger during a car drive, all of the aforementioned tasks would increase the onset and severity of MS even more.

MS is a syndrome that typically emerges from exposure to movements by vehicles and causes discomforting symptoms such as nausea. It is studied that around 60% of the population has experienced MS symptoms from car travel, whereas approximately one third has vomited in cars before the age of 12 [9]. Additionally, multiple experiments have shown that MS can cause not only significant discomfort but also performance degradation (of driving tasks) during motion [10, 11], hours [12, 13] or even days [14] after exposure. Although several theories have been proposed, the exact cause of MS is not known. For acceptance [15] and widespread use of future autonomous vehicles [5] and for our own comfort and safety, it is thus important that we find the cause of and design solutions to this problem.

To study autonomous driving comfort and to understand and predict how and when MS develops, ideally driving tests are performed under everyday driving conditions. However, considering the high costs of naturalistic driving studies, the impossibility of creating identical experimental conditions on repeated experimental trials and for multiple participants, and safety and ethical concerns that arise from experimenting with humans in real traffic, such studies are often not feasible. Here, driving simulators could present a solution. A driving simulator is a tool for research or training purposes designed to approximate actual driving scenarios and consists of a stationary or motion-based platform,

and a visualization system. A simulator offers i.a. controllability over external variables and reproducibility [16], but more importantly, it ensures the safety of participants as testing with MS has an often unknown time course and effect consequently experiments can be terminated immediately. Two limitations with simulator studies are, firstly, that it is uncertain to what extent findings of MS generalize to real vehicles [17], i.e. the validity of a simulator is questioned. The second limitation pertains to the often occurring *Simulator Sickness* (SS). This second issue frequently impacts quality measurement with for example dropouts [18] and is dependent on the specific simulator or particular immersion. SS complicates furthermore to what extent comfort-related findings from the simulator reflect the related but not identical MS. The first limitation will be elaborated on in this section.

## 1.2 Simulator validation studies

**The goal of this study is to determine the validity of the driving simulator for MS research.** However, the term *validity*, is not precisely defined. The validity of a driving simulator often refers to the ability to accurately represent real-world driving in terms of research outcomes [19]. Generalization of validity is limited because this is applicable to a particular task or simulator. For almost every validation study, major differences in the type of simulators, subjects, protocols and methods exist. Several methods [20, 21] to assess the validity of a simulator have been proposed, of which the most common assessment for validity distinguishes between absolute and relative validity [22]. *Absolute validity* occurs when research outcomes from a simulator match in absolute terms with outcomes obtained in a real vehicle, thus when signals obtained from two scenarios show no significant differences. *Relative validity* occurs when outcomes from both environments show the same patterns or (main) effects but differ for example in magnitude or onset time of MS.

Validation of simulators as compared to real driving behaviour has been researched extensively regarding driving performance measures, e.g. driving speed [19]. Research is very limited, however, when we are interested in subjective experiences such as comfort. Bellem et al. (2017) [17] validated parameters of driving *comfort* for a passive driving task by means of comparison between an on-road drive and several moving-base simulator configurations. The authors showed validity for one configuration and concluded that the moving-base simulator could be a useful research tool to study driving comfort, but that comfort is highly dependent on scaling factor parameters. Similarly, Galante et al. (2018) [23] validated a driving simulator compared to on-road measures regarding driver *workload*. Their research found mixed support for the validity, although more familiarization with the simulator might increase its validity. To the best of our knowledge, the validity of simulators to study MS as an alternative to naturalistic driving has only been evaluated in a single study, by Kuiper et al. (2019) [24]. The authors showed that MS could be researched using a motion-base simulator, as the average level and increase of MS over time was found comparable to population group responses. However, the generalizability of their results is questionable, as a pure lateral (predictable) manoeuvre is conducted and all visual cues are removed. Moreover, comparison of sickness is done with statistically calculated data instead of on-road driving and no objective parameters representing MS are validated. Consequently, it is as of yet to be determined whether results of MS experiments obtained from a simulator still hold true in real vehicles.

## 1.3 Motion Sickness

Motion sickness or *Carsickness* (CS) is a syndrome, consisting of a set of symptoms (e.g. headaches, sweating, dizziness, fatigue, nausea, vomiting) that emerge from exposure to certain movements such as abrupt, periodic or unnatural accelerations. These physical stimuli are sensed by the semicircular canals or the otolith organs within the vestibular system of an individual. These three canals detect angular acceleration of the head around the vertical, lateral and fore-aft axis. The otolith organs detect changes in acceleration in the horizontal and vertical direction but are unable to distinguish between sensed gravitational and inertial accelerations. To improve multisensory perception, the central nervous system (CNS) combines information from the vestibular system with the information of the visual system and the proprioceptive system.

The exact cause of MS is unknown and several theories exist [25–29]. The *Postural Instability Theory* by Riccio and

Stoffregen (1991) [29] explains MS as a result of prolonged postural instability. Unfamiliar situations would make it difficult to maintain a natural body posture and relearning to control balance is necessary in order to reduce sickness. The most prominent theory, however, is the Sensory Conflict or *Neural Mismatch Theory* [27] by Reason and Brand (1975). Several perspectives on how this conflict influences MS exist [30–32]. The general theory explains the conflict between sensory inputs of self-motion and self-tilt and expected neural inputs on the basis of previous experience, the neural storage. While continuous exposure to similar conditions updates the neural storage and reduces the mismatch and sickness symptoms, changes in conditions create a new mismatch thus additional sickness. Conflicts often occur between modalities (e.g. visual-vestibular), but can also occur within modalities (e.g. visual-visual). A method to mitigate MS consistent with predictions of this theory is to minimize conflicts by for instance focusing on a reference point along the horizon, rather than reading a book when travelling by car [33, 34]. A limitation of this theory is that no straightforward method exists to determine the magnitude of the conflict for a specific combination or particular provocative motion condition. Consequently, it is not yet possible to identify conflicting factors in environments such as simulators [35].

Motion is considered the number one contributing factor to MS and a complex dependency on time, frequency, amplitude and direction of movement with MS exists. For translations, MS is dependent on frequency and acceleration and this relation is best described by McCauley et al. (1976) [36]. Their research shows a peak sickening frequency of approximately 0.2 Hz, also proven by other studies [37–40], and higher sickness susceptibility following higher accelerations. The latter explains why modes of transportation with higher accelerations (e.g. cars as compared to trains) or provocative driving typically cause more sickness [36, 41]. Although several sickening combinations of and with rotational stimuli exist (e.g. Cross-Coupled Coriolis Perturbation), these movements alone are typically less sickening than translational stimuli [42]. Earth-vertical yaw rotation, such as typical rotation on curved tracks, is on its own not sickening [43]. Pitch and roll motion can be sickening [30, 42], but these stimuli are generally limited in normal car driving behaviour. Besides physical, also visual motion can contribute to MS. Whereas classical MS is attributed to sickness that occurs in response to conditions where motion is felt, but possibly not (correctly) seen, *Visually Induced Motion Sickness* (VIMS) is implied when motion is seen, but not (correctly) felt [44, 45]. VIMS for example occurs when visiting a large screen cinema or driving a fixed-base simulator. In these cyber-environments, inertial motion is absent but the human vestibular system as described above still senses gravitational motion. As the vestibular system is unable to distinguish between inertial and gravitational motion, the sensed motion remains constant and causes a conflict with the conflicting, visually sensed motion.

Furthermore, a large inter-subject vulnerability with respect to MS exists [46] and research has shown various extraneous factors affecting this susceptibility [47]. Some well-known factors differentiating MS susceptibility are gender [48], age [49], ethnicity [50] and health condition (blind people do suffer from MS [51] whereas deaf people do not [33]). Also, cognitive functioning contributes, as mental distraction and anticipation could decrease MS. A clear example of this can be seen in the fact that drivers suffer considerably less from carsickness than passengers do, irrespective of being exposed to the same motion [52].

In addition to MS and VIMS, the present technology has also necessitated the introduction of *Simulator Sickness* (SS). SS describes the discomfort that may accompany immersion in a simulator, which is principally similar to MS in terms of causes (theories and motion stimuli) and symptoms. This issue is sometimes reported by more than 80% of participants in simulator studies [53, 54]. Literature reflects some disagreement regarding the disclosure of this phenomenon. Often, SS is categorized as VIMS. However, simulators over the past years continued to become increasingly advanced. Hence, modern moving-base simulators can also provide non-visual (acceleration) cues that might induce sickness. One can thus imagine that SS could include aspects from both VIMS and traditional MS that cannot always be clearly assigned to one of the two. Moreover, SS is in the vast majority of literature simply denoted as the sickness (MS) obtained in a simulated environment. Also, this assumption acquires nuance. Although sickness in a simulator can originate from stimuli analogous to carsickness, it can additionally develop from specific technological factors introduced to compensate for the limited motion envelope and the artificial visualization modality of the

simulator. Respective literature addressing SS actually refer to something we, in the remainder of this study, will refer to as *Absolute Simulator Sickness* (ASS), describing a combination of both *Simulated Carsickness* (SCS) and *Simulator Induced Sickness* (SIS).

## 1.4 Quantifying sickness

MS and SS are multi-factorial, meaning that they elicit a wide range of symptoms that are not unique as they are often associated with other ailments as well. Consequently, there is no single reliable parameter that represents MS in a comprehensive way [55], resulting in validation of sickness in literature with numerous measures. These measures can generally be subdivided into subjective self-reported questionnaires and objective (physiological) measures. Questionnaires used to assess the degree of MS exist in one-dimensional rating scales and multi-dimensional symptom lists [42]. Examples of well-known multi-dimensional questionnaires obtained from literature are the Motion Sickness Susceptibility Questionnaire (MSSQ) [27], meant to establish past individual susceptibility to MS, the Simulator Sickness Questionnaire (SSQ) [56], specifically for SS and the Motion Sickness Assessment Questionnaire (MSAQ) [57]. The latter is used after immersion to rate how accurately (1 being not at all to 9 being severe) 16 symptoms describe the experience of subjects. An often used one-dimensional rating scale is the Misery Scale (MISC) [58]. This continuous rating method addresses a higher number (0-10) to a higher degree of sickness. The MISC is used during immersion, while the aforementioned questionnaires are used before or after immersion. Continuous rating scales convey the advantage of creating a time course of sickness and the safety of monitoring the current state of the participant. However, such a scale brings uncertainty of people to rate how they are feeling and mainly provides the order between its items, while multi-dimensional questionnaires provide insight into different physiological processes.

In the literature, also objective metrics can be related to MS and SS. There is evidence for a correlation between sickness and decreased human posture [29, 35, 59] and lower limb muscle activity [60], several frequency bands of brain activity [61], increased stomach and nerve activity [62], increase in heart rate (variability) [62, 63], eye movement and pupil dilation [62], increased skin conductance [64], decrease of blood pressure [65] and decrease in skin temperature [66]. The combination of different subjective measures and objective measures could present a wide range of the sickness time course and eliminate some of the limitations that each individual measure possesses. It is therefore suggested by several studies [67–69], to compose a subset of one- and multi-dimensional and continuous and discrete subjective measures and diverse objective physiological measures.

## 1.5 Outline

Based on what is known about validity and MS on the basis of the literature in this introduction, the research question, hypotheses and experimental design to solve the problem statement, are presented in section 2. In section 3, the research method will be explained. Section 4 and 5 will respectively feature the results and discussion of the results. The conclusion can be found in section 6. Lastly, a meta-analysis study on the relation between simulator fidelity and SS is presented in appendix A.

# 2

# Theory and design

In this chapter, the theory and experimental design necessary to determine to what extent findings of MS in a simulator generalize to real vehicles, will be addressed. This chapter first elaborates on literature regarding the second limitation of simulator validation studies, namely SS. Findings and interpretations thereof are then converted into a research proposal and the research question and hypotheses are afterwards presented and justified. Lastly, the comparison between MS and SS will be employed as a method to design a relevant empirical experiment.

## 2.1 Simulator Sickness

Similarity between MS and SS cannot only be found in elevated symptoms [56], but also in essence of origin. Although never arranged in this way, (the subcomponents of) SS could be explained by means of the Neural Mismatch Theory. Simulated Carsickness (SCS) could then originate from the conflict between sensed and expected neural inputs, and possibly modulated by a conflict between visual stimuli and the appropriate vestibular motion. Its definition therefore suggests similarity with regular MS. In contrary, Simulator Induced Sickness (SIS) could solely originate from the former perception-expectation conflict.

### 2.1.1 Differentiation

In relevant research, the term MS is almost always used interchangeably with SS. Nevertheless, findings from literature suggest SS to be different from MS regarding several aspects. Even though there is generally a lack of within-subject validation in many of the following studies [24, 70, 71], their conclusions provide a reason to question if SS is similar to MS.

*(i) Different environment.* The Simulator Induced Sickness (SIS), as a part of Absolute SS, acquired from simulator technology factors used to replicate motion and vision cues, will never occur in real life. The origin of this additional sickness is either hardware- or content dependent [72, 73]. Acknowledged artificial factors contributing to sickness in a simulator within the motion modality are, firstly, the choice of Motion Cueing Algorithm (MCA) and the tuning of its parameters [17]. Cueing is used to map the vehicle inertial motions onto the limited simulator motion space. MCAs come in various configurations, of which the baseline often exists of scaling gains, filters and tilt- and rate coordination. Generally, when the artificial motion then exceeds the human perception thresholds with a cue that is perceived as unnatural, so-called scaling or roll rate errors and false and missing cues arise [74, 75]. These errors constitute sensory conflicts and thus sickness [76]. Secondly, there are some factors related to the visualization modality. Its intensity, brightness (flicker perception) and contrast could influence SS [77, 78]. The use of stereoscopic 3D visualization as compared to 2D projection [79], the addition of a Head-Mounted Display (HMD) [59, 80] and an increased Field of View (FoV) accompanying stationary simulators [81, 82] are also proven to increase sickness. Mechanical

and visual latencies within each modality or delays between modalities furthermore constitute SS [83], as these can a different ratio in perception from the real world. The additional sickness (SIS) described here is very often underestimated, nonetheless one of the reasons why SS can be very different from MS.

Supporting this difference in environment, three more discrepancies regarding susceptibility have been found in literature. They can be summarized as a *(ii) Difference in neural store*.
*(ii.a) Different time course with respect to age*. After a peak around 10 years old, MS susceptibility monotonically decreases with increasing age [49, 52] and (consequently) with increasing years of licensure [84]. Conversely, SS increases with increasing age [84, 85] and years of driving licensure [86]. Since both age and years of licensure are often accompanied by driving experience, this difference strongly inclines the theory that the neural storage of the elderly is perfectly updated regarding perception in actual vehicles. However, a large conflict is then created in the artificial vehicle with different, yet unknown dynamics.
*(ii.b) Different reaction to control*. MS mostly affects passengers and not drivers, irrespective of being exposed to the same motion [52, 87]. SS, however, may affect drivers of driving simulators equally or more than it affects passengers [88]. On top of the visual-vestibular conflict that could be caused while being a passenger or a driver in a simulator, for the driver an extra conflict is caused. The internal model of the driver only recognizes the motion dynamics of a real vehicle and how it moves as an extension of the driver behaviour [89], which is not equal to that of a simulator.
*(ii.c) Differences due to training or repeated exposure*. The last example contributing to this difference is the fact that experienced pilots suffer more from SS than inexperienced trainee pilots, while the students suffer more easily from MS in a real aircraft after being trained on a simulator [30]. When the pilots increase their time spent on a simulator, SS decreases again as neural inputs are updated [90, 91]. These observations suggest furthermore that both sicknesses could be similar, but the simulator is (for most people) a new environment. This difference causes the need for stored inputs to be updated to the new visual cues and artificial motion cues first, in order to reduce SS again.

*(iii) Different ratio between visual and vestibular cues*. In a non-artificial environment, the CNS provides integration of multisensory information which is in accordance with one another, an indicative 1:1 ratio between visual and vestibular perception. The study by Gracio et al. (2014) [92] showed that when in a simulated environment, vestibular information requires downscaling as compared to the visual imagery in order for the scenario to be perceived as natural. The ambiguities of the visualization modality somewhat withhold the perception of speed [93, 94] and amplitude [89] as compared to how these are perceived in the real world. This difference suggests that the gain of visual motion is larger in simulated environments using artificial imagery.

*(iv) Different peak sickening frequency*. Although not empirically verified, it is assumed that the peak sickening frequency shifts from 0.2 to 0.35 Hz when displacement is fixed, such as in a driving simulator with a limited motion envelope [24]. This assumption is based on the calculations for human exposure to whole-body vibrations causing MS (ISO 2631-1 (1997)) [95].

*(v) Different clinical entities*. In a study by Kennedy et al. (2010) [70], symptom profiles between three conditions (sea, space and simulator sickness) are compared. The fixed-base SS condition shows progressively higher oculomotor disturbance scores (focusing problems, blurred vision, headache), whereas both non-artificial conditions show the highest nausea scores. The same study compared sickness for 8 different simulators, and concluded all of their profiles (oculomotor > nausea > disorientation) as similar. Another study [71] compared VIMS and SS and concluded that nausea is not the dominant symptom for either artificial condition, whereas for MS this is almost unanimously the primary sign [96]. Gavgani et al. [68, 97] did show high correlations of nausea scores between a real and artificial environment during advanced stages of sickness.

### 2.1.2 Spectrum

In order to ensure whether MS or (SIS or SCS) SS is encountered during this study and thus to know what exactly is being measured, differentiation of MS and SS is important. The terms MS, SS (SCS and SIS) and VIMS are therefore

separated in a suggested sickness spectrum in figure 2.1. The blue area represents the presence of motion stimuli and the grey area the absence of motion stimuli. The separation between the blue area and the red area denotes the difference between actual stimuli and artificial stimuli. When looking at this figure 2.1, SS can from left to right be subdivided into more specific configurations: a motion-base simulator where visuals are turned off, a motion-base simulator accompanied by a visualization system and a fixed-base simulator. All spheres have similar shapes, to indicate the possibility for similarity in all dimensions. In order to propose research to differentiate MS from SS, it is important to take previous literature regarding this differentiation into account.

The study by Gavgani et al. (2018) [68] exposed subjects to both a cybersickness (grey) scenario and a vestibular immersion in a rotational drum (blue) and concluded mostly similarities between MS and VIMS. A recent study by Bos (2021) [89] suggests that it is impossible to study actual carsickness in a fixed-base simulator, as a result of the absence of actual motion. The area in which the fixed-base simulator lies is grey and will be excluded from this research. The earlier addressed study by Kuiper et al. (2019) [24] already concluded the resemblance between MS and SS in a motion-base simulator without the visualization system and with eyes closed. The authors suggested the possibility to study carsickness with the use of this simulator configuration. In other words, they validated the simulator configuration for similarity of MS with SCS, whereby SIS was removed. The space which the motion-base without visuals covers, is therefore also blue. A combination of SIS and SCS, the complete red area, has never been compared to MS (blue area).

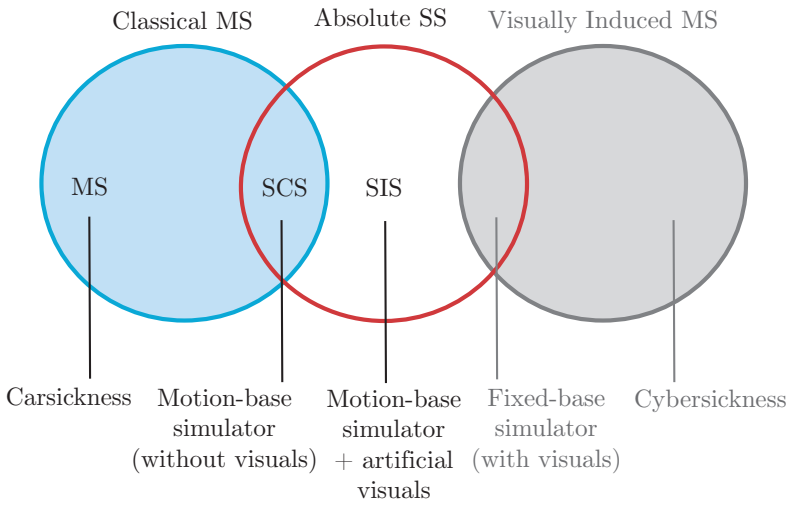


Figure 2.1: A suggestion of SS (red) and VIMS (grey) on the MS (blue) spectrum, whereby the blue area presents the presence of motion and the grey area presents the presence of visualization systems

Little is known on the exact contribution of individual simulator factors (especially SIS originating from the motion modality or the visual modality) such as latencies or false acceleration cues, to the amount of absolute SS. Vice versa, it is not yet possible to determine exactly from what specific factors sickness symptoms elicit. Hence, it is impossible to distinguish SCS from SIS in a complete, simulated environment. In order to better address differences between sickness in the real and in the artificial world, a theory is proposed on the basis of simulator fidelity.

### 2.1.3 Fidelity theory

*Fidelity* or realism refers to the extent to which (specific technical aspects of) simulators realistically emulate driving in the real world, denoted by the arbitrary classification "low", "medium" and "high" [19, 98]. The two aforementioned limitations of the use of a simulator, namely the lack of uniform validation regarding comfort issues and the sparsely

researched separation between MS and SS, are both related to the simulator fidelity. Firstly, validity (of research outcomes) of a simulator is often linked to its fidelity. There is limited empirical research that directly tests the extent to which fidelity contributes to validity [19] but it is shown that validity is generally greater in high-level fidelity simulators than mid-level or low-level simulators [99]. To produce higher levels of validity, several other studies propose increased realism or fidelity [100–102].

Secondly, a suggested method that could roughly estimate contributions of conditions to SS, also involves the simulator fidelity. As described above, many fidelity aspects of the simulator are in some way related to SS [103]. Appendix A suggests a theory on the relation between SS and simulator fidelity. In summary, the theory and accompanying figure 2.2 suggest that for multimodal systems, able to provide multiple types of sensory information (such as a motion-base simulator), the combination of SIS and SCS could approach MS. When such a multimodal simulator reaches perfect fidelity, errors in the visualization modality are minimized and realistic (perception of) motion is maximized, visualized by the decrease of SIS and increase of SCS with increasing fidelity. Thus, the theory and figure 2.2 do not only address the necessity of pursuing higher levels of fidelity to study MS, it also gives an impression of which factors are present when comparing SS to MS in artificial systems like a simulator. An elaborated version and outcomes of this meta-study can be found in appendix A. Our experimental study will aim for maximized fidelity, whereby SIS will be minimized.

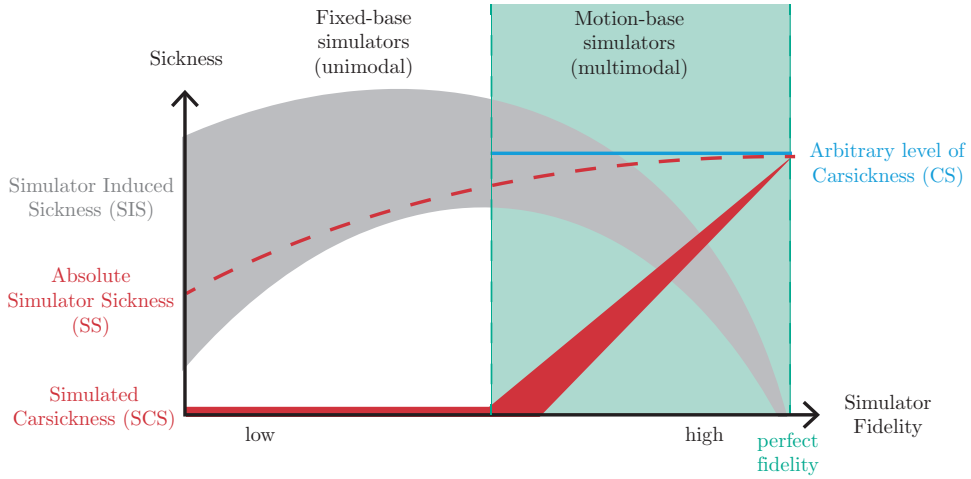


Figure 2.2: Simulator fidelity (left to right: low fidelity fixed-base simulator, high fidelity fixed-base simulator, low fidelity motion-base simulator, high fidelity motion-base simulator, perfect fidelity) versus different components of SS

## 2.2 Research question

To the best of our knowledge, the relation between SS and MS therefore the validity of simulators to study MS as an alternative to naturalistic driving, has only been evaluated by Kuiper et al. (2019) [24]. Whereas their study used an incomplete simulator setup and provided no within-subject comparison, we will perform a within-subject experiment using a realistic on-road scenario and a simulator motion base, attributed by a visualization system. With the differences between MS and SS (divided into SCS and SIS) in mind, and with the goal of (absolutely or relatively) validating the simulator for MS research, the research question states: To what degree can Motion Sickness in real road driving be approached by Simulator Sickness in a complete motion-base driving simulator? This question can be subdivided into the following subquestions: **Are symptoms as a result of SS, different from classical MS? Does a relationship exist between individual sensitivity to MS and SS?** And: **Are temporal aspects of sickness (time to onset, increase and decrease of sickness, and amplitude of sickness) different in the car and in the simulator?**

The research question is supported by the blue area (MS) versus the red area (SS) in figure 2.1. These two *scenarios*, the car and the simulator, form the main independent variable. A within-subjects experiment with a counterbalanced cohort will be conducted, whereby possible *order* effects, first MS or first SS, should be solved by randomizing conditions. The intensity of the sickness in both scenarios is manipulated by different *driving styles*. This additional independent variable, consisting of normal and provocative driving, is added to generate sufficient sickness. Also, creating a time signal or trend of sickness that has a distinctive shape as a result of different driving styles, makes a comparison between scenarios more founded. For validation completeness, for compensation of disadvantages that each of the above-mentioned measures exhibits, and as a result of the inter-variable nature of MS, this study will compare a subset of objective and subjective measures, the dependent variables.

## 2.3 Hypotheses

As described above, SS can originate from a conflict similar to MS. However, some notable differences between both phenomena exist. It is expected that values for absolute SS are generally *lower* than for MS obtained in a real car. Firstly, the (medium-high) multimodal simulator used in this study has a limited motion envelope as compared to the car and is possibly not able to produce equal amplitudes of accelerations. SCS will thus be lower than MS. This fact will be elaborated on in the experimental design and methods of this study. Secondly, by diverting attention from the (ambiguities) of the system, SIS elicited from the visualization modality can be minimized. When drawing an imaginary vertical line within the multimodal area of figure 2.2, again lower values for SS as compared to MS are expected. Lastly, this expectation can also be supported by means of *Differences (i)-(iv)*. SS is hypothesized to be lower than MS in similar conditions as false cues as a result of *(i)* are minimized, the effect coming from *(ii)* can be controlled by selection of the subject population and observation of their simulator experience and *(iii)* and *(iv)* suggest lower SS levels as compared to MS. *Difference (v)* indicates that symptoms might show similarity between both scenarios, but sickness profiles could be different. Based on (interpretations of) literature, it is hypothesized that:

**1. At least relative validity can be obtained when comparing subjective and objective measures of SS and MS.**
MISC will be consulted to compare individual sickness sensitivity, number of dropouts, and temporal aspects of sickness over time. MSAQ is consulted to obtain sickness symptoms and profiles (order of sickness symptoms). As validated indicators of sickness, EGG and GSR will be consulted to compare sickness trends. Each measure is expected to show relative validity between the car and the simulator.

**2. Time courses for SS and MS in both scenarios are (relatively) similarly affected by driving style.**
By creating a route that consists of normal and provocative driving, a pattern can be distinguished in the rate of sickness, for apparent (visual) comparison. It is expected for both scenarios that the rate of sickness (measures) is higher during provocative driving [41] than during normal driving as some recovery occurs [58].

**3. Order of the perceived scenarios has no interaction effect on subjective sickness outcomes**.
By conducting both scenarios for participants on separate days with several days in between in counterbalanced (randomized) order, no effect of the order of perceived scenarios should be found.

## 2.4 Experimental design

Based on figure 2.1, a validation of the simulator for MS research ideally consists of (at least) two experimental conditions to retrieve the relationship between SS and MS. The first scenario (A) relates to the blue area of figure 2.1 and addresses carsickness obtained during a real-road drive in a (partly automated) car. The second scenario (B) relates to the red area of figure 2.1 and addresses sickness obtained in a simulated environment with similar (sickening) stimuli. It is decided to exclude from this study the small blue area referring to a motion-base simulator without visuals (solely SCS) and the grey area, a fixed-base simulator with visuals, as these have already been studied [24, 89]. An experimental design is proposed in figure 2.3 whereby participants undergo both scenarios in randomized (balanced) order.

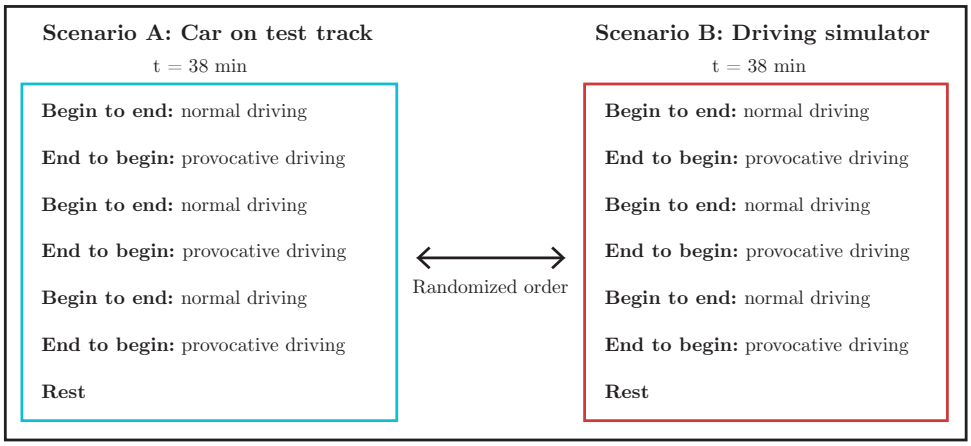


Figure 2.3: Experimental within-subject design consisting of two scenarios in randomized order

Such a within-subjects experiment ensures statistical power over a between-subjects design, due to the assumption that individual participant behaviour is usually consistent across conditions and participants thus act as their own control scenario [104]. This design also accounts for the large inter-participant susceptibility to MS, which is said to differ significantly per individual [46]. Since no effect sizes are present in literature as no such comparison between SS and MS has been made, no sample size calculation can be performed beforehand. Instead, the most acceptable way of determining the effect size based on the convention is using a 10-samples-per-variable ratio [105], resulting in a minimum of 20 participants.

### 2.4.1 Route

The route that will be driven and artificially created for the simulator condition is located in Valkenburg in the South-Holland province in the Netherlands and shown in figure 2.4. Courtesy of Rijkswaterstaat, this route is a closed track surrounding the previous military airport. The experiment could not be interfered with by other road users or unpredictable traffic scenarios such as stoplights, thus creating optimal reproducible conditions. Pilot drives performed at other possible cooperative locations in the Netherlands, demonstrated to be mainly suitable for predictable motion exposures, such as continuous slaloming. Studies showed that such predictable motion reduces sickness [6, 87] and natural vehicle motion rather than repetitive motion is expected to result in higher sickness levels [106]. The track in Valkenburg is chosen for this study as it offers the most diverse and longest route and replicates most closely actual everyday cardriving. Live surveillance and monitoring by means of cameras is present at this location, so that the experiments can be conducted safely.

When the route is driven begin-end, we refer to *normal driving* and the path in figure 2.4 taken from point I to point XIII. When referring to end-begin, it is meant that the route driven from point XIII to point I and the driving style is *provocative driving*. The one-way route as shown in figure 2.4 is 2.5 km long and has an approximate road width of 3 m. At both ends, a three-point turn will be performed. Driving the route as shown in the figure once (without turning) lasts approximately 5 min when driving at an appropriate speed of 30 km/h. The route will at most be driven 3 times back and forth which will last for approximately 33 min. Not only is this amount of time typical for a daily drive [108], also from comparable driving studies is concluded that 33 minutes is sufficient to obtain usable levels of MS [36, 109, 110] as well as SS [37, 59] for frequencies and accelerations obtained in normal driving behaviour.

### 2.4.2 Motion stimuli

Raw acceleration data recorded with an accelerometer during normal driving is analyzed in the time-domain first. Figure 2.5 (top) shows longitudinal (red) and lateral (green) acceleration data. Turning points I-XIII from the route


Figure 2.4: The 2.5 km route section driven in both scenarios that is located at Valkenburg Airport. The map is obtained from Space Office Netherlands [107].

(figure 2.4) are added by means of vertical lines. The grey area represents the three-point turning manoeuvre, directly after which the provocative manoeuvre will be performed. Accelerations in the horizontal plane typically centre around 0 m/s$^2$. Heave accelerations, which have been precluded from figure 2.5 for intelligibility and because the route is mostly flat, centre around 9.8 m/s$^2$ attributed to gravitational acceleration. Roll and pitch of the original route are centred around 0 deg/s$^2$ while the yaw varies due to cornering behaviour of the car along with the points I-XIII. During this normal driving section, a constant speed of 30 km/h is maintained with the help of the cruise control function. Consequently, turning behaviour along the route results in lateral accelerations of approximately 4 m/s$^2$.

Even though the original route shown in figure 2.5 (top) resembles the real world regarding several dimensions, it may not be sickening enough to obtain sufficient data. Half of the total excitation is therefore induced with additional, provocative driving [111]. As addressed above, a frequency of 0.2 Hz and an adequate amount of acceleration amplitude result in the highest MS sensitivity [36, 42]. It is studied that out of three manoeuvres (braking, shallow curved and lane changes), braking showed the strongest correlation with self-reported sickness scores [76]. Also, research concluded that lane-changes and deceleration-acceleration behaviour belong to the most frequently occurring manoeuvres in future automated driving [112]. Hence, the provocative driving sections will consist of slaloming (representing lane changes) and stop-and-go manoeuvres (representing deceleration and acceleration behaviour behind a slower vehicle) both at 0.2 Hz. A metronome is used to drive the manoeuvres at this fixed frequency. Figure 2.5 (bottom) shows the new provocative section of the route. It can be seen from this raw acceleration data that even though the car is driven manually, amplitude and frequency of accelerations in the horizontal plane are consistent throughout the time of the drive. Slalom acceleration motion (figure 2.5 lateral, green) is obtained with a constant driving speed of 20 km/h. Stop-and-go acceleration behaviour (figure 2.5 longitudinal, red) is executed at speeds between 10 and 20 km/h.

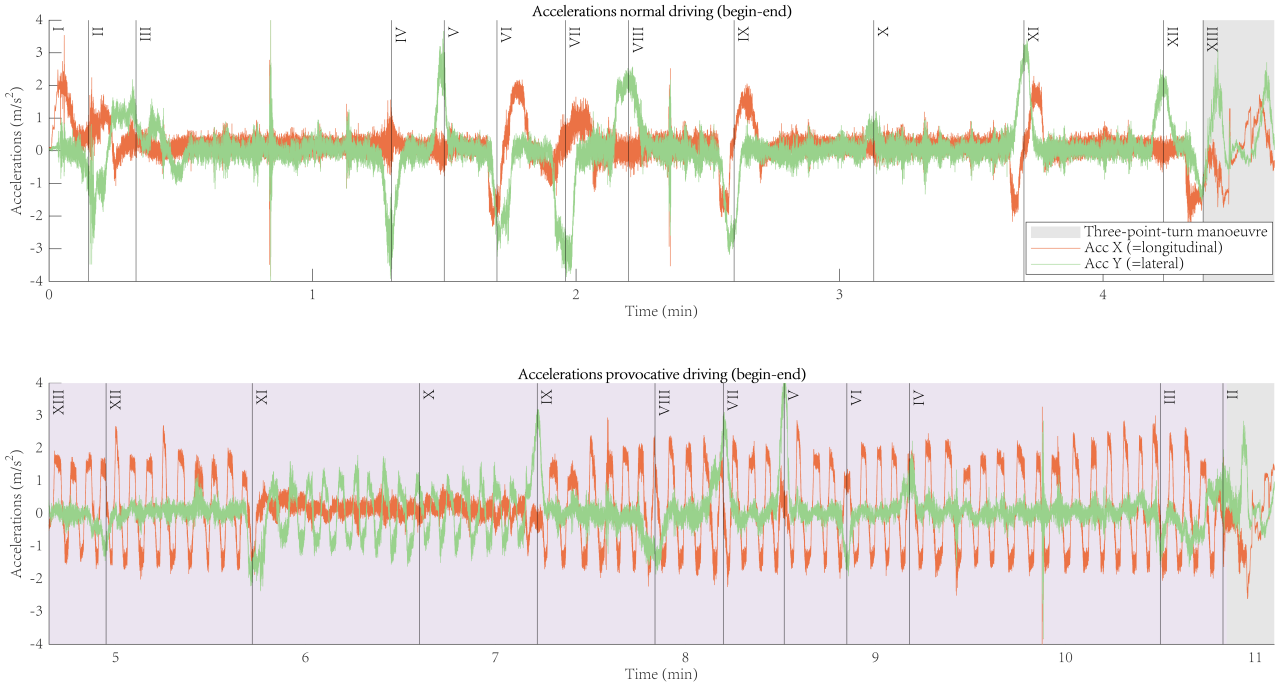


Figure 2.5: Raw longitudinal and lateral acceleration data (m/s$^2$) versus time (min) for normal driving (top), obtained at a speed of 30 km/h, and provocative driving (bottom), a result of driving speeds between 10-30 km/h. The combination of both sections is in total performed three times.

It is interesting to be able to observe a trend in increase and recovery in sickness over time for changing motion stimuli. Figure 2.5 (top and bottom) is combined into figure 3.2, which shows the sequence of normal and provocative driving three times, resulting in the complete experimental exposure. At the end of the experiment, 5 min of rest is implemented mainly for safety reasons but recovery from sickness may possibly be observed as well.

### 2.4.3 Task

Since we are interested in replicating a real-life scenario occurring for example in a SAE level 3 car or higher, participants in this experiment will be passengers. In both scenarios, the participants are motivated to believe that the (artificial) car is driving autonomously without the need for driver-related interaction, while they are able to engage in a secondary task. The (virtual) vehicle surroundings were displayed continually outside of the vehicle cabin and subjects were asked to minimize the amount of time spent looking outside. When every subject minimizes this aspect, not only are conditions kept constant, also the sickness elicited from ambiguities in the visualization modality (SIS) is minimized.

To reach a higher level of sickness [33] and to recreate a realistic secondary task that users of future automated vehicles could engage in, participants of this experiment are solving a Sudoku puzzle. To create motivation and similar focus between participants for this task, a prize is awarded to the quickest Sudoku solver. This task is chosen as it could always be solved, it needs little explanation and a ranking between participants could easily be made. Furthermore, Sudoku is considered a cognitive task [113]. Cognitive tasks require for example attention, memory and pattern recognition, all similar variables to working, discussing or reading in future autonomous cars.

### 2.4.4 Simulator design

As briefly explained in the introduction, a driving simulator is a tool often used to study and develop vehicle functionalities, to test human factors responses and to (re)create driving scenarios [114]. A schematic overview of a simulator is presented in figure 2.6. As is furthermore explained in the interpretations of literature, it is necessary to use a motion-base simulator with accompanying visual modality to study MS. The Cruden driving simulator that is used for this study, is operated by Cruden Panthera software that integrates vehicle dynamics, vision, audio, and motion platform control. To empirically compare MS to SS and validate the simulator regarding MS, the simulator requires to be designed to replicate the real world drive as close as possible regarding perceptual cues and accelerations. Some design choices regarding this perceptual fidelity are directly related to SS and will therefore be explained per modality or setting.

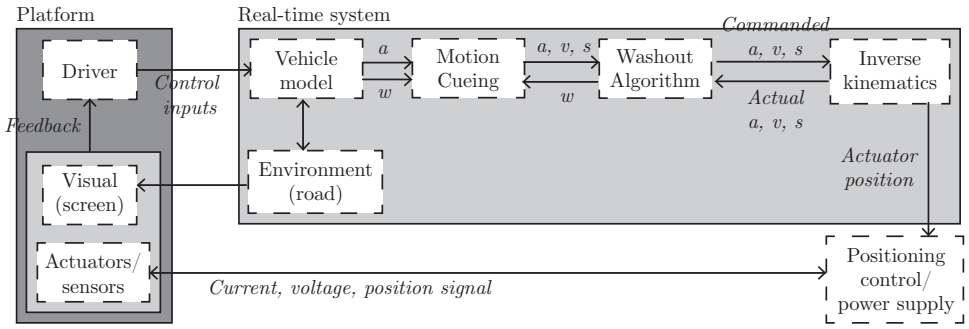


Figure 2.6: Schematic display of a driving simulator and its modalities

#### 2.4.4.1 Workspace and motion cueing

A Cruden hexapod motion-base will be used in this research, consisting of 6 actuators that are able to simulate motion in 6 Degrees of Freedom (DoFs). Hence, the workspace is always bound by the physical stroke of each actuator. As no additional yaw table or X and Y rails accompany the motion-base, the most salient limitation is its linear displacement. By means of a Motion Cueing Algorithm (MCA), the vehicle inertial motions (specific force and rotational velocity vectors) can be mapped onto the simulator motion space. This cueing is not only used to cope with the limited space available, it is also said to increase the realism of car accelerations [94, 115, 116]. Many different algorithms have been introduced [117], several of which use the *classical cueing model* as a baseline. This cueing algorithm is schematically displayed in figure 2.7. Since it is out of scope for this research to design a dedicated MCA and parameterization and for the aim of reproducibility of future research, the classical MCA is chosen. This algorithm is representative of what is currently used in many simulators [118] and very well documented [119].

For classical MCAs, inputs are real-time linear and angular accelerations. The first column of figure 2.7 shows three *scaling* blocks. Often 1:1 motion from on-road driving scenarios transferred to a driving simulator is perceived as too intense and drivers tend to underestimate simulator motion cues [92]. A scaling factor is then used. The driving simulator validity research by Bellem et al. (2017) [17], mentioned in the introduction, separated scaling factors for two scenarios (a slalom task and lane-change manoeuvre). They showed that the scaling of 0.5 for lateral and 0.6 for longitudinal, resulted in absolute validity for comfort measures when compared to on-road driving. Several other studies show a realistic perception of motion or even high validity, whereby scaling factors of 0.5-0.75 are used [17, 120, 121]. The Cruden classical MCA has default scaling gains of 0.7 in the longitudinal direction and 0.5 in the lateral direction, which will be used in this study.

After scaling of the inputs, signals are *filtered* as shown in the second column of figure 2.7. It is important to note here that gravitational acceleration is deviated from the translational input signals, as this allows to acquire the gravity vector components of the desired specific force. *Tilt coordination* is often added to classical MCAs, as it is said to improve

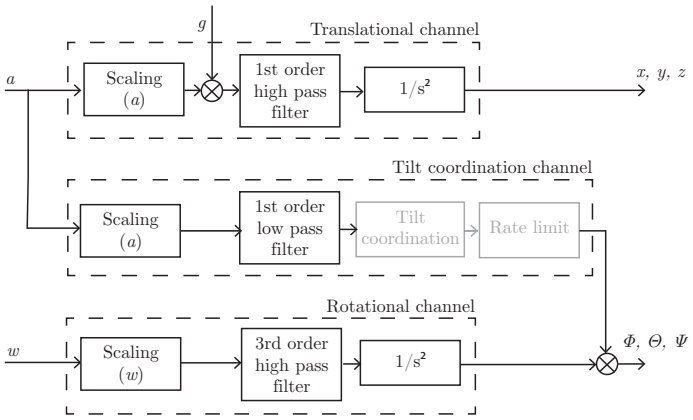


Figure 2.7: Baseline classic cueing model used, without tilt coordination and rate-limiting

motion cueing quality [122]. The tilt coordination tilts the simulator cabin to allow for longer sustained acceleration or curves. By tilting the platform (but not the visuals), it creates a roll or pitch angle whereby the gravitational acceleration is partly projected on the horizontal plane. This tilt is ideally slow enough for humans not to perceive and fast enough to keep the total acceleration as intended. After the tilt, the resultant roll motion is restricted by *rate-limiting*, to guarantee for the tilt to be below human perception thresholds [123]. The use of tilt coordination inherently creates a false cue and several studies have shown an effect of tilt coordination on sickness. Cleij et al. (2017) [75] observed that no limitations (regarding tilt coordination) lead to large rotational errors but that also limitation of the tilt and rotation rate coordination leads to false and missing cues and small rotational errors. Baumann et al. (2021) [115] stated that reduced tilt coordination (being well below perception thresholds) leads to a significant reduction of occurrence of MS. Generally, a large latency is introduced when using tilt coordination, due to the low pass filtering preceding the tilt block and that sensitive drivers or users notice being rotated when the real car would never rotate [119]. The ride in this study consists mainly of quick slaloming and stop-and-go manoeuvres, thus tilt would not contribute to the motion space necessary for this. Based on these findings, it is decided to leave tilt coordination (and accompanying rate-limiting) out of this study.

Motion cueing alone is not enough to prevent the platform from reaching the workspace boundaries (i.e. hitting the actuator endstops), which is why an additional *adaptive washout algorithm* is necessary. The motion base used in this study is therefore equipped with the Direct Workspace Management algorithm [124]. This algorithm uses a low frequency, low amplitude acceleration signal opposing a motion displacement to return to the centre of the workspace while remaining below the human perception threshold. The algorithm creates an acceleration in opposite direction from the intended direction, thus a false cue. Minimizing this false cue would require a more aggressive washout, which is unwanted. Instead, the option arises to push the platform forward in the workspace as a form of *pre-positioning*, especially helpful for repetitive braking and accelerating such as in this research. This feature pushes harder as driving speed increases, thus minimizing the false cue whilst creating more workspace during surge motion.

#### 2.4.4.2 Visual environment

As opposed to the study by Kuiper et al. (2019) [24], in this study visual cues are present. Even though participants will mainly be looking at the Sudoku throughout during the rides in both scenarios, visual content is evidently present in the peripheral vision of participants in the real world. Since the addition of peripheral vision cues is said to influence sickness [125, 126], these are ideally also present in the simulated world to replicate the reality as close as possible.

The Valkenburg route is one without accurate point cloud data available, such as LiDAR data. Acceleration, gyroscope and GPS data are recorded with an accelerometer and used to plot the trajectory. By means of (inaccurate) available

Figure 2.8: Snapshots of the visual content presented in scenario A (left) and artificially created for scenario B (right)

OpenStreetMap geodata, an overlay for the plotted trajectory is created. An elevation map, which is another input for creating a realistic scenario, is obtained from Shuttle Radar Topography Mission, a NASA world elevation application. These three sources are then combined to create the detailed trajectory that closely resembles the route in the actual environment.

Using MathWorks VectorZero RoadRunner 2019.0.5 software, a 3D environment suitable for driving is created. The created scene exists of the driveable road sections and junctions; simplified nearby buildings, fences, trees and bushes and surrounding terrain to closely replicate the real scenario. The objects in the scene are accompanied by textures and materials. Driving layers and object files exported from RoadRunner can then be imported into the Unity 2019.4 software. Unity is another development platform for creating 2D and 3D interactive experiences, mainly used for gaming industries. The results of aforementioned steps are shown in figure 2.8. By means of Panthera-compatible packages and settings, the Unity scene can be imported into the dedicated Panthera 2021B software.

#### 2.4.4.3 Experiment recording

In this recreated environment, a similar drive as in the real world is performed with a 6 DoF vehicle model. Logged data thereof is collected at a frequency of 1000 Hz and integrated into one single recording afterwards. The parameters (e.g. regenerative braking force) of the model are tuned to ensure that participants will be exposed to similar acceleration stimuli as in the real world. The recording is converted into a Matlab (2019B) Simulink model, compatible with the Panthera software of the simulator. Due to the limited motion envelope of the simulator and parameterization of the MCA, inevitably ambiguities will arise in the acceleration output of the simulator as compared to the real road drive. The results section shows where the recorded drive deviates from the real world drive.

# 3

# Methods

In this chapter, the methods designed to answer the research question, will be explained. As the experimental design has already been addressed in the previous section, it will be left out of this chapter. Firstly, participant demographics will be discussed. Used apparatus are explained next. Lastly, both subjective and objective measures and their data processing methods will be explained.

## 3.1 Participants

In total, 25 young, healthy participants took place in the experiment (mean = 25.3 years, std = 3.4 years). Of the 25 participants, 7 were female and 18 were male. Both students and non-students from the Delft University of Technology took part. All participants are in possession of a drivers license, have normal or corrected to normal vision, reported no vestibular disorders and have either no simulator driving experience at all (72%) or have never driven the simulator setup used in this experiment. All participants were asked to refrain from recreational drug consumption including alcohol and caffeine for 24 hours prior to the experiment. The main researcher who drove all participants was not a participant herself.

The MSSQ-Short form that is used as an online version in this study, can be found in [127]. The mean MSSQ of 13.1 (std = 11.4), indicates a percentile conversion of 57. The participants in this experiment have a higher sickness susceptibility than 57% of the population. The MSSQ was not used for participant selection .

### 3.1.1 Ethics statement

All participants provided written informed consent prior to participation and after the task had been explained. The experimental protocol was approved by the Human Research Ethics Committee of the Delft University of Technology, The Netherlands, under application number 1765 and data were collected and stored anonymously. Participants were free to terminate participation in the experiment at any time without any consequences. They received 5 €/hr for their participation and travel time in the form of vouchers. All experimental procedures were conducted in accordance with the ethical guidelines of the Declaration of Helsinki [128] and conform to the current COVID regulations in force as defined by the Dutch National Institute for Public Health and the Environment (RIVM).

## 3.2 Apparatus

### 3.2.1 Car

A Hyundai Kona Electric (2019), equipped with a.o. automatic transmission, three levels of regenerative braking, Virtual Engine Sound System and cruise control and instrumented with a 6 DoF Xsens MTi-G-700 accelerometer, has

been used for the experiment. The inside temperature is kept constant at 18 °C. The experimental setup of the car scenario with the Hyundai on the route can be seen in figure 3.1 (left). Participants are seated in the passenger seat. The researcher drives manually and the frequency of slaloms and stop-and-go manoeuvres was maintained with the help of a metronome. Accelerations, rotations and GPS fix position are recorded with the accelerometer located at the passenger seat H-point. Recording with this device is done at a frequency of 400 Hz during the experiments.

Figure 3.1: The Hyundai Kona on the Valkenburg route (left) and the Cruden motion-base driving simulator (right), both in its starting position (I) of the route

### 3.2.2 Simulator

In figure 3.1 (right), also the setup of the simulator scenario is shown. This Cruden simulator used, consists of the AS2 (eM6-640-1500) 6 DoF hexapod motion base, of which the workspace specifications are displayed in table 3.1. The moving base cutoff frequencies are 12 rad/s (1.9 Hz) and 9 rad/s (1.4 Hz), for longitudinal and lateral motion respectively. The cutoff frequencies for both DoFs are thus far above the peak in sickening frequency. On top of the motion base platform, a vehicular cab is mounted.

| **Axis** | **Position** | **Velocity** | **Acceleration** |
|---|---|---|---|
| Surge | -0.48 m to 0.60 m | 0.8 m/s | 13 m/s$^2$ |
| Sway | ±0.50 m | 0.8 m/s | 13 m/s$^2$ |
| Heave | ± 0.41 m | 0.6 m/s | 14 m/s$^2$ |
| Roll | ± 23.8° | 35°/s | 500°/s$^2$ |
| Pitch | -23.7° to 26.0° | 35°/s | 500°/s$^2$ |
| Yaw | ± 25.4° | 40°/s | 800°/s$^2$ |

Table 3.1: Specifications of the Cruden 640 hexapod motion system performance with accelerations based on the standard generic dynamic frame payload (800 kg)

The visualization modality consists of 3 Norxe AS 101-1007 P1+ WQXGA IR projectors with a resolution of 2560 × 1600 and a refresh rate of 120 Hz. Their brightness is 3000 lumens. The projectors create a surrounding vision with a radius of 3 m and FoV of 200° horizontally and 50° vertically. The total projection height is 2.2 m. Additionally, three LCD displays were mounted on the place of the side and rear-view mirrors, in order to simulate mirror images. Auditory road and wind noise are provided by three surrounding speakers attached, generated by the Panthera software. Although subjective, this simulator fidelity can be classified according to studies by Kaptein et al. (2010) and Wynne et al. (2019) [19, 98]. After adding all of the above-mentioned fidelity aspects, this simulator receives a *medium to high*

*fidelity* classification. The driving simulator is CE-certified.

Participants in the simulator are seated in the driver's seat, as no passenger seat is present. However, as a result of the offset in position as compared to the passenger seat in the car condition, perceived accelerations could be different. This difference is compensated for, by adding an offset to the lateral distance from the sensor in the virtual world (Center of Gravity of the vehicular cab), to the H-point of the subject, within the MCA.

## 3.3 Procedure

Prior to participation, each subject filled out an (online) MSSQ-Short form. This is done several days beforehand to prevent the interference of possible higher sickness levels as a result of expectancy [129]. A briefing about the Sudoku, the rating of the MISC and the experiment phases is given to each participant at the start of each scenario. The experiment phases for both scenarios are similar and illustrated in figure 3.2. A spare Sudoku puzzle of the same difficulty level was attached to the original Sudoku, in case a participant was very quick solving the first. Attachment of the electrodes measuring physiological responses, as will be explained in the following section, is done next. The participant is driven from begin-end (white areas) with normal driving behaviour and after a three-point-turn (grey areas), driven from end-begin (purple areas), with provocative driving manoeuvres. This subsequent combination of manoeuvres will be performed three times, dividing the total immersion into 7 time epochs. Participants are instructed to minimize looking out of the window and focus on their puzzle. After the last excitation manoeuvre or whenever a MISC of 7 is reached (or when termination is requested otherwise), a resting period (dark grey) of 5 min is started. Throughout this motion exposure and rest period, participants had to write down their MISC scores every 30 sec. After the rest period, the participant is asked to fill in the MSAQ. Both conditions are tested on different days with several days in between (mean = 13.2 days, std = 6.7 days) to prevent the influence of expectancy, habituation and possible interference by Mal Debarquement [130].

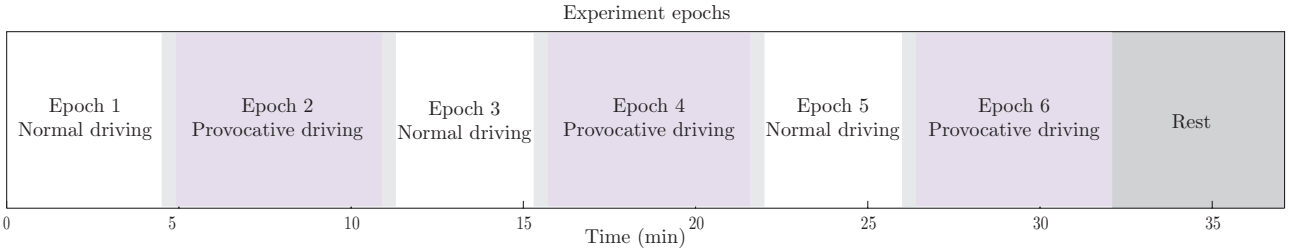


Figure 3.2: The complete experimental immersion procedure, existing of normal driving (white), provocative driving (purple), three-point-turn manoeuvres (light grey) and resting (dark grey), versus time (min)

## 3.4 Measures of sickness

In order to answer the research question, the severity of sickness must be quantified. In this study, MISC and MSAQ are recorded as subjective sickness measures. For subjective questionnaires, it has to be noted that the understanding of symptoms is subjective, that it is never a matter of fact that every human body adheres to a certain sequence of symptoms and especially for continuous rating training might be necessary in order to obtain consequent results. In order to pose a consistent analysis of the differences and similarities between SS and MS, also objective measures are evaluated. Galvanic Skin Response (GSR) and Electrogastrography (EGG) are validated indicators of MS.

### 3.4.1 MISC

The MIsery Scale (MISC) [58] features a one-dimensional ordinal rating scale from 0 (no problems) to 10 (vomiting), used in several researches [67, 131] to indicate sickness levels of participants. The MISC table used in this study can be found in [132] and in appendix B. The MISC provides a measure of subjective sickness severity by means of

symptoms that are similar between participants [110]. This quantification method is continuous and therefore safer for participants as sickness can be monitored real-time. MISC addresses the progression of symptoms and presents a time course or trend of sickness over time for individual subjects. A disadvantage is however that continuous rating might interfere with the experimental task and especially at lower sickness levels, there is the uncertainty of people to rate how they are feeling.

#### 3.4.1.1 Recording

Participants are familiarized with the MISC rating during the briefing preceding both scenarios. By means of a metronome also used to obtain stop-and-go and slaloming behaviour at 0.2 Hz, every 30 sec the researcher receives an auditory signal and asked aloud *MISC?*, after which the participant is required to write down a single integer representing their level of sickness. After every three-point-turn manoeuvre succeeding a time epoch, a new row with MISC scores is started. This way, onset of individual epochs between MISC scores for the car and the simulator can be matched. Also, GSR and EGG signals can be synchronized with specific time epochs as a result of this MISC rating system. The scoring table in combination with the Sudoku, is presented in appendix B. The experiment is terminated (and the rest period started) when MISC = 7, is reached. This level indicates that a participant is "rather nauseated", thus sufficient data could be obtained. It is furthermore is deemed an appropriate and ethical threshold for the experiment.

#### 3.4.1.2 Data processing

As shortly described in the hypotheses, four metrics related to the MISC scores will be used to validate a simulator for MS research. The first metric is the individually (averaged) MISC, one single MISC value for each participant throughout the whole exposure time and a single MISC value for the final time instance of the experiment, preceding the rest epoch. The latter value is calculated by extrapolating with a MISC of 7 if a participant requested termination of the experiment prematurely. This method has been used by numerous other studies [7, 110, 133] and accounts for the difference in experiment times between subjects. The second metric represents the time to all (7/10) sickness levels. The time to onset of these levels is given maximum value when a certain level is never reached. The third metric is the mean MISC group response. The mean MISC refers to the average of MISC scores of all participants for every time step during the full motion excitation in both scenarios. If a given participant wished to terminate the exposure prematurely (as a result of severe nausea), again a value of 7 is used in the averaging for the remainder of the experiment duration. The rest period, succeeding every experiment, is left out of this calculation as the onset time of this epoch varies between participants. Lastly, another important metric to compare trends of SS and MS is the MISC rate. The MISC rate provides an indication of how quickly sickness has increased or decreased in a time epoch. The rate is calculated by subtracting two time points, being the points I and VIII for each excitation manoeuvre and divided by its time interval.

### 3.4.2 MSAQ

The Motion Sickness Assessment Questionnaire (MSAQ) is a multidimensional symptom list, first introduced by Kennedy et al. (1993) [56] and later modified by Gianaros et al. (2001) [57] to the questionnaire used in this study. This second well-known rating method correlates well to several other MS questionnaires and is used in many types of research. The MSAQ divides 16 symptoms of MS into 4 categories, by Likert rating them with a number between 1 (not at all) to 9 (severe). The categories are Gastrointestinal (nausea, vomiting), Central (lightheaded, dizzy, spinning), Peripheral (sweaty, feeling warm) and Sopite-related (irritated, drowsy, tired). Even though the questionnaire provides only a single timestamp data point, it provides information about symptoms obtained as a result of the exposure. MSAQ subscale results can furthermore create a symptom profile that could be compared between both scenarios.

#### 3.4.2.1 Recording

Directly after the experimental excitation of both scenarios, the participant is asked to rate symptoms of the MSAQ of the completed immersion, on paper. The MSAQ is presented in appendix C

#### 3.4.2.2 Data processing

Several studies evaluate the overall and subscale MSAQ scores [57, 67, 68, 134]. According to these studies, the total and categorical subscores can respectively be calculated using the following equations:

$$\text{Overall } MSAQ \text{ score } = \frac{\text{total points from all items}}{144} \cdot 100 \tag{3.1}$$

$$\text{Subscale } MSAQ \text{ score } = \frac{\text{total points from all subscale items}}{\text{number of questions related to subscale} \cdot 9} \cdot 100 \tag{3.2}$$

### 3.4.3 EGG

Of all the symptoms that accompany MS, nausea is the most often and usually first symptom to be reported [135]. As nausea directly relates to gastrointestinal distress, MS could be detected by means of an electrogastrography. Electrogastrography (EGG) is a method to measure electrical stomach and nerve activity, which consists of slow waves and spike potentials. A slow wave is a sinusoidal signal with a Dominant Frequency (DF) between 2-4 cpm (0.05 Hz) and a typical amplitude of 50–500 μV in healthy subjects [136]. A DF below and above 3 cpm, are respectively called bradygastria and tachygastria. It is expected for the DF of the slow-wave and its power to increase as a result of motion sickness [69, 136–139], suggesting an increase in power in the tachygastric area (4-9 cpm) while experiencing MS. Whether bradychastria also decreases is less clear [69, 140]. It has to be noted that there exists a large intersubject variability in EGG data, as the DF wave is affected by for example age and gender [141], by anthropometrics such as height and weight and by individual differences in susceptibility for MS. EGG is noninvasive and uses surface electrodes positioned on the abdominal surface, which provides a reason why it is sensitive to motion artefacts and electrical interference from other internal organs.

#### 3.4.3.1 Device and placement

The EGG signal is recorded with a proprietary 4-channel recording NeXus 4 amplification device and Ag/AgCl cutaneous electrodes from Ambu Blue Sensor, sampling the output at a rate of 1024 Hz. Various possibilities to place the electrodes on the abdominal surface exist [136, 138, 142, 143]. Several of these studies use dedicated EGG sensors accompanied by a multichannel setup with for example 10 electrodes [144]. However, the NeXus 4 allows up to two sets of bipolar leads (initially intended for EMG/ECG/EEG), resulting in the following setup for measuring gastric activity as can be seen in figure 3.3 (left). Before placing the electrodes, the skin has to be cleaned with alcohol to ensure good connectivity with the electrodes.

#### 3.4.3.2 Data (pre-)processing

EGG data in the time domain is not easy to interpret and explains little about the sickness of an individual during a particular time epoch or scenario. Instead, the frequency domain could be analyzed. The extensive review of Yin & Chen (2013) [136] suggests as metrics for 2-channel devices specifically the extraction of DF, of relative power change, and the percentages of DF-, tachygastric- and bradygastric power distribution. In a similar fashion as is done in [69, 136, 144, 145], the percentage of band power is in this study calculated as follows:

$$\%\text{ tachygastric band power} = \frac{\text{power in 4-9 cpm band}}{\text{total power in 1-10 cpm band}} \cdot 100 \tag{3.3}$$

$$\%\text{ bradygastric band power} = \frac{\text{power in 0.5-2 cpm band}}{\text{total power in 1-10 cpm band}} \cdot 100 \tag{3.4}$$

Data is first downsampled from initial recording with 1024 Hz, to 32 Hz for calculation purposes. As interest lies within the frequency range of slow waves (1-10 cpm), the signal is bandpass filtered with a Butterworth filter of 1st order and a frequency of 0.01-0.1 Hz, in line with similar studies [62, 136, 137]. Filtering in this range minimizes interference from other physiological processes such as respiration, heart rate and removes motion frequencies or frequencies as a result of contracting muscles as a possible counteraction of manoeuvring at 0.2 Hz.

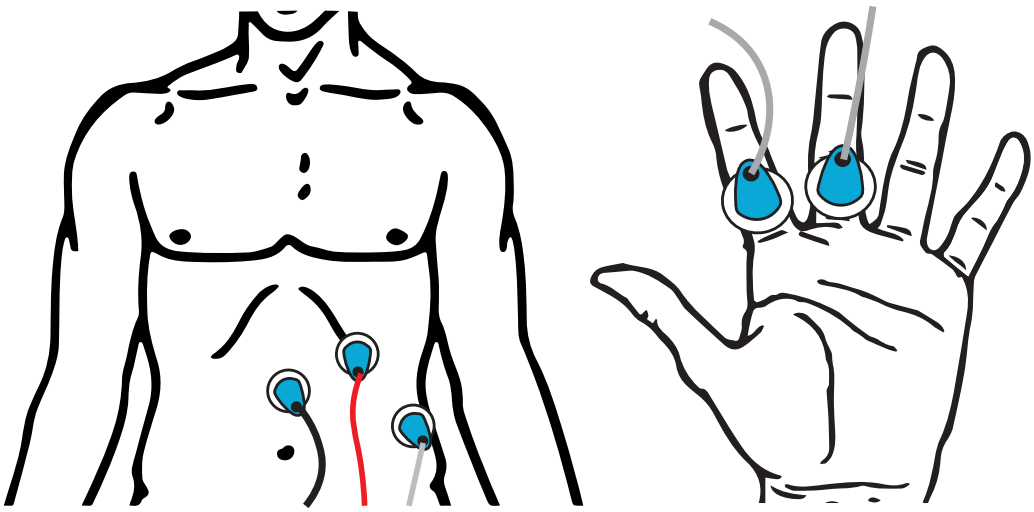

Figure 3.3: Left: Placement of electrodes for EGG. Right: Placement of electrodes for GSR.

### 3.4.4 GSR

Another of the many symptoms accompanying MS is that of (cold) sweating [27]. Electrodermal Activity (EDA) and Galvanic Skin Response (GSR) are both terms for the same property of the human body that causes continuous variation in the electrical characteristics of the skin. Two types of sweat glands exist, eccrine and apocrine glands. The eccrine sweat glands, which are mainly responsible for changes in skin response, can be found everywhere on the body. Their density is, however, highest on the hand palm and the foot. Therefore, most studies measure GSR from the hand. Raw GSR data consists of a tonic, low-frequency Skin Conductance Level below 0.16 Hz and a phasic, high-frequency signal referred to as Skin Conductance Response [146]. Both tonic and phasic GSR increase as a result of MS and comparable studies provide the combination of both components [62, 69, 137, 147, 148]. Typical values for GSR vary between 2-20 μS and the onset of changes in GSR is often delayed by a few seconds. GSR is easy to use and noninvasive and is furthermore been employed in numerous studies regarding MS and SS [64, 110, 149]. Sweating is an autonomic response of the body to various internal and external inputs, such as stress. Therefore, increases in GSR might as well be a response to another cause.

#### 3.4.4.1 Device and placement

The GSR signal is recorded with the same proprietary 4-channel recording NeXus 4 amplification device and Ag/AgCl cutaneous electrodes from Ambu Blue Sensor, sampling the dedicated skin response output at a rate of 32 Hz. Various possibilities to place the electrodes on the abdominal surface exist, such as hand, wrist and foot [149]. However, as the hand has shown promising results in similar studies, the following setup for measuring GSR can be found in figure 3.3. The electrodes are placed on the inside of the index and middle finger of the non-dominant hand.

#### 3.4.4.2 Data (pre-)processing

Before analyzing, it is important to remove the most common types of noise or artefacts, being high-frequency noise and rapid-transient artefacts [150]. Filtering can be done in many ways [146], in this study a 3rd order lowpass Butterworth filter with a cutoff frequency of 2.1 Hz [151, 152], is used to capture both components of the GSR signal [153].

GSR in μS is averaged over each experimental epoch for each subject. These epoch values are then normalized by the baseline resting epoch to account for individual differences in resting skin conductance and for differences in environmental conditions such as humidity and temperature. In line with several other studies [69, 147, 154], this is

done accordingly:

$$\Delta\, GSR = \frac{\text{average per epoch - average baseline}}{\text{average baseline}} \tag{3.5}$$

## 3.5 Data analysis

Absolute validity, when research outcomes show no differences between both scenarios, is in statistical terms supported by the absence of significant main and interaction effects. Relative validity is given when significant differences of the dependent variable in both conditions are of the same order, direction or magnitude. The factor order separates the cohort of subjects into two groups. If a main effect of order or an interaction effect of order on the comparison between the car and the simulator exists, no validity is assumed. Correlation analysis will support the outcomes of statistical tests by providing the strength and direction between MS and SS measures. Metrics that are normally distributed, will be evaluated by means of repeated measures Analysis of Variance (ANOVA), using order (first car or first simulator), driving style (normal or provocative) and scenario (car or simulator) as factors. Time and participant were included as constant and random effect, respectively. For metrics of which residuals violate the Gaussian assumption, the Aligned Rank Transform (ART) [155, 156] ANOVA will be used to evaluate validity. In contrast to classic nonparametric statistical tests (e.g. Friedman) or rank transformations (e.g. Rank Transform), ART ANOVA allows for multiple factors and their interactions to be analyzed. The procedure is as follows: For each raw dependent variable, residuals are computed. Then, estimated effects are computed for main and interaction effects. Afterwards, the aligned response is calculated by adding the results of the previous steps. In statistics, data alignment is often used to estimate effects as marginal means which are then stripped from the response variable, so that all effects but one is removed [155]. The fourth step is to assign averaged ranks to columns of aligned observations. Lastly, a Linear Mixed Effect Model including all main and interaction effects can be used to perform a full-factorial ANOVA [157]. Concluding results from either the ART ANOVA or regular ANOVA are in essence similar. Before the analysis is performed, it is verified by means of Residual Diagnostics (histogram, Quantile-quantile plot and Kernel density estimate), whether data are approximately normally distributed and the Gaussian innovation assumption holds. Cross-correlation ($R_{XY}$) or coherence ($C_{xy}$) is used when comparing patterns of sickness of both scenarios over time. Spearman's $\rho$ correlation test is used for the data with non-Gaussian distribution or when dependent variables are ordinal, Pearson correlation is used when data is normally distributed. Alpha significance level is set to $p < 0.05$. Whence more than one statistical test per dependent variable is used, the chance of committing a Type I error increases, thus increasing the likelihood of coming about a significant result by chance. Bonferroni Type Adjustment is then applied in addition to the original $p < 0.05$, to adjust for the number of comparisons. Averaged values are given as *mean* ± *standard deviation* (std), unless stated otherwise. Analysis and visualization of the data is done with MATLAB R2020b software.

# 4

# Results

In this chapter, the results of the experiment will be presented. First, an analysis of the perceived motion stimuli in either scenario is presented. Secondly, subjective MISC and MSAQ results are discussed. Lastly, objective GSR and EGG results are presented. For each measure, the (type of) validity is specified.

## 4.1 Motion stimuli

### 4.1.1 Vestibular motion

As a consequence of approximately 33 min of manual driving by one single researcher in the car scenario, accelerations and thus perceived sickening stimuli might vary between subjects. In the simulator scenario, no deviation in perceived stimuli between participants exists, as a result of replaying one pre-recorded drive. Moreover, as a result of motion cueing for the limited simulator workspace, accelerations may differ from those obtained in the car. Hence, this section will analyze the degree of similarity between accelerations obtained in the simulator and the car accelerations, and whether participants in the car perceived comparable acceleration stimuli.

| | **Car** | **Simulator** (vehicle model) | **Simulator** (platform output) |
|---|---|---|---|
| Complete drive | Longitudinal: 0.88 ± 0.06<br>Lateral: 0.79 ± 0.04 | Longitudinal: 0.91<br>Lateral: 0.81 | Longitudinal: 0.19<br>Lateral: 0.08 |
| *Normal driving* | Longitudinal: 0.49 ± 0.05<br>Lateral: 0.84 ± 0.06 | Longitudinal: 0.42<br>Lateral: 0.97 | Longitudinal: 0.06<br>Lateral: 0.08 |
| *Provocative driving* stop-and-go | Longitudinal: 1.16 ± 0.09<br>Lateral: 0.66 ± 0.04 | Longitudinal: 1.29<br>Lateral: 0.54 | Longitudinal: 0.29<br>Lateral: 0.05 |
| *Provocative driving* slalom | Longitudinal: 0.29 ± 0.10<br>Lateral: 0.88 ± 0.07 | Longitudinal: 0.37<br>Lateral: 0.89 | Longitudinal: 0.04<br>Lateral: 0.15 |

Table 4.1: Car and simulator (vehicle model and platform output) accelerations in ($m/s^2$), versus sections in the experiment. Each value for the car represents a RMS (± std) acceleration participant average.

Firstly, in order to evaluate deviations between the recorded simulator drive and the car drive, the Root Mean Square (RMS) is calculated for every (sickening) manoeuvre and shown in table 4.1. The small std values for the car scenario show that differences in perceived accelerations between participants are small. Although intended, the results show mainly longitudinal motion for stop-and-go behaviour behaviour. Similarly, slaloming shows primarily lateral motion. Heave RMS (± std) values for the car result in 9.73 ± 0.06 m/s$^2$ and 0.16 m/s$^2$ for the simulator, of which the output mean is 0. Heave accelerations are in the remainder of these results left out of for legibility purposes.

Secondly, a comparison between recorded accelerations in the car for every participant and pre-determined output accelerations of the simulator, the *vehicle model output*, will be made. This comparison is visualized in figure 4.1, by means of the blue line (car) and the thin red line (simulator) and by means of the second and third column of table 4.1. First, data with the lowest sampling rate (car, 400 Hz) is resampled so that cross-correlation ($R_{XY}$) can be performed with data from the simulator. Cross-correlations provide for each subject the location of the maximum value, thus time leads or lags between scenario acceleration data. Signals are afterwards lined up by clipping the vectors with longer delays. Results show varying correlation coefficients between participants, with a minimum of r = 0.35 and a maximum of r = 0.86 (mean = 0.51) for longitudinal motion. For lateral motion these correlations vary between r = 0.28 and r = 0.92 with a mean of 0.42. The variations can be attributed to the fact that these correlations do not account for shifts and speed differences within the signals, for example when omitting one slalom turn or taking several seconds longer to perform the three-point-turn manoeuvre while driving one participant.

Thus, experienced car stimuli provide strong correlations with the pre-determined simulator accelerations. However, as explained in chapter 2, these accelerations still require to be subjected to the scaling and washout in order to fit onto the limited simulator workspace. In a similar fashion as described above, experienced car stimuli are now compared to the perceived simulator *platform output* accelerations. This comparison is visualized in figure 4.1 as the comparison between blue and the bold red lines and the comparison between the second and fourth column in table 4.1. Results show again varying correlations between participants, with a minimum of r = 0.18 and a maximum of r = 0.56 (mean = 0.27) for longitudinal motion. Correlations for lateral motion vary between r = 0.14 and r = 0.75 with a mean of r = 0.18. These now weak to moderate correlations between perceived accelerations in both scenarios are a direct result of the simulator MCA and washout.

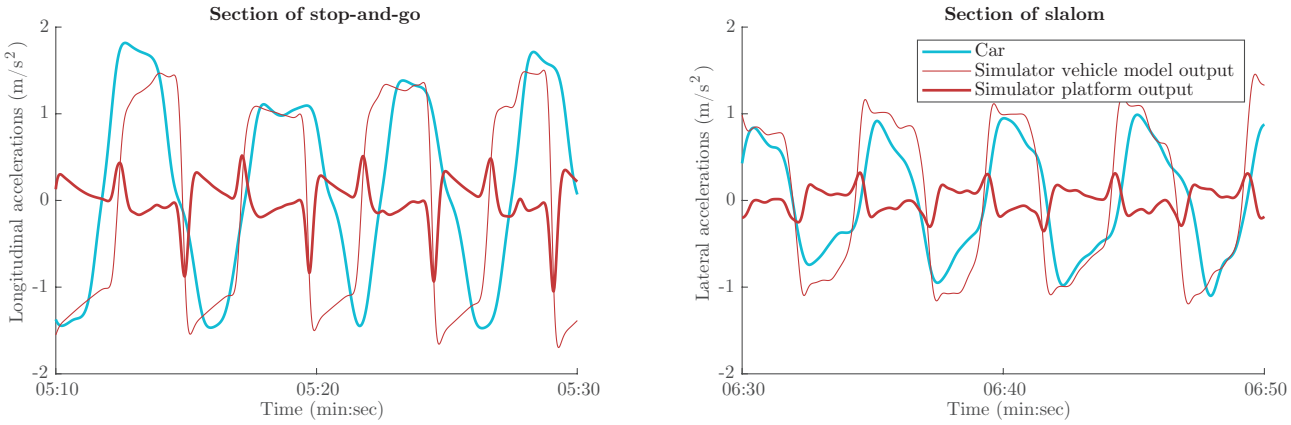


Figure 4.1: Left: Longitudinal accelerations (m/s$^2$) representing a section of the stop-and-go behaviour, versus experiment time (min:sec). Right: Lateral accelerations (m/s$^2$), representing a section of slalom manoeuvres, versus experiment time (min:sec). In both figures, the perceived car (blue) and the perceived simulator (red) motion is shown in bold. The vehicle model output of the simulator is added by means of a thin red line. The shape thereof is a consequence of model particularities. The difference between the thin and bold red lines denote the scaling and washout of the MCA.

Since the peak for sickness in the horizontal plane is mostly situated around frequencies of 0.2 Hz [38, 42], power spectrum analysis could give further insight if stimuli experienced in both conditions are theoretically able to cause similar sickness. Figure 4.2 shows normalized power for longitudinal (bold) and lateral (dotted) accelerations for the car (blue) and the simulator platform output (red). The figure demonstrates a peak at 0.2 Hz for both DoFs and conditions. It can also be seen that the simulator (MCA) attenuates low-frequency accelerations [17], whereas in the car these frequencies are preserved. While correlation is effectively used to determine the extent to which time-domain signals covary, spectral coherence can be a robust measure to assess the similarity between signals in the frequency domain [158]. The magnitude squared coherence estimate ($C_{xy}$) confirms that the experienced accelerations in the car and the simulator have a main correlated component around 0.2 Hz for every participant. The subject average coherence between the car and the simulator is $C_{xy}$ = 0.89 for longitudinal accelerations and $C_{xy}$ = 0.87 for lateral accelerations.

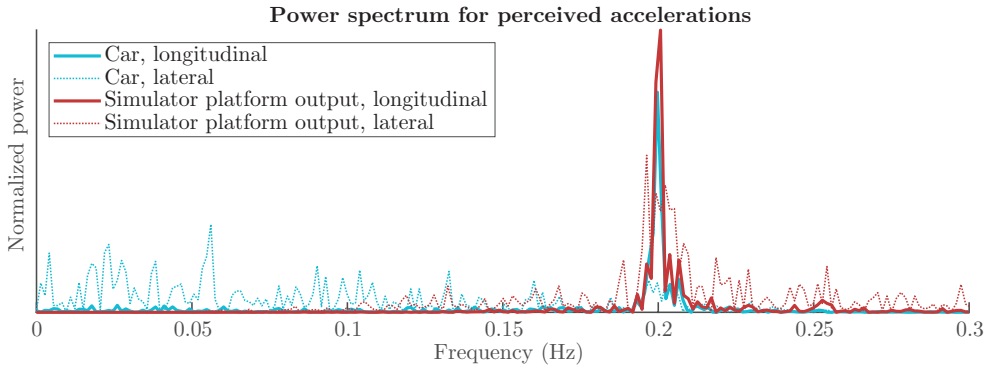


Figure 4.2: Normalized power spectrum of longitudinal (bold) and lateral (dotted) accelerations in frequency (Hz) for the complete drive, shown for the car (blue) and the simulator (red).

### 4.1.2 Visual motion

Participants were instructed to minimize their viewing outside the windows and focus on the Sudoku placed on their lap. Therefore, the only visual cues that could be perceived, were peripheral cues. It is assumed that as no emphasis is placed on the specific (ambiguities of the) visualizations, sickness obtained as a result of conflicts in the (peripheral) visual modality [125, 126], is similar for both the car and the simulator. It is thence also assumed that vestibular motion cues, as compared to visual motion, are prominent in the development of sickness in this study.

## 4.2 MISC

The raw MISC responses are plotted per participant for both scenarios in order to provide individual comparison between MS and SS in figure 4.3. The results for participant NL are left out of this data as only the simulator scenario has been completed, thus no comparison is possible. The rest epoch for participant TI is not recorded. Since 5 statistical tests are conducted on the MISC dependent variable, the Bonferroni type adjustment suggests statistical significance for $p < 0.01$. Table 4.2 at the end of this section shows all concluding results regarding the car and the simulator scenario for all evaluated MISC metrics.

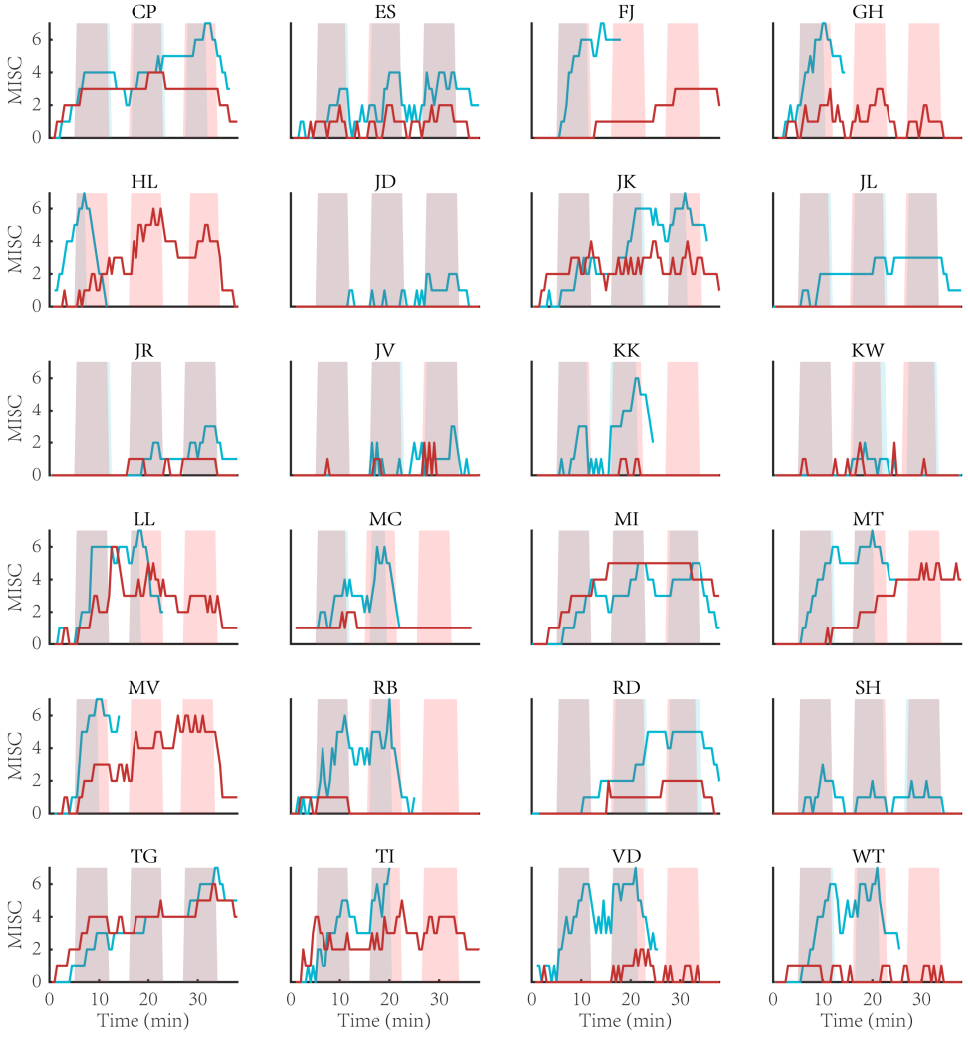


Figure 4.3: Individual sickness responses for the car (blue) and simulator (red) of 24 participants with the MISC score (0-7) versus experiment time. Red areas denote provocative driving sections in the simulator scenario, blue areas denote provocative driving in the car scenario. When an area is (partly) double coloured, it means that the participant experienced the specific part of the route in both conditions. From the overlay of provocative driving sections (coloured areas) of both scenarios, it can be seen that given MISC scores for a single participant only difference between the car and the simulator with a maximum of 1 min. Every exposure is ended with 5 min of rest.

From these individual figures, a few things can be noted. Firstly, the highest MISC score was recorded in the car. Secondly, what can be observed is the large difference in susceptibility to the stimuli between participants. Some partic-

ipants reached high levels of sickness within minutes, whereas others endured the complete drive without obtaining any symptoms. Also, the sensitivity to the provocative epochs in the route is diversified. Whereas some participants increase sickness levels during provocative driving (coloured areas) and decrease during normal driving (white areas), for some participants the type of driving style does not appear to influence their rating. Inevitably, a percentage of participants show no effect to the provocative stimulus [6, 24, 159]. The phenomenon as addressed by these studies can also be seen in this research, as half of the subjects did not report any severe sickness symptoms during the entire excitation. Lastly, what can be observed from the individual sickness trends is that for all participants except KW and MI, the peak level of sickness intending the highest MISC score reached throughout exposure, is obtained in the car.

### 4.2.1 Individually averaged MISC

Figure 4.4 (left) shows the *individually averaged MISC*, calculated as a single mean MISC value for every participant. Scores of participants are higher in the car, as compared to the simulator if scatters are located under the red MS = SS reference line. The figure shows two marginal outliers, participants MI and TG, whose average MISC score was higher in the simulator than in the car. Raw MISC scores (figure 4.3) for participant MI shows similar onset of sickness in both scenarios, but a plateau in SS and decrease in MS in normal driving sections. For TG, a very similar sickness over time for the car and the simulator can be seen, resulting in his scatter point falling just outside the red line. The average MISC throughout the immersion time was 3.43 ± 1.94 for the car and 1.33 ± 1.26 for the simulator. Two-way ANOVA (scenario × order) for individually averaged MISC scores, results in no significant main (p = 0.817) and interaction effects (p = 0.210) of order. Averaged MISC values were found to be significantly higher in the car as compared to the simulator [f(1,44) = 30.23, p = 0.000]. Also, a strong significant correlation of r = 0.57 (p = 0.004) between the car and the simulator exists.

MISC values at the last timestep of the experiment (t = 33 min) were 5.42 ± 2.32 for the car and 1.88 ± 1.85 for the simulator. Another two-way ANOVA for the individual MISC scores at the end of the experiment (t = 33 min), reveals again no significant main (p = 0.751) and interaction effects (p = 0.219) of order. Again, MISC responses are significantly higher in the car as compared to the simulator scenario [F(1,44) = 54.26, p = 0.000]. Accompanied by figure 4.4 (right), is a strong positive correlation of r = 0.52 (p = 0.009). Results from both these two metrics show relative validity between the car and the simulator and above all a strong relationship between both.

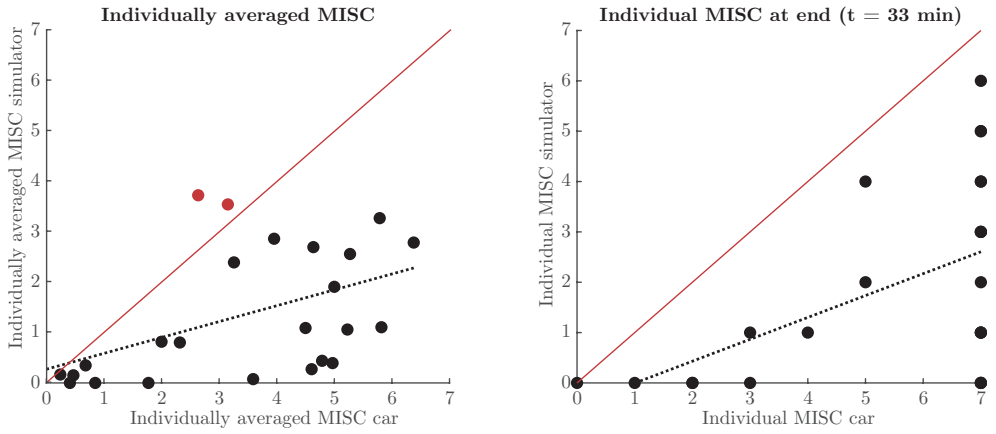


Figure 4.4: Left: The individually averaged MISC over the total experiment time (without rest epoch) for every participant. Right: Individual MISC scores at the end of the experiment (t = 33 min) for every participant. Scores for the car are displayed on the x-axis and for the simulator on the y-axis. Dotted lines represent least-squares lines to the scatterplot, red lines represent the values where scores in the car and the simulator are equal. The data are shown for 24 participants, an apparent smaller number of points on the graph is due to overlap as scores were integer values.

### 4.2.2 Time to sickness

Figure 4.5 (left) shows the percentage of subjects with onset of moderate nausea versus time, or *dropout rate* for both the car (blue) and the simulator (red). This rate, the number of participants wishing to terminate the experiment due to nausea (MISC = 7), was 14 (58%) at the end of the car drive and 0 (0%) for the simulated drive. Except for participant FJ, all dropouts occurred during provocative driving (coloured areas) of the car drive (blue line). Two participants requested termination of the experiment as a result of severe nausea, with final rated levels of MISC = 5 and MISC = 6. In the calculation of figure 4.5 (left), this MISC score is used to determine the dropout time. The relatively high percentage of dropouts in this condition as compared to the simulator suggests that obtained SS in this simulator configuration is lower than actual MS.

Figure 4.5 (right) shows the percentage of participants experiencing first real symptoms of sickness, versus the onset time of these symptoms. This *Sickness Incidence* (SI) [24, 160], is defined as the percentage of subjects rating a MISC score ≥ 3, as this score accompanies the first distinctive symptoms of MS. Subjects are more certain about their symptoms starting at MISC = 3. Participant LL went from MISC = 2 to MISC = 6, the time instance of the first scored 6 is taken to calculate the onset of sickness graph. It can be seen that during the first sickening section of the experiment, in both conditions the onset of first symptoms for several participants occurred, after which it stabilizes and reaches a plateau in both conditions. A lower percentage of the participant subset reached the Sickness Incidence in the simulator (46%) than in the car, where 92% developed these symptoms.

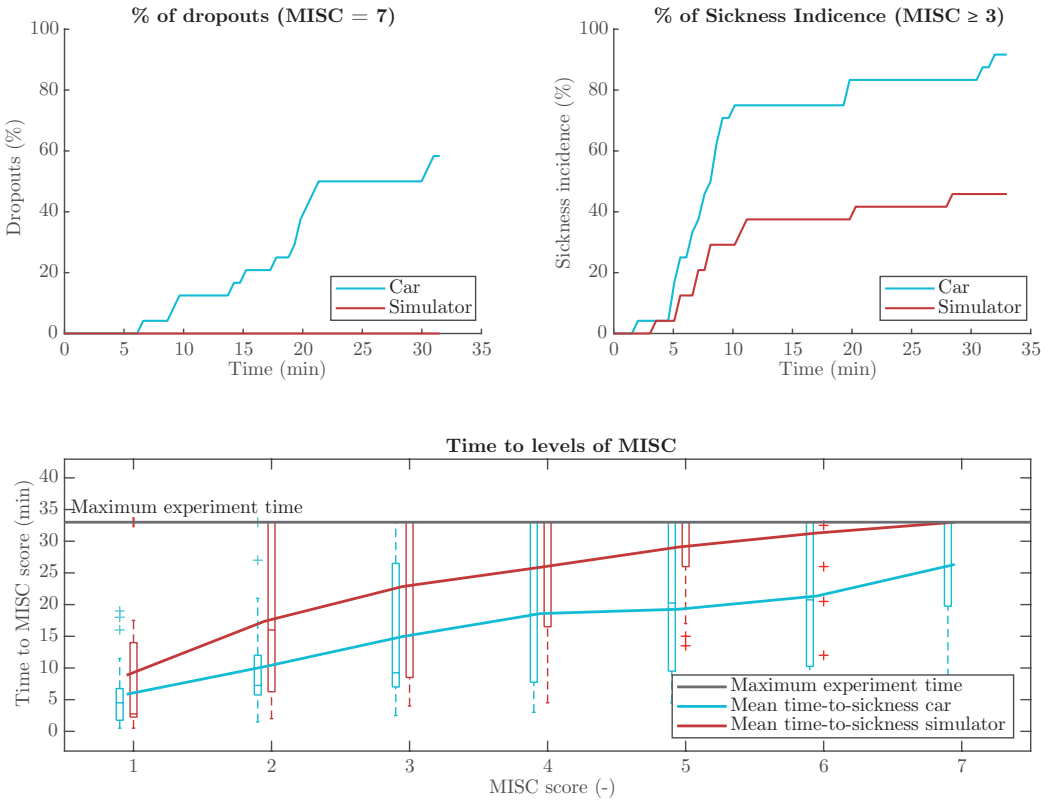


Figure 4.5: Top left: Plot of %-dropouts (MISC = 7) over experiment time for scenario A (blue) and scenario B (red). Top right: %-Sickness Incidence (MISC ≥ 3) over experiment time. Bottom: Time-to-sickness for 24 participants for each of the 7/10 MISC levels used in this study, bold lines represent subject averages per scenario. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually with the '+' marker symbol.

Both these indicators (dropouts and sickness incidences) and the *time to sickness* for other MISC levels for 24 participants are visualized in figure 4.5 (bottom). Again, for participant LL who proceeded from MISC = 2 to MISC = 6 in one timestep, no time data is available between 2 and 6 and these times are not used in the boxplots. From this figure it can be observed that time to sickness increases almost linearly over different MISC levels for both scenarios. However, from the seemingly constant offset between times of both scenarios, it is shown that shorter times are obtained in the car. Lower intercepts for the car drive, suggests that this real road drive was experienced sickening, more quickly than in the simulator.

A three-factor ANOVA (MISC level × scenario × order) is performed. The ANOVA shows that there exists no significant main ($p = 0.565$) or interaction effect ($p = 0.142$) of order. No other significant interaction effects are present. Times to sickness differ significantly between MISC levels [$F(6,308) = 17.14$, $p = 0.000$]. Also, it takes participants significantly longer to reach certain sickness levels in the simulator, as compared to the car [$F(1,308) = 4.80$, $p = 0.029$], for original $p < 0.05$. As a result of Bonferroni corrections, this finding becomes non-significant. Cross-correlation on the averaged outcomes for both scenarios as visualized in figure 4.5 (bottom), results in synchronization at an offset of 0. A strong significant correlation of $R_{XY} = 0.97$ exists between the car and the simulator. The non-significant difference (after Bonferroni correction) between the times to sickness in the car and the simulator and strong correlations between both, indicate absolute validity.

### 4.2.3 Mean MISC group response

The *mean MISC* over all participants per time instance is visualized in figure 4.6 for the car (blue) and the simulator (red). Throughout the experiment time, the MS scores (blue line) are higher than SS scores (red line). It can be observed in figure 4.6 that mean MISC scores over time are comparable during the first time epoch (0-5 min), further increase during the provocative driving sections (coloured areas), and decrease or plateau during normal driving for the car scenario. For the simulator scenario, a similar but less noticeable effect is present. One participant wished to terminate the car scenario due to severe nausea at MISC = 5 (instead of 7). It is decided to extrapolate his MISC scores after the termination timestamp also with scores of 7 as this value is probably closer to what he felt than when extrapolating with 5. Furthermore, extrapolation of the remainder of the experiment time with 5 would cause an underestimate of the real mean sickness.

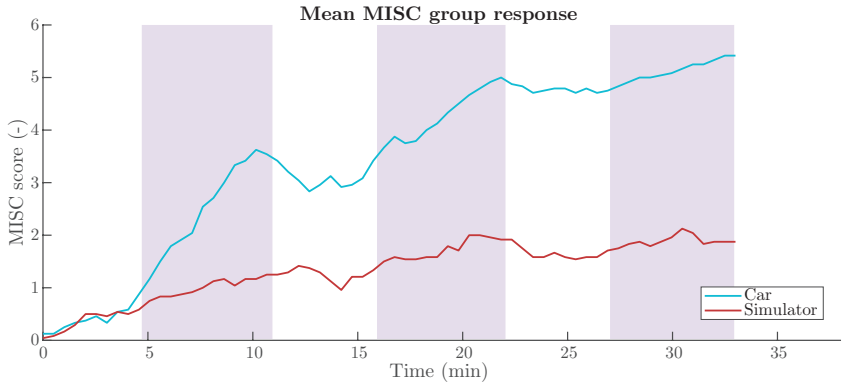


Figure 4.6: Mean MISC group response versus time (min) for the car (blue) and the simulator (red) during normal (white) and provocative (coloured) driving.

A two-way ART ANOVA (scenario × order) is conducted. The ART ANOVA shows no main ($p = 0.122$) or interaction effect ($p = 0.311$) of order. It is found that mean MISC values are significantly higher in the car than in the simulator [$F(1,259) = 297.55$, $p = 0.000$]. Cross-correlation on the outcomes for both scenarios is most synchronized at an offset of 0. Again, a very strong significant correlation of $R_{XY} = 0.96$ is found. The combination of a significant difference and a high correlation between the car and the simulator, results in relative validity for the mean MISC group response.

### 4.2.4 MISC rate

The individual *MISC rate*, increase or slope for each time epoch in the experiment, is visualized by means of the boxplots in figure 4.7. An effect with increases and decreases during respectively provocative and normal driving is shown in this figure, its effect is very obvious for the car and less clear for the simulator. As a result of the integer nature of MISC scores, mean results for epoch 1 and epoch 5 are identical for both scenarios. A three-way ART ANOVA (driving style × scenario × order) is conducted. Results for the ART ANOVA show no significant main ($p = 0.749$) or interaction effect ($p = 0.520$) of order. The magnitude of the MISC rate is significantly different between the car and the simulator, as a result of driving style [$F(1,248) = 19.15$, $p = 0.000$]. Further analysis of this spreading interaction effect reveals that MISC rate for normal driving epochs is almost equal for the car and the simulator. It is found that the MISC rate was significantly higher during provocative driving, than during normal driving [$F(1,248) = 14.60$, $p = 0.000$]. The rest epoch was not used as a driving style in this calculation. Again, the rate of sickness was significantly higher in the car, as compared to the simulator [$F(1,248) = 7.07$, $p = 0.008$]. Cross-correlation analysis between averaged MISC rates for both scenarios results in a very strong positive correlation of $R_{XY} = 0.95$ at a lag of 0. The combination of a significant difference between rates of sickness in the car and the simulator and strong correlations, indicate relative validity for this metric.

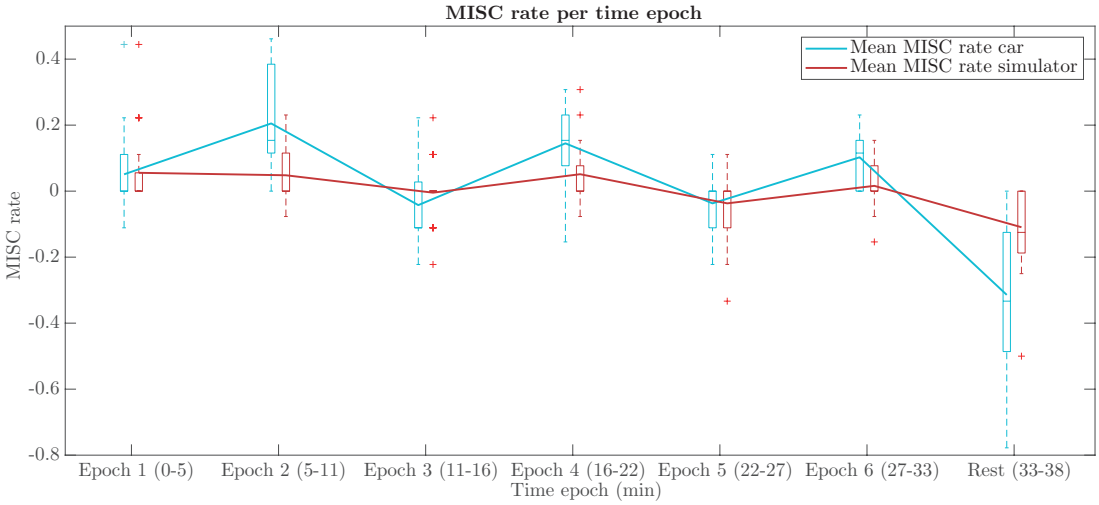


Figure 4.7: MISC rate per time epoch for 24 participants, bold lines represent subject averages per scenario. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually using the '+' marker symbol.

| | **Car** mean ± std | **Simulator** mean ± std | (ART) ANOVA | Correlation |
|---|---|---|---|---|
| Individually averaged MISC | 3.43 ± 1.94 | 1.33 ± 1.26 | $F(1,44) = 30.23$, $p = 0.000$** | $r = 0.57$, $p = 0.004$** |
| Individual MISC at end | 5.42 ± 2.32 | 1.88 ± 1.85 | $F(1,44) = 54.26$, $p = 0.000$** | $r = 0.52$, $p = 0.009$** |
| Time to MISC levels | | | $F(1,308) = 4.80$, $p = 0.029$* | $R_{XY} = 0.97$ |
| Mean MISC group response | 3.43 ± 1.94 | 1.33 ± 1.26 | $F(1,259) = 297.34$, $p = 0.000$** | $R_{XY} = 0.96$ |
| MISC rate | | | $F(1,248) = 7.07$, $p = 0.008$** | $R_{XY} = 0.95$ |

Table 4.2: MISC metric statistics for both scenarios, * $p < 0.05$ (original). **$p < 0.01$ (Bonferroni corrected)

## 4.3 MSAQ

Figure 4.8 shows the individual MSAQ post-experiment responses plotted per participant for both the car (blue) and the simulator (red), for 23 participants. Data for participant KW and NL is left out of this figure as MSAQ data for only one scenario is recorded thus comparison is not possible. Again, large differences between subjects can be seen for several subscale scores, indicating a large variety in experienced sickness. Since 5 statistical tests are conducted on the MSAQ dependent variables, the Bonferroni type adjustment suggests statistical significance for $p < 0.01$.

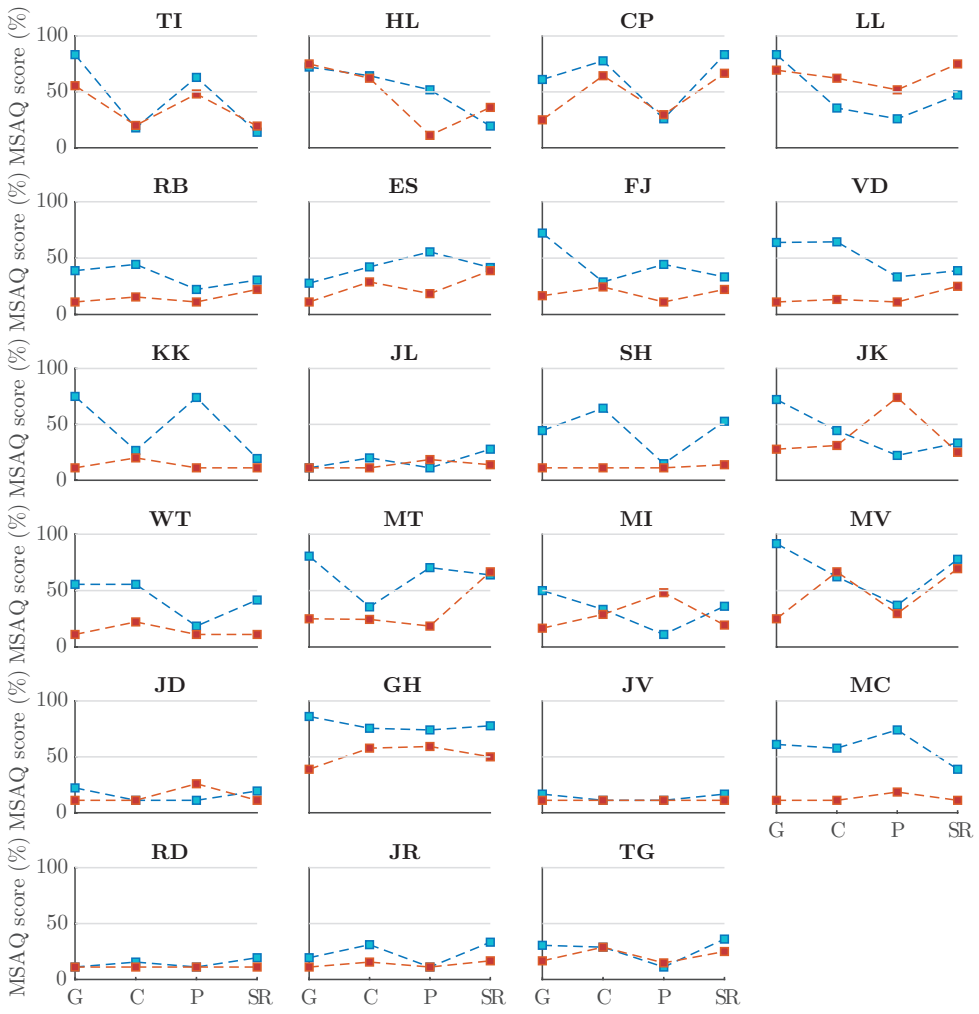


Figure 4.8: Individual sickness responses of 23 participants with the post-experimental MSAQ subscale scores (%) versus categories G (Gastrointestinal), C (Central), P (Peripheral) and SR (Sopite-Related). Dotted lines are added between different subscale scores for visual interpretation.

### 4.3.1 Overall MSAQ score

A two-way ANOVA (scenario × order) on the interaction between the two scenarios assessed directly after the experiments is performed. The *overall MSAQ score* in the car is rated with a mean (± std) of 42.5 ± 17.7 and for the simulator as 26.5 ± 16.4. No significant main ($p = 0.744$) and interaction effect ($p = 0.139$) of order was found. MSAQ total scores were significantly higher in the car as compared to the simulator [$F(1,42) = 27.53$, $p = 0.000$]. Figure 4.9 (right) shows

this holds for every participant, except for participant LL. This outlying result can be explained by looking at individual question scores. LL specifically rated her fatigue and dizziness levels very high in the simulator as compared to the car, possibly attributed to another cause. A strong correlation of r = 0.64 (p = 0.002) was found between MSAQ scores in the car and the simulator. The combination of significant difference and a high correlation, implies relative validity between both conditions.

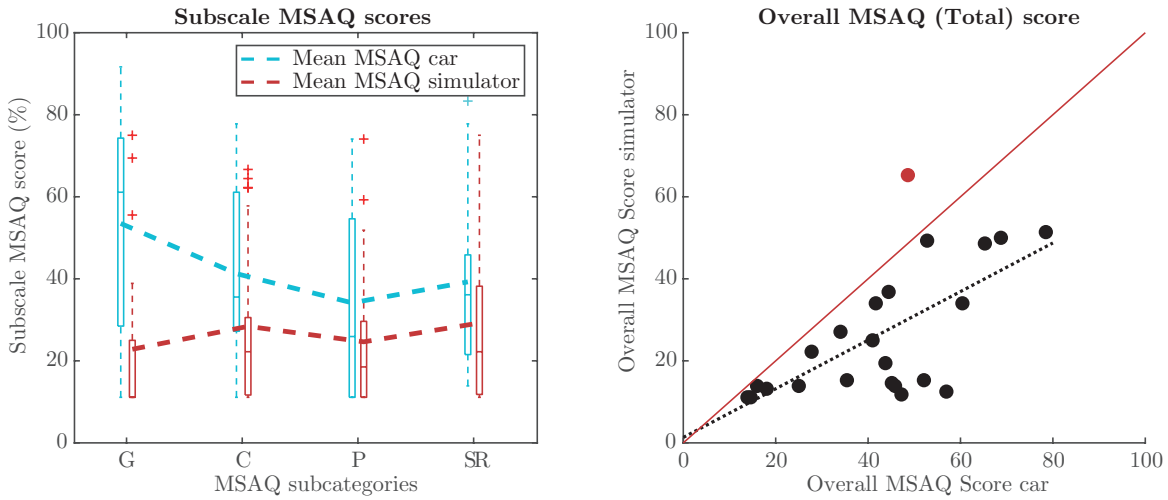


Figure 4.9: Left: Subscale MSAQ scores (%) averaged over 23 participants versus MSAQ subcategories. Bold dotted lines are added between different subscale scores for visual interpretation. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually using the '+' marker symbol. Right: Correlation of post-experiment overall MSAQ scores within participants for both scenarios with the car on the x-axis and the simulator on the y-axis. Dotted lines represent least-squares lines to the scatterplot, red lines represent the values where scores in the car and the simulator are equal.

### 4.3.2 Subscale MSAQ score

Figure 4.9 (left) shows *Subscale MSAQ scores* for the four categories of all participants in boxplots. It can be seen that for the car, the averaged symptom profile results in Gastrointestinal > Central > Sopite-Related > Peripheral. For the simulator, this profile shows Sopite-Related > Central > Peripheral > Gastrointestinal. Only responses for Gastrointestinal suggest a large variance. Figure 4.10 provides more insight into correlations between the car and the simulator for individual subcategories. From the figure can be seen that for most participants, scores in the Gastrointestinal and Central category are significantly higher in the car than in the simulator. Regarding the Peripheral category, neither the car nor the simulator obtained the highest scores as individual responses lie under and above the red line. For the Sopite-related category, responses in the car and the simulator are fairly equal.

A two-way ANOVA (scenario × order) for MSAQ subscores is performed. For MSAQ-G, significantly higher results are found for the car than for the simulator scenario [$F(1,42) = 23.37$, $p = 0.000$]. No significant effect of the order in which the scenarios are experienced ($p = 0.314$) and the interaction of order and scenario ($p = 0.563$), on the MSAQ-G scores of the scenarios can be concluded. For MSAQ-C, significantly higher results are found for the car than for the simulator scenario [$F(1,42) = 15.83$, $p = 0.000$]. No significant effect of order ($p = 0.677$) or interaction effect between the order of scenarios experienced and the scenario scores ($p = 0.058$). In the MSAQ-P category, significantly higher results are found for the car than for the simulator scenario [$F(1,42) = 9.11$, $p = 0.004$]. A non-significant effect for order ($p = 0.306$) and a non-significant interaction between the order and scenario ($p = 0.017$) is found. For MSAQ-SR, significantly higher results are found for the car than for the simulator scenario [$F(1,42) = 7.94$, $p = 0.007$] and no significant effects for order ($p = 0.220$) or interaction between experienced order of conditions and the scenario ($p = 0.678$) is found. The accompanying table 4.3 shows the correlations for each category. It can be seen that 3 out of 4

subcategories obtain significant correlations. This results in relative validity for all but MSAQ Peripheral (sub)scores. These results suggest that participants who obtain high scores rating a symptom in the simulator, would do so as well in the car.

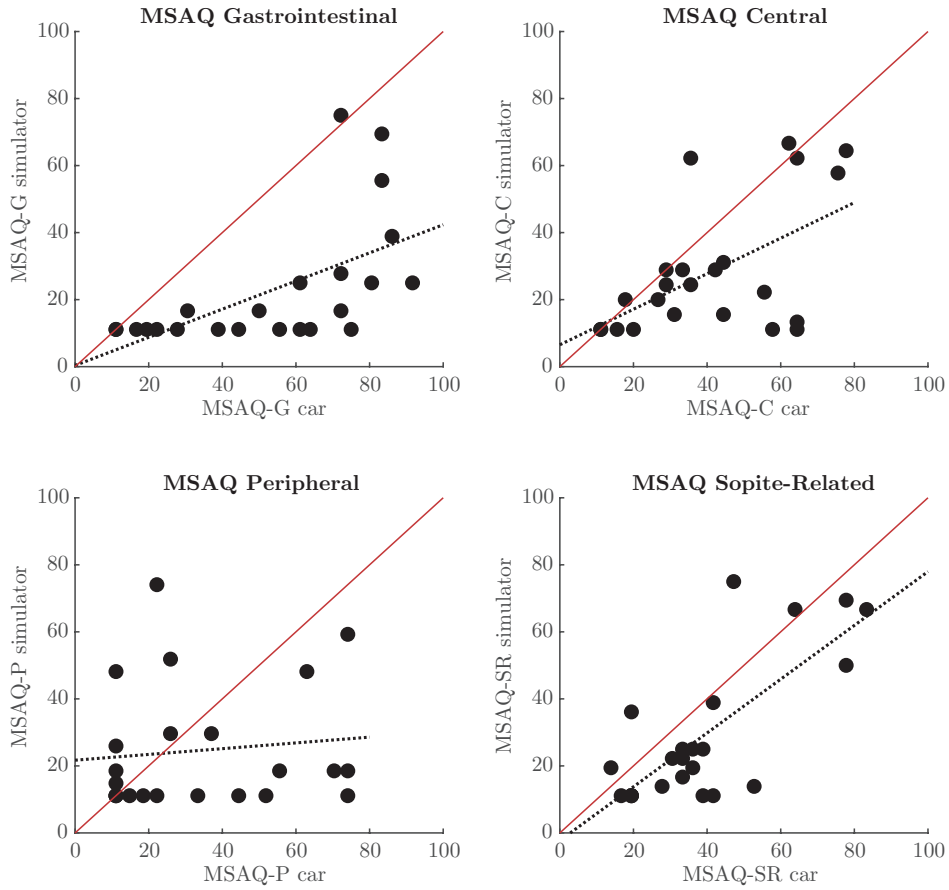


Figure 4.10: Subscale MSAQ scores (%) in 4 scatterplots for 23 participants in the car (x-axis) and the simulator (y-axis). Accompanying correlations can be found in table 4.3.

| | **Car** mean ± std | **Simulator** mean ± std | ANOVA | Correlation |
|---|---|---|---|---|
| Total MSAQ | 42.5 ± 17.7 | 26.5 ± 16.4 | F(1,42) = 23.53, p = 0.000** | r = 0.61, p = 0.002** |
| Gastrointestinal | 53.5 ± 26.4 | 22.8 ± 19.1 | F(1,42) = 23.37, p = 0.000** | r = 0.73, p = 0.000** |
| Central | 51.6 ± 25.7 | 35.5 ± 24.5 | F(1,42) = 15.83, p = 0.000** | r = 0.48, p = 0.020* |
| Peripheral | 25.6 ± 17.9 | 18.5 ± 14.0 | F(1,42) = 9.11, p = 0.004** | r = 0.18, p = 0.402 |
| Sopite-Related | 39.3 ± 20.0 | 29.2 ± 21.4 | F(1,42) = 7.94, p = 0.007** | r = 0.61, p = 0.002** |

Table 4.3: MSAQ total and subscale statistics for both scenarios, * $p < 0.05$ (original). **$p < 0.01$ (Bonferroni corrected)

## 4.4 GSR

Normalized raw GSR data are plotted per participant for both scenarios can be found in figure 4.11. The original data signal retrieved from participant TI and JV cannot be recovered due to a Bluetooth configuration artefact for one scenario. For participant NL, no physiological data in the car has been recorded. For above-stated reasons, no comparison between both scenarios for these participants can be made and they are excluded from further results.

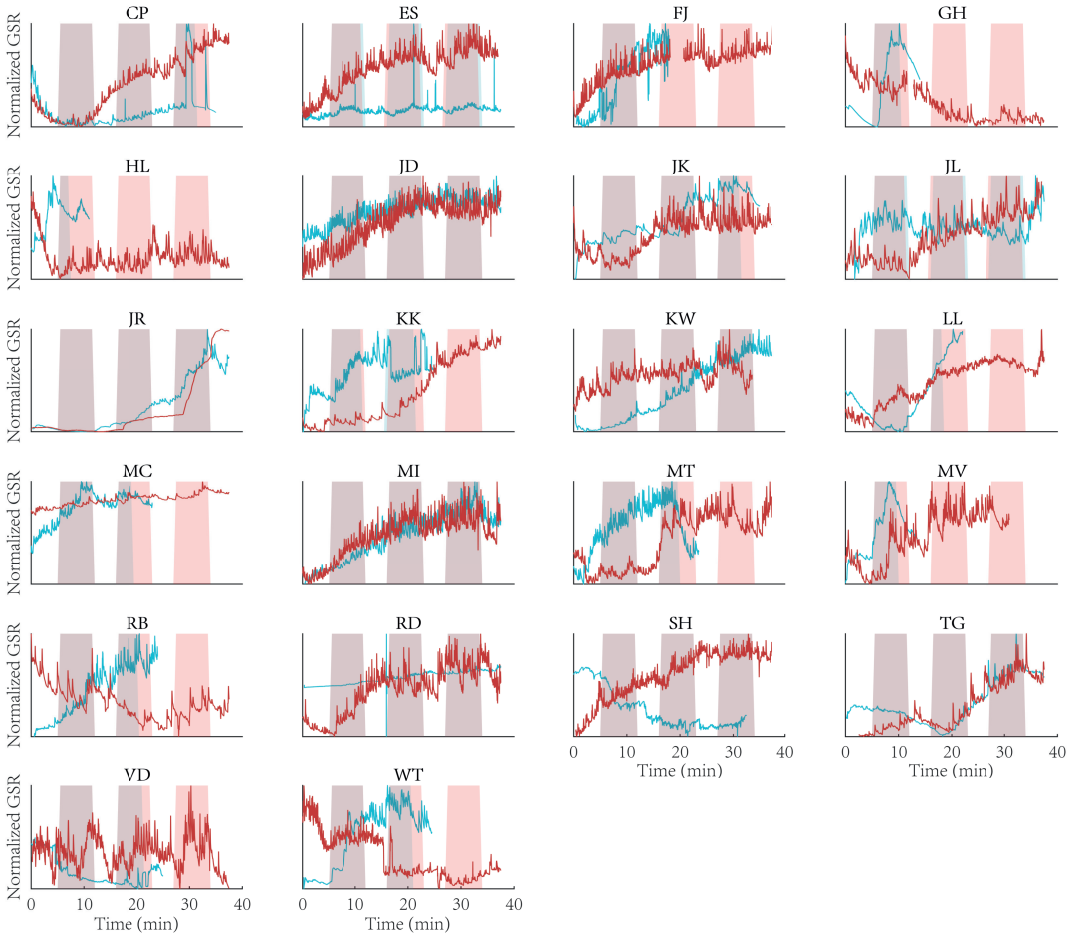


Figure 4.11: Individual sickness responses for the car (blue) and simulator (red) of 22 participants with the raw (normalized) GSR recordings versus experiment time. Red areas denote provocative driving sections in the simulator scenario, blue areas denote provocative driving in the car scenario. When an area is (partly) double coloured, it means that the participant experienced the specific part of the route in both conditions. Every exposure is ended with 5 min of rest.

The MISC metric in section 4.2 shows clear alterations of sickness response over time and driving styles. Figure 4.11 shows that GSR changes as a result of driving style or scenario, are not as clearly visible. However, several studies state skin response to be a very reliable and indicative measure of MS [64, 110, 149]. Figure 4.12 shows GSR and MISC data combined for the car (blue) and the simulator (red) for three representative participants. It can be seen that for some participants (e.g. MV), normalized skin responses behave almost identical to their self-reported MISC levels. For some participants (e.g. MI), the signal is still very noisy after filtering which makes it hard to detect temporal changes. Finally, for some participants (e.g. KK) GSR data increased, whereas their synchronized MISC scores did not. This effect is particularly clear for the simulator scenario (red), where the GSR signal doubled in value (increased with

10 µS) and no changes in MISC are reported. This can be the result of many other external or internal physiological responses, such as stress.

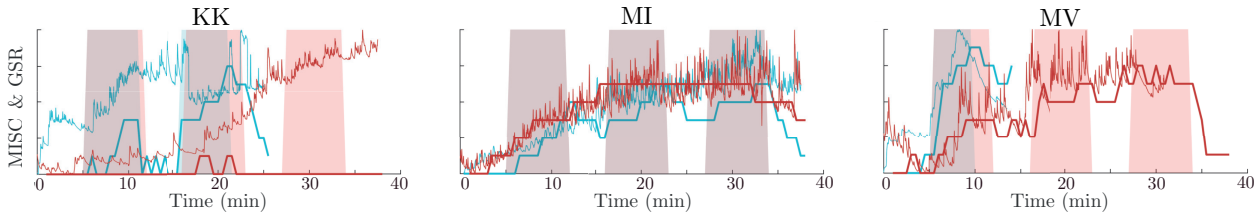


Figure 4.12: Individual MISC responses for the car (blue) and simulator (red) of 3 participants are shown in bold lines, synchronized with their raw (normalized) GSR recordings, versus experiment time. Red areas denote provocative driving sections in the simulator scenario, blue areas denote provocative driving in the car scenario. When an area is (partly) double coloured, it means that the participant experienced the specific part of the route in both conditions.

*Normalized skin responses*, calculated as described in the previous chapter, are for 22 participants per epoch of both scenarios shown in figure 4.13. The bold lines in this figure represent the mean values over all participants. The data show lower GSR values for the simulator, compared to the car. A two-way ART ANOVA (scenario × driving style) test is conducted. Results indicate a significant interaction effect between scenario and driving style [$F(1,230) = 5.55$, $p = 0.019$], whereby the rest epoch was not used as a driving style in this calculation. Further analysis of this spreading interaction effect shows that for both the car and the simulator the Δ GSR increases, but the slope thereof is steeper for the car condition as compared to the simulator. No significant difference can be found between normal and provocative driving [$F(1,230) = 2.95$, $p = 0.087$], but GSR results showed to be significantly higher in the car condition, compared to the simulator [$F(1,230) = 15.58$, $p = 0.000$]. Cross-correlation on the outcomes for both scenarios is most synchronized at an offset of 0. A very strong significant correlation of $R_{XY} = 0.96$ is found at this offset. From these results, relative validity for GSR can be concluded.

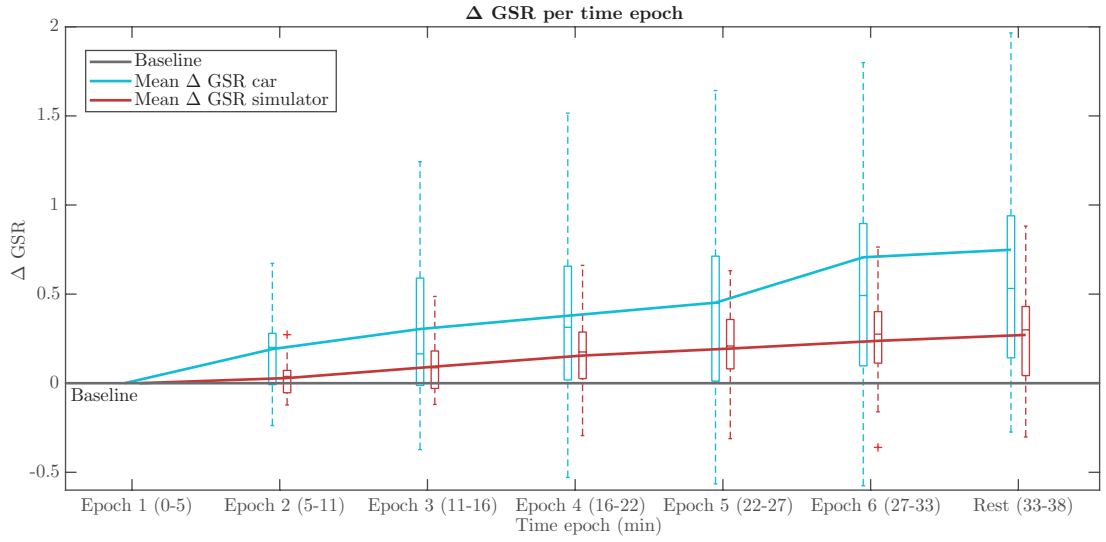


Figure 4.13: Plot of Δ GSR over time epochs (min) for the car (blue) and the simulator (red) for 22 participants. Subject averages are shown with bold lines for both scenarios. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually with the '+' marker symbol.

## 4.5 EGG

Power for raw EGG data in the frequency domain can be found for all participants in figure 4.14. Each line in each subplot represents a time epoch response, where blue lines indicate time epoch responses for the car and red for the simulator. Thus, if a participant completed the whole experiment time, 7 blue and 7 red lines are shown. Light coloured lines represent early epochs, darker lines indicate the final time epochs of the experiment. The healthy DF region is represented by vertical lines between 2-4 cpm, left and right of which occur bradygastria and tachygastria, respectively. As a result of the same artefact as mentioned for the GSR data, no EGG data of participants JV and TI for the car drive are analysed, and car data for participant NL is not recorded. These three participants are left out of the analysis. In line with literature that states tachygastric (4-9 cpm) power to increase with increasing sickness [69, 139, 161], for most subjects normalized EGG power should obtain higher values for blue lines (car) than red lines (simulator). And, as sickness increased throughout the exposure time for most people, darker coloured lines for these participants should obtain higher values (in the 4-9 cpm region) than lighter coloured lines per scenario. Both of these effects can only be found for a few participants, for example for participant VD.

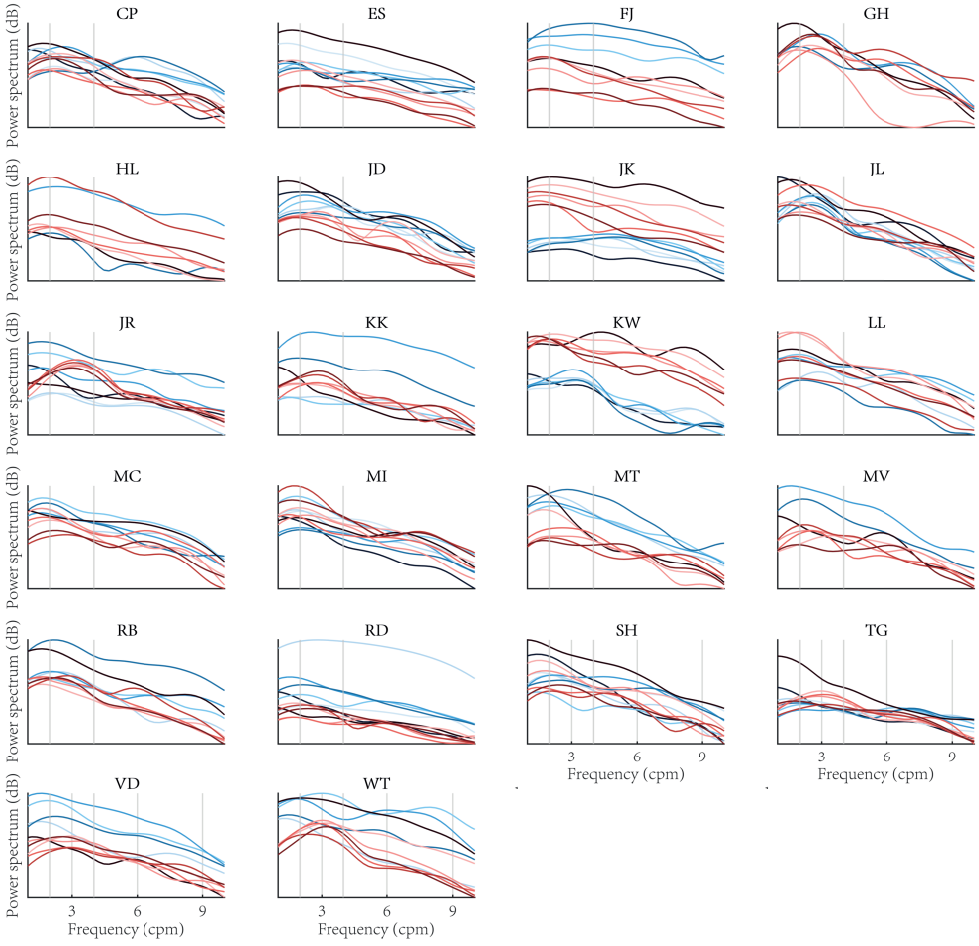


Figure 4.14: Individual normalized raw EGG power spectrum responses (dB) versus frequency (cpm) for 22 participants. Each line indicates a time epoch response for the car (blue) and the simulator (red). A lighter coloured line corresponds to an earlier time epoch and darker lines represent later stages in the experiment. Vertical grey lines represent the healthy frequency range (2-4 cpm).

A closer examination of EGG amplitude, power spectrum and band power of individual participants provides insight into this matter. Figure 4.15 (top) shows the data-trace for participant ES during the first and final immersion epoch of the experiment for both scenarios. Literature suggests that higher amplitudes can be observed when a participant experiences increased sickness [137]. However, the increased amplitude can also hint at talking, coughing, repositioning on the car seat or delayed digestion of previously consumed goods. It is manually observed from several experimental drives that some of these events occurred and that these factors can only be limited to a certain extent. Especially in the first epoch (0-5 min) and during the rest epoch, many participants moved, coughed or talked. The result hereof can be seen in figure 4.15 (bottom left), which shows the power spectrum for the same participant in the same experimental epochs for both scenarios. As a result of extensive coughing in the simulator (red) during the first minutes, this epoch for participant ES obtains the highest power. The suggested effect of increase of 4-9 cpm frequency power with increased MS, cannot be seen. Figure 4.15 (right) shows the percentage of band power in the tachygastric region over time over the total 1-10 cpm power, where a dependency on the driving style and an overall increase over time (if experienced sickness increases), is expected but not observed.

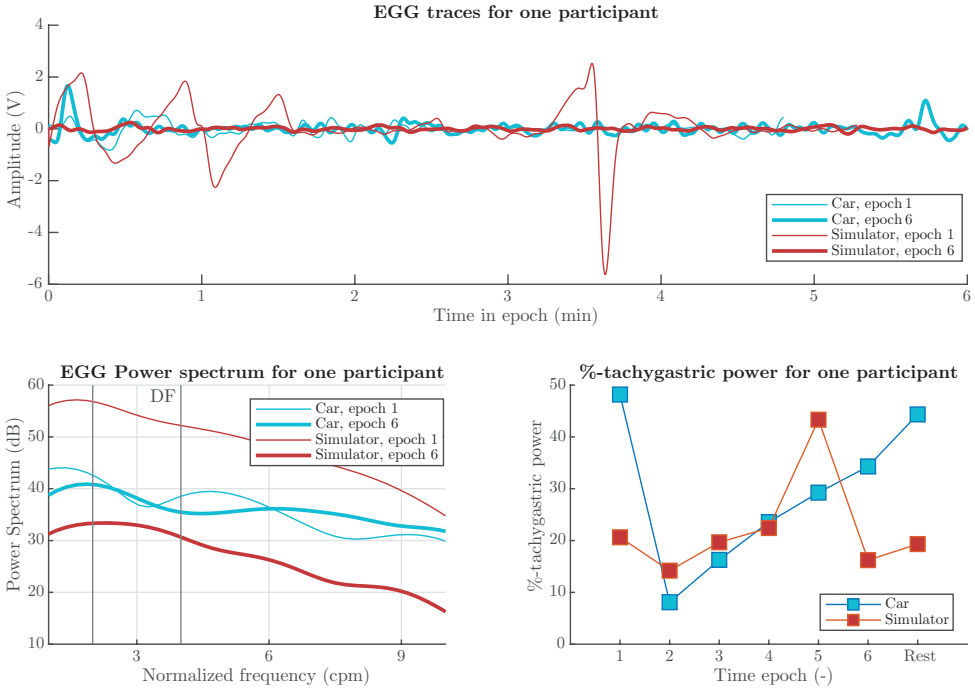


Figure 4.15: Top: Plot of raw EGG traces from one subject during epoch 6 (27-33 min) for scenario A (blue) and scenario B (red). Left: Accompanying power spectra of the data for epoch 6 (bold), compared to epoch 1 (regular). Right: Normalized power in the tachygastric (4–9 Hz) band throughout the whole experiment (epochs 1-7).

It is decided to remove for all participants, these first and rest epochs. During the remaining time (5-33 min), participants were mostly focused and artefacts are mostly minimized. *Percentage tachygastric and bradygastric power* over time (epochs) for all participants per epoch of both scenarios, is shown in figure 4.16. The bold lines in this figure represent the mean values over all participants. When averaged, the data do show increased tachygastric band activity for the car as compared to the simulator. A two-way ANOVA (scenario × driving style) on these 5-33 min experiments is conducted. As two tests are conducted, the Bonferroni Correction suggests significance for $p < 0.025$. Tachygastric (figure 4.16 (top)) values do not significantly differ for normal driving as compared to provocative driving [$F(1,186) = 1.21$, $p = 0.273$] and no significant differences are found for the interaction effect of the driving style with the car or simulator condition [$F(1,186) = 0.26$, $p = 0.610$]. Responses for EGG tachygastria were significantly higher in the car as

compared to the simulator [F(1,176) = 7.62, p = 0.006]. For bradygastria (figure 4.16 (bottom)), ANOVA analysis shows no significant interaction effect [F(1,186) = 0.05, p = 0.819] and no significant effects for either driving style [F(1,186) = 0.18, p = 0.672] or scenario [F(1,186) = 0.24, p = 0.628]. Cross-correlation on the outcomes for both the car and the simulator, are most synchronized at an offset of 0. A moderate negative correlation of $R_{XY}$ = -0.39 is found when comparing both in the tachygastric band range and a positive correlation $R_{XY}$ = 0.36 in the bradygastric band range. After removal of epoch 1 and the rest epoch, correlations change to $R_{XY}$ = -0.10 and $R_{XY}$ = 0.70 for respectively tachygastria and bradygastria.

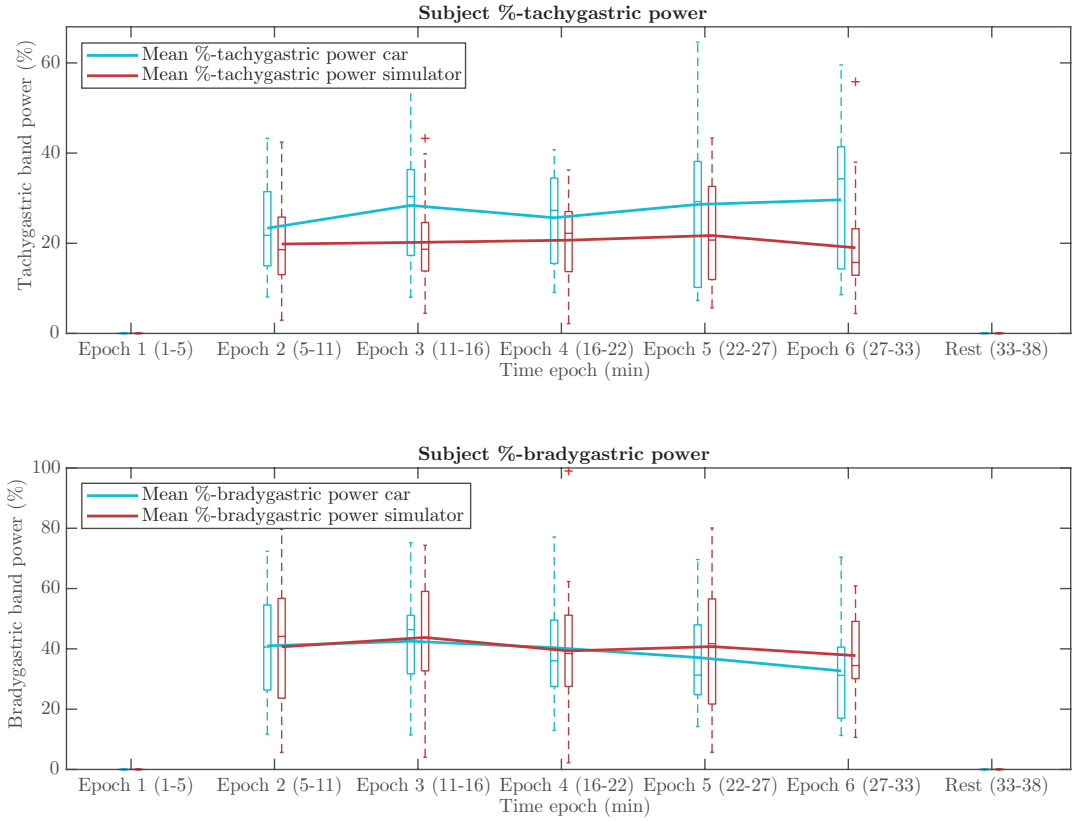


Figure 4.16: Plot of tachygastric (top) and bradygastric (bottom) band power (%) over time epochs (min) for the car (blue) and the simulator (red) for 22 participants. Subject averages are shown with bold lines for both scenarios. Time epoch 1 (0-5 min) is not shown as a result of consequent outlier data. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually with the '+' marker symbol.

An increase in %-tachygastric power and the higher values thereof in the car, do show the expected effects. The presence of significantly higher values in the car as compared to the simulator, accompanied by a moderate correlation, results in relative validity. Also, the seldomly validated %-bradygastric power decreases with increasing sickness for both the car and the simulator, as proposed by Dennison et al. (2016) [69]. We thus share the opinion of the authors, that the decrease of bradygastric band power is present but less clear than the tachygastric power increase. Absence of significant differences between the car and the simulator for bradygastria and a moderate positive correlation result in absolute validity. However, considering the occurrence of the aforementioned events and the absence of a validated relation between MS and bradygastric activity, this conclusion should be interpreted with caution. Magnitude of both the effects (increase and decrease) is subtle, nevertheless in accordance with studies with comparable experiment times [62, 69].

# 5

# Discussion

This study investigated the validity of a driving simulation environment for research on motion sickness, by comparing sickness obtained in a complete motion-base simulator (SS) to motion sickness (MS) in a real (autonomous) car. Although either sickness could be caused by a similar neural conflict, fundamental differences between both complicate the validation of such studies. To the best of our knowledge, this is the first study where the same participants (who represented a broad spectrum of MS susceptibility) are subjected to both sicknesses in a real and in a simulated scenario, designed to perceive identical stimuli. A cohort of Sudoku solving "passengers" was exposed to vestibular and peripheral visual cues in both the car and the simulator condition in a balanced order. Both drives existed of a combination of normal and provocative (slalom and stop-and-go) manoeuvres to create a distinctive pattern in (increasing and decreasing) sickness over time. Sickness was quantified using two subjective and two objective measures, which were then tested on absolute, relative or no validity at all. For all responses, (at least) relative validity for the simulator can be concluded, thus adopting the first hypothesis that stated this. Also the second hypothesis, stating higher sickness as a result of provocative driving and recovery as a result of normal driving, was found true for 2 out of 3 measures. No main and interaction effects of the order of the experienced scenarios were found, adopting also the third hypothesis. In the remainder of this chapter, findings in relation to each of the subquestions will be discussed. Lastly, the setting and conditions under which MS can be studied using a motion-base driving simulator will be discussed.

## 5.1 Are symptoms as a result of SS, different from classical MS?

The frequently employed MSAQ [56, 57] is used to obtain a profile existing of (a sequence of) four symptom categories. In this study, post-experiment mean MSAQ values (figure 4.9) show the highest scores for Gastrointestinal (nausea, vomiting) for the car scenario and the highest scores for Sopite-related (tired, uneasy) and Central (dizzy) in the simulator. These findings are in line with the studies by Kennedy et al. (2010) [70] and Stanney et al. (1997) [71], who concluded nausea as most prominent for the real environment but not for simulated environments. Regardless of this correspondence with previous research, we believe that the difference in gastrointestinal scores in theirs and our study underlies the severity of the compared conditions. A low averaged individual MISC score of 1.33 ($\pm$ 1.26) in the simulated environment, indicates that participants had simply not progressed to severe sickness, thus had not (yet) obtained high MSAQ nausea ratings. Nausea and vomiting (being components of the gastrointestinal category) are typically the most notable "everyday" symptoms of MS [135]. Not only is vomiting and nausea what people associate with carsickness (from past experiences), also vomiting lies unanimously at the end of MS rating scales [27, 42, 58]. Furthermore, it can be observed that when sickness in an environment is indeed severe, such as in the study by Gavgani et al. (2018) where participants lasted around 6 min, then Gastrointestinal is also the dominant category in the simulated condition [68, 97].

Strong significant correlations were found between the car and the simulator for Gastrointestinal, Central and Sopite-related (3 out of 4) MSAQ categories. From the reference lines in figure 4.10, it can be seen that Gastrointestinal scores are approximately twice as high in the car, and that Central and Sopite-related categories obtain more comparable scores when comparing the car to the simulator. The sole non-significantly correlated category Peripheral concerns being sweaty, clammy or warm. A possible explanation for the only odd result addresses the (outside) experimental conditions. Even though the temperature in the car was intended to be kept approximately constant for every participant, the simulator workshop temperature was harder to control. Outside temperatures varied during this experiment period between 20°C to just above the freezing point, suggesting the possibility that a participant experienced both conditions in very different ambient temperatures. Concluding, the answer to this question can be answered as depending on the specific condition and perceived severity of sickness. It is speculated that when the latter is similar in both the car and the simulator, not only symptoms but also symptom profiles will show resemblance.

## 5.2 Does a relationship exist between individual sensitivity to MS and SS?

Individual sensitivity can be evaluated by comparing correlations of sickness levels for both the car and the simulator scenario, over the whole exposure period. Measures in this study that represent this individual susceptibility, are post-experiment overall (total) MSAQ scores and the individually (averaged) MISC scores. Strong positive correlations regarding the overall MSAQ score, individually averaged MISC scores and the final MISC score, were found. These significant correlations in sensitivity measures imply that participants who would report high scores succeeding the simulator drive, would do so as well following the car exposure. This finding is in line with the study by Gavgani et al. (2018) [68], who concluded a significant relation between sensitivity to VIMS and MS. For these three subjective measures, scores obtained in the car are approximately twice as high as in the simulator. This can be observed by comparison of least squares regression lines (black) with the MS = SS reference line (red) in figures 4.4 (both) and 4.9 (right).

A percentage of participants in this study show no (or very little) effect to either provocative stimulus. Several studies [6, 24, 159] concluded it to be inevitable that a part of subjects is insensitive to the exposure. Also, a percentage of participants show only effect to the car stimuli and not to the simulator motion. We again believe that this effect underlies the severity of sickness of both conditions and that their "threshold" [32, 162, 163] beyond which disease develops, has not yet been reached in the simulator. Nevertheless, the answer to the question of this section is affirmative. Although levels of self-reported MS are significantly higher than those for SS in this study, a strong relationship between sensitivity in the car and the simulator exists.

## 5.3 Are temporal aspects of sickness different in the car and in the simulator?

### 5.3.1 Time to onset of sickness

How rapidly an individual experiences sickness, could be an important indicator to take into account when studying MS in a driving simulator. In order to answer the question if the time to onset of SS and MS are different, the time-to-MISC metric will be consulted. Results thereof show a significant difference between both scenarios for the original alpha value ($p < 0.05$), the result however turns non-significant after Bonferroni corrections have been applied. Therefore, even though mean time-to-sickness values are unanimously higher for the simulator with an almost constant offset (figure 4.5 (bottom)), the difference in onset times between both scenarios cannot be concluded. This result is supported by figure 4.5 (right), where it can be observed that a higher percentage of subjects reached distinctive MS symptoms (MISC $\geq 3$) in the car, but that its onset time interval shows resemblance with the simulator. Absolute validity is given to the time to onset of sickness, as a difference between the car and the simulator scenario for this metric cannot be concluded. It is suggested that MS does not necessarily commence earlier or quicker than SS.

### 5.3.2 Increase and decrease of sickness

We expected to find increased sickness during provocative driving sections and a decrease in sickness response during normal driving, in both scenarios. This effect can be validated by evaluating the main and interaction effects of the independent variable driving style for GSR, EGG and MISC rate. A significant main effect of driving style exists for MISC rate, indicating that sickness responses are significantly higher during provocative driving, as compared to normal driving. For both MISC rate and GSR, also a significant interaction effect between scenario and driving style exists. It was found that the size of the effect of executed driving style is higher for the car than for the simulator. For MISC rate, the effect can clearly be seen in an alternating positive and negative slope for both conditions in figure 4.7. For GSR this effect resulted in a higher and lower positive GSR slope. The interaction effect can be explained by the observation that for the normal driving epochs in the simulator, very little motion can be perceived. During these epochs, a constant speed is maintained, resulting in practically no motion in the simulator. Even though auditory and physical road and wind noise was generated, it is expected that immersion was very low during these epochs. For both EGG metrics, no significant main and interaction effect with driving style exist. This can be a result of the aforementioned sensitivities and artefacts (e.g. coughing, talking, repositioning) of the EGG signal, which is why we propose future research to request participants not to eat or drink at all for at least 12 hours preceding the recording. The absolute validity can also be a direct result of the ratio between the very low frequency of stomach contractions and the relatively short lengths of the time epochs. Since normal and provocative time epochs last respectively 5 and 6-7 min and the frequency of interest for gastrointestinal activity is located around 4 cpm, the amount of data per epoch is limited. Therefore, a distinctive increase or decrease in EGG activity would be difficult to observe. Gruden et al. (2021) [137] suggested more expressive results for these metrics for longer (> 15 min) trials and Dennison et al. (2016) [69] concluded that alternative, faster measures would be more successful in sickness estimation. Thus, for future studies interested in increasing and decreasing low frequency (EGG) signals, longer time sections are furthermore proposed. Various studies have shown an increase in motion sickness slope during dynamical (provocative) driving [36, 41, 137], such as slaloming and stop-and-go manoeuvres [17]. It is furthermore known that a decrease of this slope (recovery) can be observed during rest or less dynamical driving periods [32, 58, 67]. With an effect size dependent on the simulator design, it is suggested that SS increases and decreases similarly to MS, as a result of driving style.

### 5.3.3 Amplitude of sickness

A difference in amplitude, intensity or peak of sickness between both scenarios could be evaluated by means of considering all of the main effects of scenario for each of the four measures. For overall MSAQ scores, for mean MISC scores, for GSR values and for EGG tachygastric response, significantly higher values in the car as compared to the simulator were found. No difference in amplitude is found for EGG bradygastria, whereby it has to be noted that a decrease in bradygastric band power is not a noted indicator of increasing sickness [140]. Also the number of dropouts as a result of severe sickness, 14 versus 0 for respectively the car and the simulator scenario supports these results. Our findings are in line with several studies [56, 68], who concluded that SS (VIMS) is generally less severe than sickness in a non-artificial environment. Several reasons could be given to explain this difference. First, we argue that this difference in amplitude is not necessarily always present. Severity of sickness differs greatly between conditions and is in this study a cause of the inevitable simulator constraints. The limited motion space and in particular the scaling factor as part of the chosen MCA might be the most insightful variable to answer this question. Even though accelerations were very similar in design (table 4.1), they are reduced in final output. Figure 4.1 showed the simulator longitudinal and lateral accelerations after washout, compared to the inertial car accelerations. As a result of a.o. scaling, the amplitudes of simulator accelerations are lower than those obtained from the car. Another explanation might be that reduced severity is caused by the shifted peak in sickening frequency (*Difference (iv)*) for a simulator with a limited motion envelope. Future empirical research is necessary to test this suggestion by Kuiper et al. (2019) and Bos (2021) [24, 89], preferably by means of a within-subjects study. It can be concluded that the amplitude of sickness for the car condition was higher in this study, suggesting MS to be more severe than SS. Even though this effect will be likely to occur when using similar motion-base simulators with comparable MCA parameterization, it does not necessarily hold true for every participant or simulated condition.

## 5.4 Can we study MS in a driving simulator?

The answer to this research question depends on the simulator. Firstly, we agree with literature stating the impossibility to study MS in a (low fidelity) fixed-base simulator [24, 89] as a result of the absence of vestibular motion. Sickness that is obtained as a result of driving such a stationary simulator, needs to be referred to as cybersickness or VIMS [44]. Secondly, all of the similarities found in this study combined, show relative validity and the possibility of a motion-base driving simulator to study MS. However, this study used a medium-to-high fidelity motion-base simulator and minimized the perception of visual motion cues, in order to reduce SIS caused by visual ambiguities. A remark should be made, that should one wish to study MS whereby participants actively view the visualization, outcomes can be different. Specific dissimilarities can then be found as a result of SIS (*Difference (i)*) and the underestimation of vestibular cues when attributed with visual cues (*Difference (iii)*). Lastly, we believe that unconditional absolute validity is obtained with the use of a (near) perfect fidelity simulator. When no errors, ambiguities or false cues in the simulator systems are able to generate SIS, MS responses will be similar to SS (SCS). To improve (up until absolute) validity, we suggest three practical conditions. These conditions are proposed in order to decrease the conflict between sensed and expected neural inputs based on previous experience with an actual car (*Difference (ii)*), to increase SCS from the motion modality (*Difference (i)*) and to decrease SIS as caused by the visual modality (*Difference (iii)*).

*(a) Repeated exposure.* Studies largely agree that excessive training or repeated exposure is necessary for humans to adapt to the new dynamics and perceptual cues of the artificial environment [23, 89–91, 164]. Since for most people, the neural storage is not up-to-date regarding driving in a simulator, a sensory conflict is caused by what we know of everyday car driving. References supporting *Difference (ii)* as described in chapter 2, imply that the cues coming from a simulator are simply too new for most people to obtain results similar to those in a car. It has to be noted that SS is most severe for the elderly [84, 85], the magnitude thereof decreasing as more multisensory information is provided [45]. Conversely, we see indicators that (younger) people who grew up with advanced artificial imagery such as games, are better adapted to the visual perception of such artificial content [164]. Repeated (gaming/VR) sessions increase the capability to downgrade the gain of the vestibular perception [54, 165] and thus reduce sickness [166]. In this study, 72% of the participants had little to no simulator driving experience and could roughly be categorized as the same generation with comparable access to modern imagery. An interesting topic for further exploration could be to conduct an experiment whereby one subgroup of participants is very experienced simulator users, in order to determine the effect of familiarization on SS outcomes.

*(b) Simulating 1:1 vestibular motion.* The second condition for absolute validity of the simulator is a necessity to simulate 1:1 motion as it can be perceived in the real world. This requires a high-to-perfect fidelity classification of the simulator motion-base hardware. The condition to provide sufficient provocative motion has already been addressed by Kuiper et al. (2019) as a "prerequisite" to study MS in a simulator [24]. Also, the study by Bellem et al. (2017) [17] concluded that validity of sickness obtained during their longitudinal motion condition was superior to the lateral condition, as it uses the largest simulator stroke and a 1:1 scaling used. In our study, several aspects of SS were very similar to MS, except for the severity thereof. This finding can be ascribed to the reduced simulator vestibular motion. It is predicted that differences in amplitude of MS and SS will show larger coherence when larger displacements are possible. In order to advance fidelity of the simulator hardware, e.g. X and Y rails and a 360° yaw table could be introduced to create larger motion space. Additionally, improving the fidelity of the software is particularly rewarding when motion is close to the actual motion perceived in the real world. MCA options such as tilt and roll coordination could contribute to creating more realistic sustained acceleration behaviour and minimize false cues when larger displacements are possible [119]. Also other MCAs, such as realistic and non-delayed Model Predictive Control (MPC) MCAs, could turn out to be beneficial solutions for some driving scenarios [167]. It is suggested that for small simulators the difference between MCAs is negligible [168], and when a simulator is generally larger, qualities of advanced MPC MCAs might be better-exploited [169].

*(c) Reduction of visual ambiguities.* The third proposed condition for absolute validity, is high-to-perfect visual fidelity, meaning that ambiguities (SIS) arising from the visual modality are reduced to a minimum. *Difference (iii)* describes

the general overestimation of visual cues [89, 92, 170] and the ability to reduce accompanying vestibular information as a result of repeated motion exposures [54]. Complicating is thus, that we expect the effect of SIS as a result of visual distortions to be larger than for physical motion. Furthermore, this condition might be the hardest to achieve as it requires considerable technological advances of which the effect on sickness is yet unknown. Current visualization methods such as displays, VR systems or holograms are still far from being perceived as is done in the real world. This is mainly attributed to the fact that these methods are unable to let human eyes dynamically accommodate and convergence, let alone that the viewer can determine and instantly influence what to focus on. In the study by Kuiper et al. (2019) [24], the visual cues of subjects were obstructed by a blindfold after which they were able to conclude the validity of the simulator regarding MS. In our study, we concluded relative validity, whereby SIS is also minimized by asking participants not to focus on the visual projections but rather on the Sudoku puzzle. The contribution of SIS as a cause of visual ambiguities to sickness responses, therefore, remains an interesting topic for future inquiry as long as new technologies are developed for it.

## 5.5 Limitations

Several issues with the use of the NeXus device for recording EGG responses arose. As addressed before, irregularities in the EGG signal can partially be avoided by instructing the subjects not to eat and drink for at least 12 hours preceding the excitation. Furthermore, comparison between time epochs of 5 minutes subsequently proved to be too short to obtain valuable results. Future research that is interested in the use of this method to record EGG data, is also advised to utilize multiple leads on one participant and choose the lead pair with the least artefacts, as is done in the study by Gruden et al. (2021) [137]. Suggested by Curilem et al. (2020) [144] is furthermore a trained person monitoring the individual subject signal and annotating all events (e.g. movement, talking) during a record. Yin et al. (2013) [136] even stated that it is very likely to be able to accurately record gastric slow waves if the recording device is not specifically designed for it.

This study explains MS and SS from the perspective of the Neural Mismatch Theory [27]. It does not investigate another popular theory, the Postural Instability Theory by Riccio and Stoffregen (1991) [29], as a cause of sickness obtained in these experiments. Future research could relate both MS and sickness as a result of exposure to simulator motion to for example head or body movements via an Inertial Measurement Unit or a balance board.

# 6

# Conclusion

In order to gain better insight into the theories, causes, and temporal aspects of MS, ideally, this would be studied in the safe and reproducible environment that is the driving simulator. Literature suggests differences between sickness obtained in a fixed-base simulator (VIMS), sickness obtained in a motion-base simulator (SS) and everyday motion sickness (MS). This study, therefore, aimed to compare sickness obtained during a real-road drive and simulator sickness obtained during a replica of the same drive in a motion-base driving simulator. With results suggesting at least relative validity for all four measures (MISC, MSAQ, GSR and EGG), an overall relative validity and the possibility for MS research in the moving base simulator can be concluded. We found that symptoms between MS and SS are similar, that a strong relation exists between sensitivity to MS and SS and that both entities are affected by time and driving style similarly. The only prominent difference found, addresses the amplitude or severity of experienced sickness between the simulator and the car. This difference can be mainly attributed to the restraints of the simulator hardware, consequently to lower the intensity of perceived vestibular stimuli. Absolute validity, however, is a long journey ahead. In order to minimize the conflict between what we know of cardriving and what we feel during simulator immersion, we need to aim for excessive simulator training, large simulator motion spaces and elimination of visual ambiguities. Until this future is realized, it is important to keep our eyes on the road ahead.

## 6.1 Research acknowledgements

This research was supported by my mother, who let me use her car to conduct scenario A of the experiment. This research was also supported by Rijksoverheid and Vliegkamp Valkenburg, who allowed for the use and surveillance of their closed-track terrain to conduct scenario A of the experiment. Lastly, this research is supported by Cruden, allowing for their expertise on and the use of the moving base driving simulator to conduct scenario B, as described in this study. The research was funded by Cruden.

# Bibliography


[1] Mario Gerla et al. “Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds”. In: *2014 IEEE world forum on internet of things (WF-IoT)*. IEEE. 2014, pp. 241–246.

[2] SAE-International. *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems (Standard J3016)*. 2014. URL: http://www.sae.org/misc/%20pdfs/automated_driving.pdf.

[3] BBC. *"Self-driving" cars to be allowed on UK roads this year*. Apr. 2021. URL: https://www.bbc.com/news/technology-56906145.

[4] Cyriel Diels and Jelte E. Bos. “Self-driving carsickness”. In: *Applied ergonomics* 53 (2016), pp. 374–382.

[5] Michael Sivak and Brandon Schoettle. *Motion sickness in self-driving vehicles*. Tech. rep. University of Michigan, Ann Arbor, Transportation Research Institute, 2015.

[6] Xiao Dong, Ken Yoshida, and Thomas A Stoffregen. “Control of a virtual vehicle influences postural activity and motion sickness.” In: *Journal of Experimental Psychology: Applied* 17.2 (2011), p. 128.

[7] Michael J Griffin and Maria M Newman. “Visual field effects on motion sickness in cars”. In: *Aviation, space, and environmental medicine* 75.9 (2004), pp. 739–748.

[8] IL Bloomington. Sept. 2016. URL: https://newsroom.statefarm.com/state-farm-releases-autonomous-vehicles-survey-results/.

[9] Michael J. Griffin and John Erdreich. “Handbook of human vibration”. In: Acoustical Society of America, 1991.

[10] Joseph Smyth et al. “Too sick to drive: How motion sickness severity impacts human performance”. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2018, pp. 1787–1793.

[11] Eric R. Muth, Alexander D. Walker, and Matthew Fiorello. “Effects of uncoupled motion on performance”. In: *Human factors* 48.3 (2006), pp. 600–607.

[12] Johnell O. Brooks et al. “Simulator sickness during driving simulation studies”. In: *Accident analysis & prevention* 42.3 (2010), pp. 788–796.

[13] Eric R. Muth and Johnell O. Brooks. “User considerations in virtual environments”. Human Factors and Ergonomics Society annual meeting, Baltimore, MD, USA. October 2002.

[14] Timothy J. Ungs. “Simulator induced syndrome: evidence for long-term aftereffects”. In: *Aviation, space, and environmental medicine* (1989).

[15] Frank Biocca. “Will simulation sickness slow down the diffusion of virtual environment technology?” In: *Presence: Teleoperators & Virtual Environments* 1.3 (1992), pp. 334–343.

[16] Joost C. F. de Winter, Peter M. van Leeuwen, and Riender Happee. “Advantages and Disadvantages of Driving Simulators: A Discussion”. In: August 2012, p. 47.

[17] Hanna Bellem et al. “Can we study autonomous driving comfort in moving-base driving simulators? A validation study”. In: *Human factors* 59.3 (2017), pp. 442–456.

[18] David R. Schreier, Christina Banks, and Johannes Mathis. “Driving simulators in the clinical assessment of fitness to drive in sleepy individuals: a systematic review”. In: *Sleep medicine reviews* 38 (2018), pp. 86–100.

[19] Rachael A. Wynne, Vanessa Beanland, and Paul M. Salmon. “Systematic review of driving simulator validation studies”. In: *Safety science* 117 (2019), pp. 138–151.

[20] Evi Blana. *Driving Simulator Validation Studies: A Literature Review*. Institute of Transport Studies University of Leeds, 1996.

[21] James J. Leonard Jr. and Walter W. Wierwille. "Human performance validation of simulators: Theory and experimental verification". In: *Proceedings of the Human Factors Society annual meeting*. Vol. 19. 4. SAGE Publications, Los Angeles, CA. 1975, pp. 446–456.

[22] Gerard J. Blaauw. "Driving experience and task demands in simulator and instrumented car: a validation study". In: *Human Factors* 24.4 (1982), pp. 473–486.

[23] Francesco Galante et al. "Validity of mental workload measures in a driving simulation environment". In: *Journal of Advanced Transportation* 2018 (2018).

[24] Ouren X. Kuiper et al. "Moving base driving simulators' potential for carsickness research". In: *Applied ergonomics* 81 (2019).

[25] William J. McNally and E. A. Stuart. *Physiology of the labyrinth reviewed in relation to seasickness and other forms of motion sickness*. American Medical Association, 1942.

[26] Jack E. Steele. "The symptomatology of motion sickness". In: *Fourth Symposium on the Role of the Vestibular Organs in Space Exploration*. 1970, pp. 89–98.

[27] James T. Reason and Joseph J. Brand. *Motion sickness*. Academic press, 1975.

[28] Michel Treisman. "Motion sickness: an evolutionary hypothesis". In: *Science* 197.4302 (1977), pp. 493–495.

[29] Gary E. Riccio and Thomas A. Stoffregen. "An ecological theory of motion sickness and postural instability". In: *Ecological psychology* 3.3 (1991), pp. 195–240.

[30] Willem Bles et al. "Motion sickness: only one provocative conflict?" In: *Brain research bulletin* 47.5 (1998), pp. 481–487.

[31] Daniel .M Merfeld et al. "A multidimensional model of the effect of gravity on the spatial orientation of the monkey." In: *Journal of vestibular research: equilibrium & orientation* (1993).

[32] Charles M. Oman. "Motion sickness: a synthesis and evaluation of the sensory conflict theory". In: *Canadian journal of physiology and pharmacology* 68.2 (1990), pp. 294–303.

[33] Jelte E. Bos, Willem Bles, and Eric L. Groen. "A theory on visually induced motion sickness". In: *Displays* 29.2 (2008), pp. 47–57.

[34] Dror Tal et al. "Artificial horizon effects on motion sickness and performance". In: *Otology & Neurotology* 33.5 (2012), pp. 878–885.

[35] L. James Smart Jr., Thomas A. Stoffregen, and Benoıt G. Bardy. "Visually induced motion sickness predicted by postural instability". In: *Human factors* 44.3 (2002), pp. 451–465.

[36] Michael E. McCauley et al. *Motion sickness incidence: Exploratory studies of habituation, pitch and roll, and the refinement of a mathematical model*. Tech. rep. Canyon Research Group Inc Goleta Ca Human Factors Research Div, 1976.

[37] John F. Golding and H. M. Markey. "Effect of frequency of horizontal linear oscillation on motion sickness and somatogravic illusion". In: *Aviation, space, and environmental medicine* 67.2 (1996), pp. 121–126.

[38] Barnaby E. Donohew and Michael J. Griffin. "Motion sickness: effect of the frequency of lateral oscillation". In: *Aviation, Space, and Environmental Medicine* 75.8 (2004), pp. 649–656.

[39] Bob Cheung and Ann Nakashima. "A review on the effects of frequency of oscillation on motion sickness". In: *Defence Research and Development Canada* (2006).

[40] Anthony Lawther and Michael J. Griffin. "Motion sickness and motion characteristics of vessels at sea". In: *Ergonomics* 31.10 (1988), pp. 1373–1394.

[41] Johan Förstberg. “Ride comfort and motion sickness in tilting trains”. PhD thesis. Institutionen för farkostteknik, 2000.

[42] Rickard Persson. “Motion sickness in tilting trains”. In: *Description and analysis of the* (2008).

[43] Giovanni Bertolini and Dominik Straumann. “Moving in a moving world: a review on vestibular motion sickness”. In: *Frontiers in neurology* 7 (2016), p. 14.

[44] Lawrence J Hettinger and Gary E Riccio. “Visually induced motion sickness in virtual environments”. In: *Presence: Teleoperators & Virtual Environments* 1.3 (1992), pp. 306–310.

[45] Behrang Keshavarz et al. “Vection and visually induced motion sickness: how are they related?” In: *Frontiers in psychology* 6 (2015), p. 472.

[46] James R. Lackner. “Motion sickness: more than nausea and vomiting”. In: *Experimental brain research* 232.8 (2014), pp. 2493–2510.

[47] Ashton Graybiel, Charles D. Wood, and Earl F. Miller (II). *Diagnostic criteria for grading the severity of acute motion sickness*. Vol. 1030. Naval Aerospace Medical Institute, Naval Aerospace Medical Center, 1968.

[48] Helen Hockx-Yu. *Invisible women, data bias in a world designed for men: Caroline Cirado Perez, Abrams Press (2019), New York*. Taylor and Francis, 2020.

[49] Aurore C. Paillard et al. “Motion sickness susceptibility in healthy subjects and vestibular patients: effects of gender, age and trait-anxiety”. In: *Journal of Vestibular Research* 23.4, 5 (2013), pp. 203–209.

[50] Sibylle Klosterhalfen et al. “Effects of ethnicity and gender on motion sickness susceptibility”. In: *Aviation, space, and environmental medicine* 76.11 (2005), pp. 1051–1057.

[51] Robert S. Kennedy et al. “Symptomatology under storm conditions in the North Atlantic in control subjects and in persons with bilateral labyrinthine defects”. In: *Acta oto-laryngologica* 66.1-6 (1968), pp. 533–540.

[52] Eike A. Schmidt et al. “An international survey on the incidence and modulating factors of carsickness”. In: *Transportation research part F: traffic psychology and behaviour* 71 (2020), pp. 76–87.

[53] Kay M. Stanney, Ronald R. Mourant, and Robert S. Kennedy. “Human factors issues in virtual environments: A review of the literature”. In: *Presence* 7.4 (1998), pp. 327–351.

[54] Jelte E. Bos. “Motion sickness, simulator sickness, and automated vehicles”. Driving Simulation and Virtual Reality Conference Europe, Antibes, France. September 2018. URL: `http://dsc2018.org/Docs/Keynotes/BOS_KEYNOTE.pdf`.

[55] Avi Shupak and Carlos R. Gordon. “Motion sickness: advances in pathogenesis, prediction, prevention, and treatment”. In: *Aviation, space, and environmental medicine* 77.12 (2006), pp. 1213–1223.

[56] Robert S. Kennedy et al. “Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness”. In: *The international journal of aviation psychology* 3.3 (1993), pp. 203–220.

[57] Peter J. Gianaros et al. “A questionnaire for the assessment of the multiple dimensions of motion sickness”. In: *Aviation, space, and environmental medicine* 72.2 (2001), p. 115.

[58] Jelte E. Bos, Scott N. MacKinnon, and Anthony Patterson. “Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view”. In: *Aviation, space, and environmental medicine* 76.12 (2005), pp. 1111–1118.

[59] Sue V. G. Cobb et al. “Virtual Reality-Induced Symptoms and Effects (VRISE)”. In: *Presence* 8.2 (1999), pp. 169–186.

[60] Marco Recenti et al. “Toward Predicting Motion Sickness Using Virtual Reality and a Moving Platform Assessing Brain, Muscles, and Heart Signals”. In: *Frontiers in Bioengineering and Biotechnology* 9 (2021), p. 132.

[61] Jiahui Li et al. “The study of driving simulator validation for physiological signal measures”. In: *Procedia-Social and Behavioral Sciences* 96 (2013), pp. 2572–2583.

[62] Young Y. Kim et al. “Characteristic changes in the physiological components of cybersickness”. In: *Psychophysiology* 42.5 (2005), pp. 616–625.

[63] Ran Liu, Eli Peli, and Alex D. Hwang. “Measuring visually induced motion sickness using wearable devices”. In: *Electronic Imaging* 2017.14 (2017), pp. 218–223.

[64] Lawrence A. Warwick-Evans et al. “Electrodermal activity as an index of motion sickness”. In: *Aviation, space, and environmental medicine* (1987).

[65] Anadil Javaid et al. “Changes in skin blood flow, respiration and blood pressure in participants reporting motion sickness during sinusoidal galvanic vestibular stimulation”. In: *Experimental physiology* 104.11 (2019), pp. 1622–1629.

[66] R. J. V. Bertin et al. “Objective measurement of simulator sickness and the role of visual-vestibular conflict situations”. In: *Driving Simulation Conference North America*. University of Central Florida Orlando, USA. 2004, pp. 280–293.

[67] Tugrul Irmak et al. “Individual motion perception parameters and motion sickness frequency sensitivity in fore-aft motion”. In: *Experimental brain research* (2021), pp. 1–19.

[68] Alireza Mazloumi Gavgani et al. “Cybersickness-related changes in brain hemodynamics: a pilot study comparing transcranial Doppler and near-infrared spectroscopy assessments during a virtual ride on a roller coaster”. In: *Physiology & Behavior* 191 (2018), pp. 56–64.

[69] Mark S. Dennison, A. Zachary Wisti, and Michael D’Zmura. “Use of physiological signals to predict cybersickness”. In: *Displays* 44 (2016), pp. 42–52.

[70] Robert S. Kennedy, Julie Drexler, and Robert C. Kennedy. “Research in visually induced motion sickness”. In: *Applied ergonomics* 41.4 (2010), pp. 494–503.

[71] Kay M. Stanney, Robert S. Kennedy, and Julie M Drexler. “Cybersickness is not simulator sickness”. In: *Proceedings of the Human Factors and Ergonomics Society annual meeting*. Vol. 41. 2. SAGE Publications Sage CA: Los Angeles, CA. 1997, pp. 1138–1142.

[72] Cyriel Diels and Peter A. Howarth. “Frequency characteristics of visually induced motion sickness”. In: *Human factors* 55.3 (2013), pp. 595–604.

[73] Alireza Mazloumi Gavgani, Deborah M Hodgson, and Eugene Nalivaiko. “Effects of visual flow direction on signs and symptoms of cybersickness”. In: *PloS one* 12.8 (2017), e0182790.

[74] A. R. Valente Pais et al. “Comparison of three motion cueing algorithms for curve driving in an urban environment”. In: *Presence* 18.3 (2009), pp. 200–221.

[75] Diane Cleij et al. “Continuous subjective rating of perceived motion incongruence during driving simulation”. In: *IEEE Transactions on Human-Machine Systems* 48.1 (2017), pp. 17–29.

[76] Cheryl Roe, Timothy Brown, and Ginger Watson. “Factors associated with simulator sickness in a high-fidelity simulator”. In: *Education* 251 (2007), 5A.

[77] Eugenia M. Kolasinski. *Simulator sickness in virtual environments*. Vol. 1027. US Army Research Institute for the Behavioral and Social Sciences, 1995.

[78] Michael Jäger et al. “Manipulations to reduce simulator-related transient adverse health effects during simulated driving”. In: *Medical & biological engineering & computing* 52.7 (2014), pp. 601–610.

[79] Yannick Forster et al. “The third dimension: Stereoscopic displaying in a fully immersive driving simulator”. In: *Proceedings of DSC 2015 Europe Driving Simulation Conference & Exhibition*. 2015, pp. 25–32.

[80] Florian Weidner et al. “Comparing VR and non-VR driving simulations: An experimental user study”. In: *2017 IEEE Virtual Reality (VR)*. IEEE. 2017, pp. 281–282.

[81] A Fleming Seay et al. “Simulator sickness and presence in a high FOV virtual environment”. In: *Proceedings IEEE Virtual Reality 2001*. IEEE. 2001, pp. 299–300.

[82] James W. Lin et al. “Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment”. In: *Proceedings ieee virtual reality 2002*. IEEE. 2002, pp. 164–171.

[83] Sherrilene Classen, Megan Bewernitz, and Orit Shechtman. “Driving simulator sickness: an evidence-based review of the literature”. In: *American journal of occupational therapy* 65.2 (2011), pp. 179–188.

[84] Robert S. Kennedy and Michael G. Lilienthal. “Measurement and control of motion sickness aftereffects from immersion in virtual reality”. In: *Proceedings of Virtual reality and medicine: The cutting edge, SIG-Advanced Applications INC, New York*. 1994, pp. 111–119.

[85] Naoko Kawano et al. “Slower adaptation to driving simulator and simulator sickness in older adults aging clinical and experimental research”. In: *Aging clinical and experimental research* 24.3 (2012), pp. 285–289.

[86] Joost de Winter and Jorrit Kuipers. “Relationships between years of licensure and driving style measured with a short simulator-based test (N= 650)”. In: *Advances in human aspects of transportation*. Springer, 2017, pp. 641–654.

[87] Arnon Rolnick and R. E. Lubow. “Why is the driver rarely motion sick? The role of controllability in motion sickness”. In: *Ergonomics* 34.7 (1991), pp. 867–879.

[88] Chih-Hui Chang et al. “Effects of physical driving experience on body movement and motion sickness during virtual driving”. In: *Aerospace medicine and human performance* 88.11 (2017), pp. 985–992.

[89] Jelte E. Bos, Suzanne A. E. Nooij, and Jan L. Souman. “(Im) possibilities of studying carsickness in a driving simulator”. In: *20th Driving Simulation & Virtual Reality Conference & Exhibition (DSC 2021 EUROPE VR)*. 2021.

[90] Joshua E. Domeyer, Nicholas D. Cassavaugh, and Richard W. Backs. “The use of adaptation to reduce simulator sickness in driving assessment and research”. In: *Accident Analysis & Prevention* 53 (2013), pp. 127–132.

[91] Adi Ronen and Nitzan Yair. “The adaptation period to a driving simulator”. In: *Transportation research part F: traffic psychology and behaviour* 18 (2013), pp. 94–106.

[92] Bruno J. Correia Grácio et al. “Perceptual scaling of visual and inertial cues”. In: *Experimental brain research* 232.2 (2014), pp. 637–646.

[93] P. Jaekl et al. “Gravity and perceptual stability during translational head movement on earth and in microgravity”. In: *Acta astronautica* 56.9-12 (2005), pp. 1033–1040.

[94] Florent Colombet et al. “Contribution of stereoscopy and motion parallax for speed perception in driving simulation”. In: *Proceedings of the Driving Simulation Conference 2021 Europe VR*. Ed. by Andras Kemeny, Jean-Rémy Chardonnet, and Florent Colombet. Driving Simulation Association. Munich, Germany, Sept. 14, 2021, pp. 91–97.

[95] ISO. “2631-1: Mechanical vibration and shock - Evaluation of human exposure to whole-body vibration”. In: *Geneva, Switzerland: ISO* (1997).

[96] J. F. Golding. “Motion sickness”. In: *Handbook of clinical neurology* 137 (2016), pp. 371–390.

[97] Alireza Mazloumi Gavgani et al. “Profiling subjective symptoms and autonomic changes associated with cybersickness”. In: *Autonomic Neuroscience* 203 (2017), pp. 41–50.

[98] Nico A. Kaptein, Jan Theeuwes, and Richard Van Der Horst. “Driving simulator validity: Some considerations”. In: *Transportation research record* 1550.1 (1996), pp. 30–36.

[99] Stuart T. Godley, Thomas J. Triggs, and Brian N. Fildes. “Driving simulator validation for speed research”. In: *Accident analysis & prevention* 34.5 (2002), pp. 589–600.

[100] Joost C. F. De Winter et al. “Driving simulator fidelity and training effectiveness”. In: *Proceedings of the 26th European annual conference on human decision making and manual control, Lyngby, Denmark*. Citeseer. 2007, pp. 20–22.

[101] Oliver Carsten and A. Hamish Jamson. “Driving simulators as research tools in traffic psychology”. In: *Handbook of traffic psychology*. Elsevier, 2011, pp. 87–96.

[102] Jeffry A. Greenberg and Mike Blommer. “Physical fidelity of driving simulators”. In: *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (2011), pp. 7–1.

[103] Thomas D Marcotte et al. “A multimodal assessment of driving performance in HIV infection”. In: *Neurology* 63.8 (2004), pp. 1417–1422.

[104] Joost C. F. De Winter and Dimitra Dodou. *Human subject research for engineers: A practical guide*. Springer, 2017.

[105] James B Schreiber et al. “Reporting structural equation modeling and confirmatory factor analysis results: A review”. In: *The Journal of educational research* 99.6 (2006), pp. 323–338.

[106] Ouren X. Kuiper, Jelte E. Bos, and Cyriel Diels. “Looking forward: In-vehicle auxiliary display positioning affects carsickness”. In: *Applied Ergonomics* 68 (2018), pp. 169–175.

[107] *Satellietdataportaal*. URL: `https://www.spaceoffice.nl/nl/satellietdataportaal/`.

[108] Andrew Gross. *New study reveals when, where and how much motorists drive*. Oct. 2020. URL: `https://newsroom.aaa.com/2015/04/new-study-reveals-much-motorists-drive/`.

[109] James F. O’Hanlon and Michael E. McCauley. *Motion sickness incidence as a function of the frequency and acceleration of vertical sinusoidal motion*. Tech. rep. Canyon Research group Inc Goleta CA Human Factors Research Division., 1973.

[110] Tugrul Irmak, Daan M. Pool, and Riender Happee. “Objective and subjective responses to motion sickness: the group and the individual”. In: *Experimental Brain Research* 239.2 (2021), pp. 515–531.

[111] Mark Turner and Michael J. Griffin. “Motion sickness in public road transport: the effect of driver, route and vehicle”. In: *Ergonomics* 42.12 (1999), pp. 1646–1664.

[112] Hanna Bellem et al. “Objective metrics of comfort: developing a driving style for highly automated vehicles”. In: *Transportation research part F: traffic psychology and behaviour* 41 (2016), pp. 45–54.

[113] Patil Ashlesh, Kishore K Deepak, and Kochhar Kanwal Preet. “Role of Prefrontal Cortex during Sudoku task: fNIRS study”. In: *Translational neuroscience* 11.1 (2020), pp. 419–427.

[114] Arben Parduzi, Joost Venrooij, and Stefanie Marker. “The Effect of Head-Mounted Displays on the Behavioural Validity of Driving Simulators”. In: *Proceedings of the Driving Simulation Conference 2020 Europe VR*. Antibes, France, 2020, pp. 125–132.

[115] Gerd Baumann et al. “Driving simulator studies for kinetosis-reducing control of active chassis systems in autonomous vehicles”. In: *Proceedings of the Driving Simulation Conference 2021 Europe VR*. Ed. by Andras Kemeny, Jean-Rémy Chardonnet, and Florent Colombet. Driving Simulation Association. Munich, Germany, Sept. 14, 2021, pp. 51–58.

[116] Herman J. Damveld et al. “Effects of motion cueing on curve driving”. In: *Proc. Driving Simul. Conf.* 2012, pp. 1–9.

[117] Konrad Stahl et al. “State of the art and simulation of motion cueing algorithms for a six degree of freedom driving simulator”. In: *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2014, pp. 537–541.

[118] Joost Venrooij et al. “Perception-based motion cueing: validation in driving simulation”. In: *DSC 2015 Europe: Driving simulation conference & exhibition*. 2015, pp. 153–161.

[119] Martijn de Mooij, Edwin de Vries, and Jelle van Doornik. “Adaptation of the Classic Cueing Algorithm for Automotive Applications”. In: *AIAA Scitech 2019 Forum*. 2019, p. 0425.

[120] Sebastian Fischer and Stephan Schwan. “Comprehending animations: Effects of spatial cueing versus temporal scaling”. In: *Learning and Instruction* 20.6 (2010), pp. 465–475.

[121] Paolo Pretto et al. “Influence of the size of the field of view on motion perception”. In: *Computers & Graphics* 33.2 (2009), pp. 139–146.

[122] Maurice Kolff et al. “Quality Comparison of Motion Cueing Algorithms for Urban Driving Simulations”. In: *Proceedings of the Driving Simulation Conference 2021 Europe VR*. Ed. by Andras Kemeny, Jean-Rémy Chardonnet, and Florent Colombet. Driving Simulation Association. Munich, Germany, Sept. 14, 2021, pp. 141–148.

[123] Eric L. Groen and Willem Bles. “How to use body tilt for the simulation of linear self motion”. In: *Journal of Vestibular Research* 14.5 (2004), pp. 375–385.

[124] Marinus Cornelis Veltena. *Movement-simulator*. Dec. 2014.

[125] Jason D. Moss and Eric R. Muth. “Characteristics of head-mounted displays and their effects on simulator sickness”. In: *Human factors* 53.3 (2011), pp. 308–319.

[126] Juffrizal Karjanto et al. “The effect of peripheral visual feedforward system in enhancing situation awareness and mitigating motion sickness in fully automated driving”. In: *Transportation research part F: traffic psychology and behaviour* 58 (2018), pp. 678–692.

[127] J. Golding. “Motion sickness susceptibility questionnaire short-form (MSSQ-Short)”. In: *Brain Res Bull* 47 (1998), pp. 507–516.

[128] General Assembly of the World Medical Association et al. “World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects”. In: *The Journal of the American College of Dentists* 81.3 (2014), pp. 14–18.

[129] Sean D. Young, Bernard D. Adelstein, and Stephen R. Ellis. “Demand characteristics in assessing motion sickness in a virtual environment: Or does taking a motion sickness questionnaire make you sick?” In: *IEEE transactions on visualization and computer graphics* 13.3 (2007), pp. 422–428.

[130] Timothy C Hain, Philip A Hanna, and Mary A Rheinberger. “Mal de debarquement”. In: *Archives of Otolaryngology–Head & Neck Surgery* 125.6 (1999), pp. 615–620.

[131] Anne Reuten, Jelte Bos, and Jeroen B Smeets. “The metrics for measuring motion sickness”. In: *Driving Simulation Conference Europe 2020*. 2020.

[132] Jelte E Bos. “Less sickness with more motion and/or mental distraction”. In: *Journal of Vestibular Research* 25.1 (2015), pp. 23–33.

[133] Nicholas A Webb and Michael J Griffin. “Eye movement, vection, and motion sickness with foveal and peripheral vision”. In: *Aviation, space, and environmental medicine* 74.6 (2003), pp. 622–625.

[134] Nidzamuddin bin Md Yusof. “Comfort in autonomous car: mitigating motion sickness by enhancing situation awareness through haptic displays”. In: (2019).

[135] AJ Benson. “Spatial disorientation–spatial aspects, common illusion and Motion sickness”. In: *Aviation medicine, Oxford, England* (1999), pp. 419–471.

[136] Jieyun Yin and Jiande DZ Chen. “Electrogastrography: methodology, validation and applications”. In: *Journal of neurogastroenterology and motility* 19.1 (2013), p. 5.

[137] Timotej Gruden et al. “Electrogastrography in Autonomous Vehicles—An Objective Method for Assessment of Motion Sickness in Simulated Driving Environments”. In: *Sensors* 21.2 (2021), p. 550.

[138] Giuseppe Riezzo, Francesco Russo, and Flavia Indrio. “Electrogastrography in adults and children: the strength, pitfalls, and clinical significance of the cutaneous recording of the gastric electrical activity”. In: *BioMed research international* 2013 (2013).

[139] Han-Chung Lien et al. “Effects of ginger on motion sickness and gastric slow-wave dysrhythmias induced by circular vection”. In: *American Journal of Physiology-Gastrointestinal and Liver Physiology* 284.3 (2003), G481–G489.

[140] Bob Cheung and Peter Vaitkus. “Perspectives of electrogastrography and motion sickness”. In: *Brain research bulletin* 47.5 (1998), pp. 421–431.

[141] Henry P. Parkman et al. “Influence of age, gender, and menstrual cycle on the normal electrogastrogram.” In: *American Journal of Gastroenterology (Springer Nature)* 91.1 (1996).

[142] Hrair P Simonian et al. “Multichannel electrogastrography (EGG) in normal subjects: a multicenter study”. In: *Digestive diseases and sciences* 49.4 (2004), pp. 594–601.

[143] Nenad B Popović, Nadica Miljković, and Mirjana B Popović. “Simple gastric motility assessment method with a single-channel electrogastrogram”. In: *Biomedical Engineering/Biomedizinische Technik* 64.2 (2019), pp. 177–185.

[144] Millaray Curilem et al. “Recognition of Nausea Patterns by Multichannel Electrogastrography”. In: *Pattern Recognition Techniques Applied to Biomedical Problems* (2020), pp. 91–108.

[145] Osamu Tokumaru et al. “Vector analysis of electrogastrography during motion sickness”. In: *Digestive diseases and sciences* 48.3 (2003), pp. 498–507.

[146] Arvind Gautam et al. “A data driven empirical iterative algorithm for GSR signal pre-processing”. In: *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE. 2018, pp. 1162–1166.

[147] Dong P. Jang et al. “Analysis of physiological response to two virtual environments: Driving and flying simulation”. In: *CyberPsychology & Behavior* 5.1 (2002), pp. 11–18.

[148] Byung-Chan Min et al. “Psychophysiological evaluation of simulator sickness evoked by a graphic simulator”. In: *Applied ergonomics* 35.6 (2004), pp. 549–556.

[149] Afsaneh Koohestani et al. “A knowledge discovery in motion sickness: a comprehensive literature review”. In: *IEEE access* 7 (2019), pp. 85755–85770.

[150] Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.

[151] Rajiv Ranjan Singh, Sailesh Conjeti, and Rahul Banerjee. “A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals”. In: *Biomedical Signal Processing and Control* 8.6 (2013), pp. 740–754.

[152] J. F. Golding and J. R. R. Stott. “Comparison of the effects of a selective muscarinic receptor antagonist and hyoscine (scopolamine) on motion sickness, skin conductance and heart rate”. In: *British journal of clinical pharmacology* 43.6 (1997), pp. 633–637.

[153] Elena Smets et al. “Comparison of machine learning techniques for psychophysiological stress detection”. In: *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer. 2015, pp. 13–22.

[154] Hongwei Wan, Senqi Hu, and Jiangyue Wang. “Correlation of phasic and tonic skin-conductance responses with severity of motion sickness induced by viewing an optokinetic rotating drum”. In: *Perceptual and motor skills* 97.3_suppl (2003), pp. 1051–1057.

[155] Jacob O. Wobbrock et al. “The aligned rank transform for nonparametric factorial analyses using only anova procedures”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2011, pp. 143–146.

[156] Lisa A. Elkin et al. “An Aligned Rank Transform Procedure for Multifactor Contrast Tests”. In: *arXiv preprint arXiv:2102.11824* (2021).

[157] Ramon C. Littell, P. R. Henry, and Clarence B. Ammerman. “Statistical analysis of repeated measures data using SAS procedures”. In: *Journal of animal science* 76.4 (1998), pp. 1216–1231.

[158] Miguel Angel Guevara and Maria Corsi-Cabrera. “EEG coherence or EEG correlation?” In: *International Journal of Psychophysiology* 23.3 (1996), pp. 145–153.

[159] Philippe Perrin et al. “Motion sickness in rally car co-drivers”. In: *Aviation, space, and environmental medicine* 84.5 (2013), pp. 473–477.

[160] Eric L. Groen and Jelte E. Bos. “Simulator sickness depends on frequency of the simulator motion mismatch: An observation”. In: *Presence* 17.6 (2008), pp. 584–593.

[161] Hirohisa Mizuta et al. “Changes in the heart rate variability and electrogastrogram due to the motion sickness”. In: *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 4. IEEE. 1996, pp. 1636–1637.

[162] Robert S. Kellogg, Robert S. Kennedy, and Ashton Graybiel. *Motion sickness symptomatology of labyrinthine defective and normal subjects during zero gravity maneuvers*. Tech. rep. Aerospace Medical Research Labs, Wright-Patterson, AFB Ohio, 1964.

[163] Björn Kufver and Johan Förstberg. *A net dose model for development of nausea*. Statens väg-och transportforskningsinstitut., VTI särtryck 330, 1999.

[164] Behrang Keshavarz and John F Golding. “Motion sickness: current concepts and management.” In: *Current opinion in neurology* (2021).

[165] Jaap Swanenburg et al. “Exergaming in a moving virtual world to train vestibular functions and gait; a proof-of-Concept-Study with older adults”. In: *Frontiers in physiology* 9 (2018), p. 988.

[166] Jelte E Bos et al. “The effect of internal and external fields of view on visually induced motion sickness”. In: *Applied ergonomics* 41.4 (2010), pp. 516–521.

[167] Mehmet Dagdelen et al. “Model-based predictive motion cueing strategy for vehicle driving simulators”. In: *Control Engineering Practice* 17.9 (2009), pp. 995–1003.

[168] Diane Cleij et al. “Optimizing an Optimization-Based MCA using Perceived Motion Incongruence Models”. In: *Proceedings of the 19th Driving Simulation and Virtual Reality Conference, Antibes, France*. 2020, pp. 9–11.

[169] *Driving simulator motion cueing optimization*. Dec. 2019. URL: `https://cs.lr.tudelft.nl/cybernetics/projects/driving-simulator-cueing/`.

[170] Th Probst, A Straube, and W Bles. “Differential effects of ambivalent visual-vestibular-somatosensory stimulation on the perception of self-motion”. In: *Behavioural brain research* 16.1 (1985), pp. 71–79.

[171] Willibald Brems et al. “Frequency response and latency analysis of a driving simulator for chassis development and vehicle dynamics evaluation”. In: *Driving simulation conference and exhibition Europe*. 2015.

[172] Joost C. F. de Winter. *Lectures Human-Machine Systems (ME41080)*. Nov. 2020. URL: `https://brightspace.tudelft.nl/d2l/le/content/279399/viewContent/2038952/View`.

[173] Thomas A. Stoffregen et al. “Postural instability and motion sickness in a fixed-base flight simulator”. In: *Human Factors* 42.3 (2000), pp. 458–469.

[174] J. D. Lee. “Simulator fidelity: How low can you go”. In: *48th Annual Meeting of the Human Factors and Ergonomics Society*. 2004.

[175] Björn Blissing and Fredrik Bruzelius. “Exploring the suitability of virtual reality for driving simulation”. In: *Driving Simulation Conference 2018*. 2018, pp. 163–166.

[176] W. D. Ledegang et al. “Simulation of motion parallax by correcting the out-the-window scenery for head position”. In: (2015).

[177] James McGuinness, J. H. Bouwman, and Jim M. Forbes. *Simulator Sickness Occurrences in the 2E6 Air Combat Maneuvering Simulator (ACMS)*. Tech. rep. Person-System Integration INC Alexandria, VA, 1981.

[178] J. W. Miller and J. E. Goodson. “Motion sickness in a helicopter simulator”. In: *Aerospace medicine* 31 (1960), pp. 204–212.

[179] Martijn L. van Emmerik, Sjoerd C. de Vries, and Jelte E. Bos. “Internal and external fields of view affect cybersickness”. In: *Displays* 32.4 (2011), pp. 169–174.

[180] Ksander N. De Winkel, Mikhail Katliar, and Heinrich H. Bülthoff. “Causal inference in multisensory heading estimation”. In: *PloS one* 12.1 (2017), e0169676.

[181] Eunhee Chang, Hyun Taek Kim, and Byounghyun Yoo. “Virtual reality sickness: a review of causes and measurements”. In: *International Journal of Human–Computer Interaction* 36.17 (2020), pp. 1658–1682.

[182] Lawrence H. Frank, John G. Casali, and Walter W. Wierwille. “Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator”. In: *Human factors* 30.2 (1988), pp. 201–217.

[183] Thomas A. Stoffregen and L. James Smart Jr. “Postural instability precedes motion sickness”. In: *Brain research bulletin* 47.5 (1998), pp. 437–448.

[184] George D. Park et al. “Simulator sickness scores according to symptom susceptibility, age, and gender for an older driver assessment study”. In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 26. SAGE Publications Sage CA: Los Angeles, CA. 2006, pp. 2702–2706.

[185] Hannsjoerg Schmieder, Katja Nagel, and Hans-Peter Schoener. “Enhancing a driving simulator with a 3D-stereo projection system”. In: *Proceedings of the driving simulator conference*. 2017.

[186] Barıs Aykent et al. “Study of the influence of different washout algorithms on simulator sickness for a driving simulation task”. In: *World Conference on Innovative Virtual Reality*. Vol. 44328. 2011, pp. 331–341.

[187] Malte Klüver et al. “The behavioral validity of dual-task driving performance in fixed and moving base driving simulators”. In: *Transportation research part F: traffic psychology and behaviour* 37 (2016), pp. 78–96.

[188] Guillaume Lucas et al. “A simulation sickness study on a driving simulator equipped with a vibration platform”. In: *Transportation research part F: traffic psychology and behaviour* 68 (2020), pp. 15–22.

# A

# Fidelity and simulator sickness: An uncanny valley

Driving simulators are becoming an increasingly important tool to develop (automated) vehicle functionalities, for rehabilitation, driver training, entertainment and to test human factors responses [171]. Various advantages accompany the use of a simulator as compared to a real-road drive, some of which are controllability, reproducibility, efficiency of data collection and safety for operators [16]. An often occurring limitation is simulator discomfort, specifically Simulator Sickness (SS) [16, 53]. The most prominent theory on the origin of simulator (motion) sickness explains a conflict between sensed and expected neural inputs on the basis of previous experience [27]. The possible cause of SS is twofold and can be subdivided into sickness as a result of motion which can also be encountered during actual driving (Simulated Carsickness, SCS), and sickness as a result of ambiguities within or between the motion or visualization modality of the simulator (Simulator Induced Sickness, SIS). Several studies showed a relation between sickness and specific simulator technology aspects such as choice of Motion Cueing Algorithm (MCA) [17] or the Field of View (FoV), intensity, brightness and contrast of the visualization system [77, 78, 81, 82].

Half a century ago, all simulators were fixed-based, hence unable to simulate actual motion. Simulators nowadays come in all sorts of forms and sizes, making it difficult to compare outcomes of different simulators. Although subjective, simulators are typically classified by means of their fidelity. *Fidelity* refers to the extent to which (specific technology aspects of) simulators realistically emulate driving (or flying or sailing) in the real-world [19]. For fidelity, sometimes the word *realism* or *correspondence* is used. Fidelity is often subdivided into physical fidelity, perceptual fidelity and behavioural fidelity [172]. *Physical fidelity* refers to the objective degree of similarity between the simulator and the real vehicle, measured using motion cues in both. *Perceptual fidelity* is the degree of similarity between subjective perception of the simulation by the operator, with respect to that of the real vehicle and *behavioural fidelity* refers to the degree of similarity of a person his or her control behaviour in the real world versus the simulator [172]. Currently, no unified and consequent approach to classify simulator fidelity is acknowledged. Studies generally use "low", "medium" and "high" fidelity without a fixed classification. The study by Kaptein et al. (1996) [98] was the first to categorize simulators based on their fidelity elements, such as the inclusion of a motion base and vehicular controls. Wynne et al. (2019) [19] later created a similar ranking table (shown in table A.2), with the possibility to rank also newer simulators on the basis of motion, visual and physical fidelity factors. As described above, several fidelity aspects are in some way related to SS. In this study, we aim to find the relation between SS and simulator fidelity as a whole, to make estimations about outcomes of future simulator experiments.

Several recent studies [16, 35, 100, 173–176] address a (counterintuitive) positive correlation between improvements in simulation fidelity and the increase in likelihood of sickness. Their statements logically impede simulator inno-

vation. Nonetheless, this relation is very often based on two findings: the results presented by McGuinness et al. (1981) [177] and those by Miller & Goodson (1960) [178]. We propose that the aforementioned statements have to be nuanced as conclusions from [177, 178] are drawn too quick. Not only are the simulators used in these studies all fixed-base simulators, also these studies are over 40 years old. Due to endless new and improved cues, "high" fidelity simulators classified back then would be "low" fidelity simulators nowadays. Moreover, Miller & Goodson themselves suggest that "SS could occur because of the absence of real motion" [178]. The most widely evaluated fidelity aspect in relation to SS, must be the display/projection Field of View (FoV). Studies using stationary simulators showed that the larger the FOV, the more convincing a possible conflict may be, resulting in significant increases of SS [81, 82, 179]. For very low fidelity fixed-base simulators, however, it is possible that people do not experience any sickness. This occurrence can be justified by the Causal Inference Principle, which suggests that multisensory integration of visual and vestibular information in the brain does not occur when visual motion is not realistic enough [180].

Though never studied in the context of (moving base) driving simulators, Chang et al. (2020) [181] showed a relation between Virtual Reality (VR) fidelity and (visually induced) motion sickness. Participants showed decreased discomfort in high fidelity virtual environments if the VR provided information in multiple sensory dimensions (right of figure A.1). Multisensory here indicates for example the addition of vestibular, auditory or proprioceptive information, to the visual motion. If the VR was a unimodal system (left of figure A.1), this effect was shown reversed. Their results are presented in figure A.1. In this study, we will present a comparable meta-analysis for SS as a result of fidelity aspects.

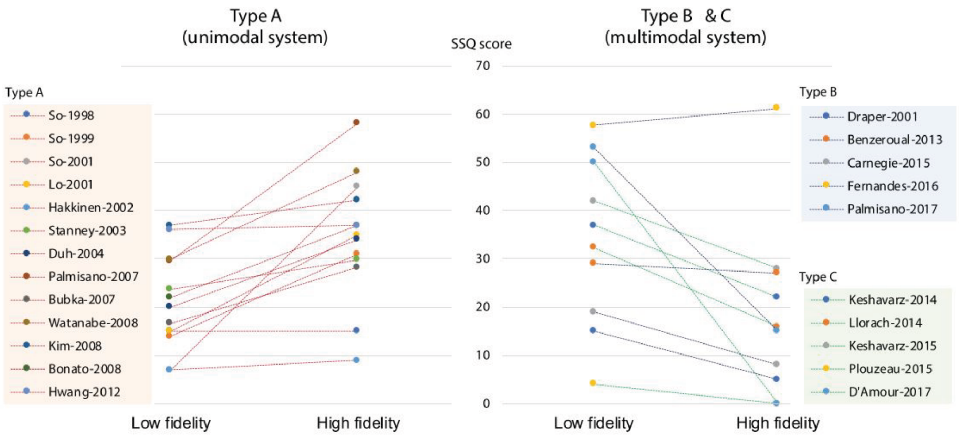


Figure A.1: The relationship between the level of VR fidelity and VR sickness depending on the multimodality of the VR scenario, retrieved from [181].

## A.1 Method

Existing (empirical) studies comparing (at least) two simulator conditions that vary some aspect of fidelity, will be used to obtain the relation between SS and fidelity. Since *fidelity* addresses such a large scope of research, literature used in this analysis is obtained from the extensive simulator review on SS by Classen et al. (2011) [83]. This study assembled, critically appraised, and synthesized the results carefully. Literature from subsequent Driving Simulator Conference proceedings until 2021 is added to also provide more recent findings.

The rating table by Wynne et al. (2019) [19] is used to calculate the total level of fidelity for each simulator setup or condition. On the basis of this table, the fidelity level of a simulator can be placed on a horizontal axis. Leftmost on this axis would be the lowest fidelity "simulator", existing of for example a single display and a seat. Though trivial, this setup does probably not provoke any sickness. Also a typical fixed-base simulator in modern applications is classified

as low fidelity [19], as the complete motion aspect is missing. Nevertheless, SS has been reported in these fixed-base simulators [182, 183]. At the rightmost end of the axis lies perfect fidelity, creating only the sickness that would also occur in real-road driving (SCS). On the vertical axis, SS is placed.

| Fidelity measure | Score | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Visual | Single PC screen | Single projector or PC screen >25 in. | <180° FOV with multiple screens | 180–270° FOV; multiple PC screens | >270° FOV; projector screens |
| Motion | No motion base | | Lower degrees of motion (<6 or partial plate) | | Motion-base on Full motion plate |
| Physical | Computer based-simulation using keyboard or joystick | Computer-based simulation with steering wheel | "Arcade" vehicle – car seat with steering wheel and pedals | Vehicular controls; no or incomplete cab | Full vehicular cab and controls |

Figure A.2: Rating system for simulator fidelity, retrieved from Wynne et al. (2019) [19].

Sickness severity between the literature findings cannot be compared due to variations in perceived motion or visual cues, individual differences, different time courses and different sickness rating methods. However, within-literature trends (if SS increases, decreases or if no difference can be found between the varying fidelity conditions), can be compared. Scores from the rating table are added, whereby a score of 3 is the minimum and 15 the maximum score that can be obtained.

## A.2 Results

McGuinness (1981) [177] compared two simulators, the 2E6 ACMS and the SAAC. Both have their motion-bases turned off, a FoV of 360° and both obtain a total score of 10. Display quality of the 2E6 is higher, as this simulator is newer. Results report an increase in SS. Seay et al. (2001) [81] compared a FoV of 60° with a FoV of 180° using the same fixed-base simulator. The first condition receives a score of 3, the second a score of 6. The results report an increase in SS. Lin et al. (2002) [82] reported increased sickness for increasing FoV (60°, 100°, 140°, 180°), of which the first condition is scored with 9, whereas the last is scored with 10. Increased SS with increasing FoV is reported. Park et al. (2006) [184] compared three simulator conditions. One condition exists of a single screen, the second uses three screens and the last is a full vehicular cab. Fidelity scores obtained are respectively 4, 6 and 10. Highest nausea scores were found for the single screen and lowest scores for the 3-monitor setup. Jäger et al. (2014) [78] showed that increased optical flow resulted in increased sickness levels. Both setups score a 9, as the table has no optical flow included. Forster et al. (2015) [79] found increased sickness symptoms for the stereoscopic 3D visualization setup as compared to a 2D setup. Both simulator setups scored a 10. Weidner et al. (2017) [80] showed that for three fixed base simulator setups accompanied by 2D visualization, 3D stereoscopic visualization and a Head-Mounted Display (HMD), the HMD condition obtained highest sickness scores. Respectively, these setups score 6, 6 and 9 points. Schmieder et al. (2017) [185] found no significant differences in sickness between two fixed-base conditions using 2D and 3D visualization

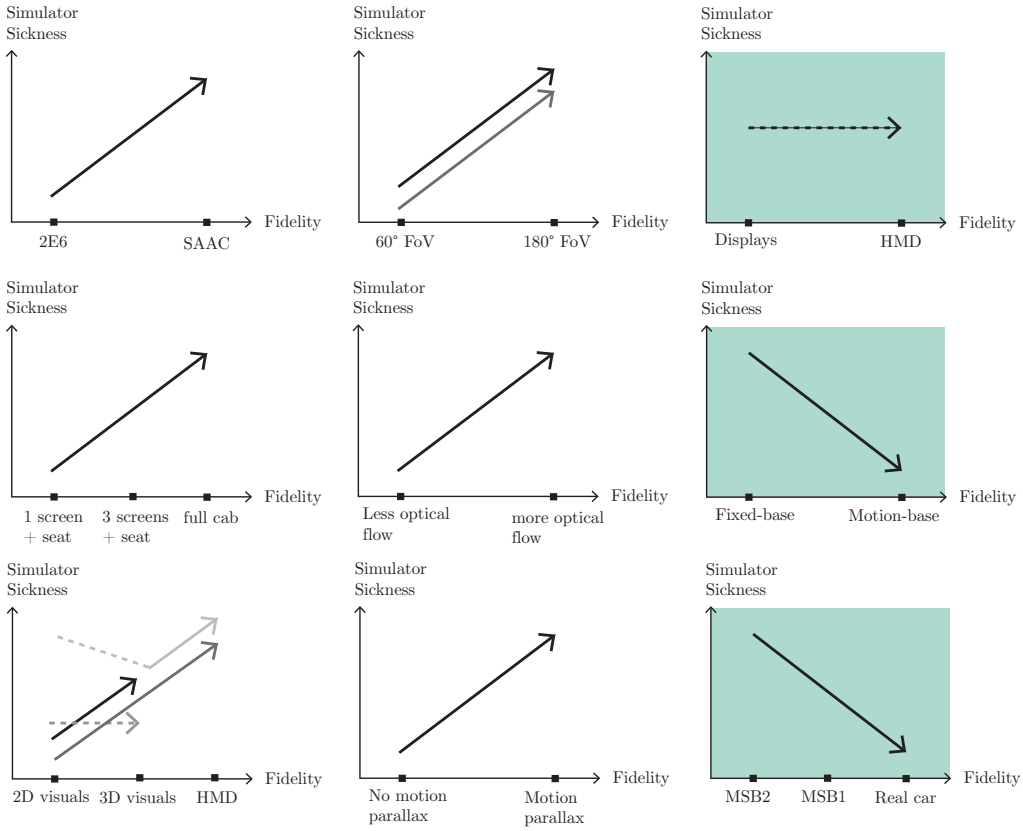


Figure A.3: Visualization of literature findings that present some relation with SS (y-axis) and fidelity (x-axis). Green graphs represent the presence of a motion-base. When multiple studies concluded similar results, this is displayed in the form of multiple arrows. Dotted lines represent unexpected findings.

setup. According to the rating table, both conditions scored a 10. Ledegang et al. (2015) [176] compared two fixed base simulator setups with and without motion parallax. Both conditions scored an 8 according to the rating table. However, motion parallax obtained higher SS scores.

Parduzi et al. (2020) [114] obtained a non-significant difference between a motion-base simulator setup with three display screens (score of 12) and one accompanied by a Head-Mounted Display system (score of 14). Aykent et al. (2014) [186] compares a static and a driving simulator and found SS to decrease significantly for the dynamic case. The static simulator setup receives a score of 10, whereas the dynamic setup receives a score of 14. Also Klüver et al. (2016) [187] compared both static and dynamic driving simulators. Their 6 setups range from 8, 9, and 10 for fixed-base simulators, to 15, 15 and a real car (perfect fidelity). For the motion-base simulators, they found that sickness decreased with increasing fidelity, for the fixed-base simulators the SSQ scores are very similar.

## A.3 Discussion

From the results, a few observations can suggest a general relationship between SS and fidelity: Firstly, it is possible to become motion sick (SS) in a fixed-base simulator. Secondly, SIS (and therefore SS) generally increases with increasing fidelity when a simulator is fixed-base. This finding is contradictory with previous literature, as mentioned in the in-

troduction. Thirdly, SS decreases with increasing fidelity for multimodal systems (e.g. motion-base simulator). These findings combined, result in an increase of (Simulator Induced) sickness for unimodal simulators and a decrease of sickness for multimodal systems, when fidelity is increased. The need for multisensory information, here in the form of motion, is furthermore supported by Pinto et al. (2008) and Lucas et al. (2020) [188]. Both studies suggest providing proprioceptive information as a need to reduce the conflict (to reduce sickness) caused by the sole availability of visual information.

This relation addresses one half of the equation: SIS only. Accompanying the conclusion, are some findings of previous studies that (obliviously) address a relationship between SCS and fidelity. Firstly, from the study by Bos (2021) [89], who states that it is impossible to elicit Simulated Carsickness in a fixed-base simulator, we infer that SCS is zero in these stationary simulators. The same study addresses the second argument of the relationship. Everyday medium-fidelity motion-base simulators (such as with a limited motion envelope) are unable to simulate motion 1:1, which is why we suggest SCS will be lower than actual MS. Higher (or perfect) fidelity motion-base simulators can approach real road MS better as the level of SCS can now be higher. Lastly, by definition, at perfect fidelity all of the sickness obtained in the simulator, would also be obtained in a real car assuming identical scenarios. SIS is then zero (there are no more latencies, visual errors, poor visual quality etc.) and SCS (SS) has approached MS perfectly. These two relations are combined in figure A.4.

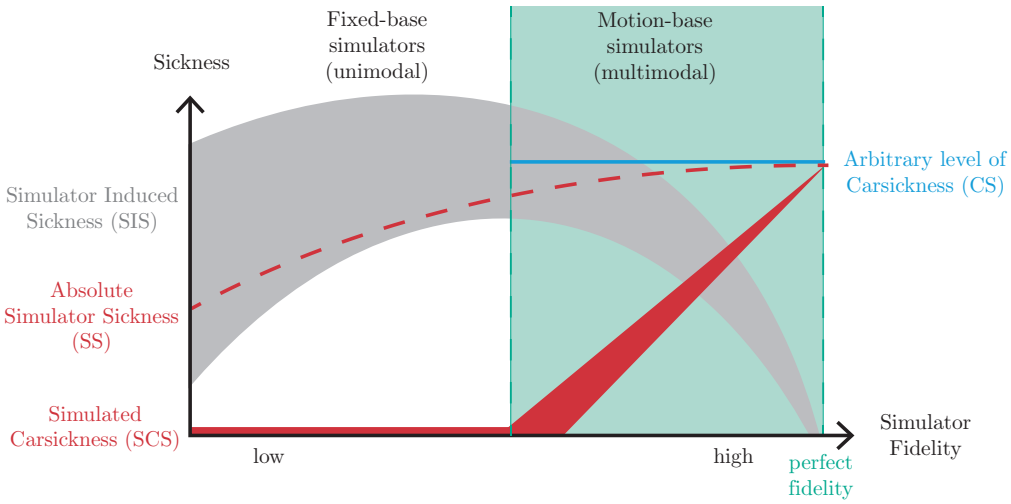


Figure A.4: Visualization of both SS components versus simulator fidelity.

We can conclude that a similar effect occurs for uni- and multimodal driving simulators, as occurs in VR analyzed by Chang et al. (2020) [181]. Therefore it can be assumed that a similar relation holds for other artificial immersive tools, such as flight simulators or boat simulators. Moreover, this relation addresses the urgency to pursue higher fidelity for simulators.

In summary, once fidelity of unimodal systems such as a fixed-base simulators increases, perceived sickness also increases. Possible neural conflicts may be more convincing with higher fidelity, resulting in significant increases of SS. However, if a system is able to provide more sensory cues than only the visual cues (e.g. a moving base simulator), Simulator Induced Sickness as a result of ambiguities in the visual or motion modality reduces, and Simulated Carsickness approaches actual Carsickness. This relation can be used to roughly estimate outcomes of simulator (sickness) studies on the basis of the fidelity of the setup.

# B

# Misery Scale rating and Sudoku

By means of Table B.1, participants were asked to rate their discomfort level every 30 seconds during the experiment. Rating was done in writing on a designated form that accompanied an informed consent, a structured briefing, the Sudoku that subjects had to solve and the MSAQ form.

| Misery Scale (MISC) rating | | |
|---|---|---|
| **Symptoms** | **Severity** | **MISC** |
| No problems | | 0 |
| Some discomfort, but no specific symptoms | | 1 |
| Dizziness, cold/warm, headache, stomach/throat awareness, sweating, blurred vision, yawning, burping, tiredness, salivation, ... but no nausea | Vague | 2 |
| | Little | 3 |
| | Rather | 4 |
| | Severe | 5 |
| Nausea | Little | 6 |
| | Rather | 7 |
| | Severe | 8 |
| | Retching | 9 |
| Vomiting | | 10 |

Table B.1: The MISC rating scale, retrieved from [58].

Participant _____ Date _____ of Scenario _____

## Sudoku and sickness rating table

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 6 | | | 8 | 7 | | | | |
| | | 8 | | | 9 | 7 | | |
| | | | | | 1 | | 4 | |
| | | | | | | | | |
| | | | 3 | | 5 | | 9 | |
| | 5 | | | 9 | | | 8 | 2 |
| | 4 | | | | 2 | 5 | 6 | |
| | | 1 | | 6 | 4 | | | |
| | | 6 | 1 | | | 9 | 3 | |

Aantal behaalde punten _____

| | MISC rating (0-7) to be written down every 30 seconds | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Begin-end | | | | | | | | | | | | | | | | | | |
| End-begin | | | | | | | | | | | | | | | | | | |
| Begin-end | | | | | | | | | | | | | | | | | | |
| End-begin | | | | | | | | | | | | | | | | | | |
| Begin-end | | | | | | | | | | | | | | | | | | |
| End-begin | | | | | | | | | | | | | | | | | | |
| Rest | | | | | | | | | | | | | | | | | | |

8/10

Figure B.1: A4 document of Sudoku puzzle and MISC recording table to be printed out for the experiment, together with the MISC rating table B.1.

# C

# Motion Sickness Assessment Questionnaire

By means of table C.1, participants were asked to rate their experience after immersion in both scenarios. Rating was done in writing on a designated form that accompanied an informed consent, a structured briefing, the MISC table and the Sudoku that subjects had to solve.

APPENDIX A

MOTION SICKNESS ASSESSMENT QUESTIONNAIRE (MSAQ).

***Instructions.*** **Using the scale below, please rate how accurately the following statements describe your experience**

| Not at all | | | | | | | | Severely |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 1. I felt sick to my stomach (G) | | | | | | | | | | 9. I felt disoriented (Q |
| 2. I felt faint-like (C) | | | | | | | | | | 10. I felt tired/fatigued (S) |
| 3. I felt annoyed/irritated (S) | | | | | | | | | | 11. I felt nauseated (G) |
| 4. I felt sweaty (P) | | | | | | | | | | 12. I felt hot/warm (P) |
| 5. I felt queasy (G) | | | | | | | | | | 13. I felt dizzy (C) |
| 6. I felt lightheaded (C) | | | | | | | | | | 14. I felt like I was spinning (C) |
| 7. I felt drowsy (S) | | | | | | | | | | 15. I felt as if I may vomit (G) |
| 8. I felt clammy/cold sweat (P) | | | | | | | | | | 16. I felt uneasy (S) |

*Note.* G; Gastrointestinal; C: Central; P: Peripheral; SR; Sopite-related.

The overall motion sickness score is obtained by calculating the percentage of total points scored: (sum of points from all items/144) × 100. Subscale scores are obtained by calculating the percent of points scored within each factor: (sum of gastrointestinal items/36) × 100; (sum of central items/45) × 100; (sum of peripheral items/27) × 100; (sum of sopite-related items/36) × 100.

Figure C.1: A4 document of the MSAQ rating table to be printed out for the experiment, retrieved from [57].


TU Delft
cruden