Enhanced Perception of Risk in a Driving Simulator

Master Thesis of Andrea L. Hauslbauer

University of Twente

1st Supervisor:

Prof. dr. ing. W.B. Verwey

Faculty of Behavioural, Management and Social Sciences (BMS), Cognitive Psychology and Ergonomics (CPE) (CPE)

2nd Supervisor.

Francesco Walker (PhD candidate)

Faculty of Engineering Technology (ET), Centre for Transport Studies (VVR)

Abstract

A crucial factor influencing the validity of VR is sense of presence. The aim of this study is, firstly, to increase sense of presence of simulator driving by introducing an anxiety-inducing risk factor. Secondly, to add clarification to the partly inconclusive relationship between sense of presence, anxiety and physiological markers. The study consisted of two groups. The control group drove in the simulator normally. The threat group expected an electric shock in case of collision. Self-report ratings (on sense of presence, anxiety and task load), behavioral (speed, steering reversal rate, brake pedal use) and physiological measurements (EDA, HR) were taken. We expected an increase in the sense of presence, arousal, steering reversals, brake pedal use, task load and anxiety ratings, and a decrease of speed in the threat group. Analyses revealed partial group differences in speed and steering reversal rate. Sense of presence was identical and notably high in both groups. This implies that either a more aversive stimulus might be needed to induce anxiety and increase sense of presence, or that it can hardly be increased at all in highly immersive VR systems.

Word count: 184

Enhanced Perception of Risk in a Driving Simulator

Due to their numerous advantages over on-road driving, driving simulators are a popular, useful virtual reality (VR) tool for a variety of purposes. These include, but are not limited to, driver training (Casutt, Theill, Martin, Keller, & Jäncke, 2014), road design (Keith et al., 2005), medical research (Helland et al., 2013) and human factors in engineering research (e.g. Slater, Usoh, & Steed, 1994; Ahlström, Anund, Fors, & Åkerstedt, 2018).

Driving simulators provide high controllability with systematic variation of variables (Godley, Triggs, & Fildes, 2002), high reliability and simple driver monitoring and recording (Nilsson, 1993). To be truly useful, VR tools like simulators need to offer adequate validity to allow inferring conclusions about the real world. To do so, the VR experience must be sufficiently authentic and lifelike (Alsina-Jurnet & Gutierrez-Maldonado, 2010). Essential for this is the sense of being there (Slater, Usoh & Steed, 1994), commonly referred to as "sense of presence". This is a subjective aspect quantified by the individual user. It has been found to be influenced by immersion and the extent of emotion presented in the virtual environment (Baños et al., 2004). Immersion refers to the objective technology that enables the VR experience, e.g. the extent of the visual field (Slater, Usoh & Steed, 1994). For example, a mobile phone is lower in immersion than an IMAX cinema. Somewhat independent of technical developments, the affective content of a virtual environment can be manipulated and is succumbed to subjectivity.

Anxiety-inducing environments have been found to enhance sense of presence. For example, Alsina-Jurnet and Gutierrez-Maldonado (2010) exposed groups of students to a virtual exam situation designed to induce anxiety, as well as to neutral virtual environments (e.g. an ordinary living room). They found that in the anxiety environment, students felt greater sense of presence than in the neutral environment, and concluded that affective content increases sense of presence in VR. The review of Diemer, Alpers, Peperkorn, Shiban and Mühlberger (2015) lists more experiments that come to the same conclusion. However, they also state that "research has not yet been able to clarify the relationship between presence and emotional experience in VR" (p. 4). Furthermore, the relationship between sense of presence and its physiological markers is debatable (Schuemie, van der Straaten, Krijn, & van der Mast., 2001; Felnhofer et al., 2014). Regarding the emotional aspect in terms of anxiety, there is a striking difference between on-road and simulator driving: The lack of risk in the latter (Espié, Gauriat, & Duraz, 2005). If the relation between sense of presence and anxiety holds, inducing anxiety might offer room for improvement of sense of presence, and therefore the validity of simulator driving (*Fig. 1*). Furthermore, insights might be gained into the controversial relationships between sense of presence, physiological and emotional markers.

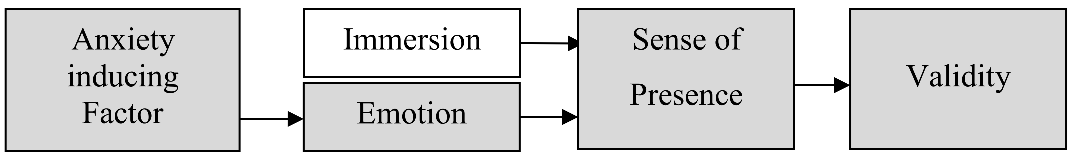

Figure 1. Simplified model to illustrate the relation from anxiety to validity by means of increasing sense of presence. For a more detailed model see Bystrom, Barfield & Hendrix (1999).

The chosen aversive stimulus to serve as a risk factor in this experiment was an electric shock. Electric shocks are pain stimuli. They are ideal for laboratory purposes due to their high controllability, measurability, and reproducibility. Furthermore, they can be easily applied anywhere on the body with electrodes, are not too physically constraining, and do therefore not interfere with other tasks, like simulated driving (Sang, Max & Gracely, 2003; Tursky, 1974).

In this experiment, anxiety was induced by the prospect of an electric shock, experienced in case of collision in the simulator. The intensity of the shock was determined by means of a 6-step calibration procedure before the driving task (see Appendix). Notably, participants were not expected to crash, and even in that case, no shock was actually administered. Our goal was to induce anxiety, not pain. The sole expectance of an aversive event has been shown as sufficient to induce anxiety (Phelps et al., 2001). Additionally, not administering electric shocks during driving avoided a possible interference with the physiological measurements.

Sense of presence and anxiety scores were collected with questionnaires. To gain more insight into the psychological experience, a questionnaire assessing task load was administered as well. All self-report scores were expected to be higher in the threat group.

For physiological measures, we selected heart rate (HR) and electrodermal activity (EDA) in the form of skin conductance. We expected both to be higher in the threat group due to increased arousal. The literature (Bassett, Marshall & Spillane, 1987; Jorna, 1993) is controversial about the link between HR and fear inducing experimental situations. However, the majority of studies reviewed are in favor of increased heart rate in anxiety inducing events (Lande, 1982; Vrijkotte, van Doornen, & de Geus, 2000; Hjortskov et al., 2004). We looked at HR data in an explorative fashion, to see if our data can help to clarify the debate. Skin conductance is described as a straightforward way of measuring any activity of the sympathetic nervous system, and hence arousal (e.g. Stern, Ray, & Quigley, 2001; Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000). The study from Meehan (2000) found a link between self-reported presence in VR and skin conductance. Participants of that study showed a simultaneous increase in sense of presence ratings and skin conductance level when they moved from a virtual training room to a virtual experimental room with a pit in the floor that one could fall into. As our goal was to induce anxiety, and therefore arousal, we expected to see increased skin conductance in the threat group compared to the control group.

Concerning the behavioral measures, we expected the threat group to drive more carefully. This would reduce the probability of a collision and therefore of experiencing an electric shock. In the data, we expected this to show in decreased speed, increased steering reversal rate and increased use of the brake pedal. For speed, it is undisputed in the literature that careful driving is associated with reduced speed (Schmidt-Daffy, 2013). A steering reversal is defined as a change in movement of the steering wheel from a clockwise to a counterclockwise direction, and vice versa (Green, 2013). Studies found that fatigued drivers tend to decrease steering movement (Feng, Zhang & Cheng, 2009; Krajewski, Sommer, Trutschel, Edwards, & Golz, 2009). Increasing workload, by adding additional activities to the driving task, has been found to correlate highly with increased corrective steering maneuvers (Nakayama, Futami, Nakamura, &

Boer, 1999). Faure, Lobjois and Benguigui (2016) also found that steering reversal rate increased simultaneously with workload. In their study, they added secondary tasks, such as calculation and reaction time tasks, to the driving task. Correspondingly, as we expected the threat group to be more alert, as well as to experience increased workload, we also expected them to make more steering reversals than the control group. Furthermore, we expected the threat group to make more use of the brake pedal. Additionally, we expected less, if at all, collisions in the threat group compared to the control group.

In conclusion, the purpose of this study was to tackle a major difference between on-road driving and simulator driving: The lacking experience of risk in the latter. This was done by adding affective content (anxiety) to the simulator experience, which was expected to increase sense of presence ratings. Furthermore, there is, still, a noted "lack of conclusive research on the relationship between presence (…) and emotional responses such as fear" (Schuemie et al., 2001, p. 199; Felnhofer et al., 2014). Through this study, we aimed to clarify the link between anxiety and sense of presence.

Our study was a between-subject design, where the participants in the threat group were at risk of experiencing an aversive electric shock while driving. The collected data was compared to the data of the control group, where there was no risk of receiving electric shocks. In this way, we tried to answer the following research question: Does the perception of enhanced risk in a driving simulator increase the sense of presence in the user?

2. Method

2.1 Participants

Thirty-two healthy students (23 males and 9 females, split equally across groups) participated in this study. Their age ranged from 20 to 27 years, (M = 23.8, SD = 4.2). The participants had to be in possession of a drivers license and confirm sufficient driving experience with at least two years of driving. They reported normal or corrected-to-normal vision and no previous experience in a driving simulator. We recruited participants through the University's online recruitment service. Participants received monetary compensation (6 € per participant) or credits. Participants had no prior knowledge of the research question, but were told that the goal of the experiment was to test the realism of the simulator. Exclusion criteria were set up to ensure participants' health and safety with regards to receiving electric shocks: Pregnancy, epilepsy, seizures, cardiac illness or related conditions; neurologic, psychiatric or psychological condition; and the implantation of a pacemaker or other electric devices. There was one initially accepted participant who experienced severe motion sickness and had to end the experiment early. This data was excluded and replaced by one extra recruited participant.

2.2 Task

Participants were randomly, but gender-balanced (so that the male-female ratio in the groups were approximately similar), assigned to one of two groups. In the control group, participants were instructed to follow traffic rules, but to otherwise drive in the simulator freely for 15 minutes. In the threat group, participants received similar instructions but were additionally told that in case of collision with another car, pedestrian or building, etc., they would receive an electric shock to the ankle. Prior to the driving task, the intensity of the shock was determined for each participant of the threat group. This was done by means of a standardized 6-step calibration procedure. The start intensity for this calibration procedure was 1 Volt. Participant rated the shock on a Likert scale from 1 to 5 (1 = "not perceptible"; 5 = "highly annoying"). According to the participant's rating, the shock intensity for the consecutive shock was increased or decreased. For instance, a rating of 2 (mildly perceptible) for the first shock resulted in an increase of 0.3 Volt for the second shock (See Appendix). The rating of this second shock was then used to increase or decrease the intensity of the third shock, and so on. The finally reached intensity was set as what participants were at risk of experiencing in case of collision.

The 15-minute driving task was split into a 5-minute familiarization phase followed by a 10-minute experimental phase. Participants were informed by the experimenter when the familiarization phase was over and the experimental phase began. Thus, the threat group was brought to believe that the shock machine could deliver shocks only from this point on. The anticipation of an aversive event has been found sufficient to induce anxiety (Phelps et al., 2001). Therefore, participants would not really receive any shock after the calibration procedure in case of collision. In both groups, participants filled in questionnaires before and after the driving task.

2.3 Self-report

ITC-Sense of Presence Inventory (ITC-SOPI). This is a standardized questionnaire, developed by Lessiter, Freeman, Keogh and Davidoff (2001). It was administered directly after completing the driving task. It is recommended by Schuemie and colleagues (2001) due to its high internal consistency coefficients (Cronbach's alphas ranging from .76 up to .94). The sense of presence in the driving simulator was measured with four subscales:

- *Sense of Physical Space* entails questions on the interaction with and physical placement in the simulated environment.
- *Engagement* assesses the tendency to feel psychologically involved with the content.
- *Ecological Validity* evaluates if the simulated environment was perceived as lifelike and real.
- *Negative Effects* relates to the adverse physiological reactions the simulation might evoke, like eyestrain or dizziness.

State-Trait Anxiety Inventory (STAI). This was administered to trace participants' level of induced anxiety throughout the experiment (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). The questionnaire consists of two subsections that can be used independently. The first section measures state anxiety, the second section measures trait anxiety. The control group filled out both sections before starting the experiment, and only the state-related section after the driving task. Additionally, the threat group filled out the state-related section directly after the calibration procedure (*Figure 2*).

NASA Task Load Index (NASA-TLX). This is a tool for evaluating subjective workload over six subscales: *Mental, Physical* and *Temporal Demands, Frustration, Effort* and *Performance* (Hart & Staveland, 1988). It was administered to both groups after the driving task.

2.4 Apparatus

Driving simulator. The simulator (running with the *SILAB* software from the company *wivw*) consisted of a skeletal mock-up car positioned in front of a visual screen. Participants experienced a 180° field of view of the screen when sitting in the driver's seat (*Figure 2*). They were driving with an automated gear box. This setup can be classified as a mid-level driving simulator (Kaptein, Theeuwes, & van der Horst, 1996).

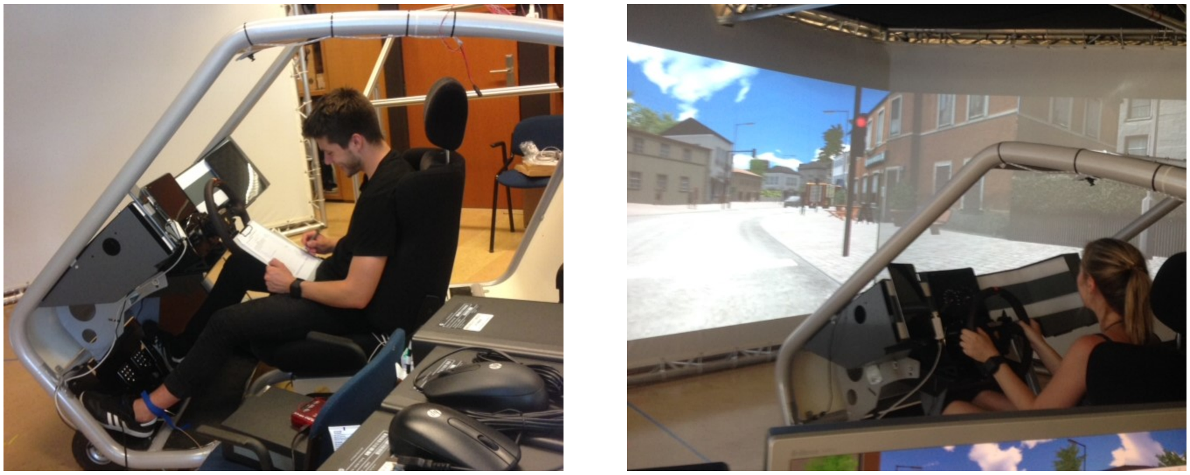

Figure 2. Participants filling out the STAI after calibration (left) and driving in the simulator (right).

The virtual driving scenario was a city scenario (*Figure 3)*. All roads were lined by buildings like houses or stores. It was designed to loop in such a way that whatever turns participants took, they experienced the same scenario. Mildly cloudy weather was simulated to not influence driving. The scenario was filled with pedestrians who walked on sidewalks or crossed the street at traffic lights. Moderate traffic was simulated. To keep participants in a demanding situation, unexpected events occurred on certain points of the scenario. These events included a firetruck with alarm sirens that would cut off the way and force the participant to brake, and parked cars suddenly pulling on the street.

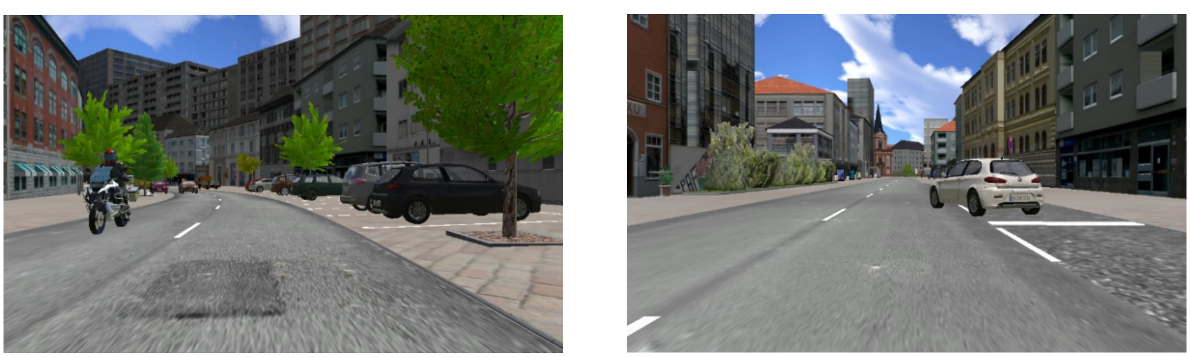

Figure 3. Illustration of the driving environment. Example of moderate traffic in the city environment (left) and a car unexpectedly pulling onto the street (right).

Data on speed, use of brake pedal and steering angle, from which we calculated the steering reversal rate, were recorded. The steering reversal rate was calculated as the number of steering wheel reversals divided by the driving time period (Green, 2013).

Digitimer. For the threat group, electric stimuli were delivered with the *Digitimer For Transcutaneous Stimulation of Nerve and Muscle Tissue* (DS5 2000, Digitimer Ltd.). The electric stimuli were delivered via a lubricated (*Signa Gel* from *parkerlabs*) standard electrode placed on the left ankle. The shock duration of the electric stimulus was 26 ms. Input intensity started at 1 Volt. In the 6-step calibration procedure, a maximum input intensity of 4 Volts could be reached. The current output was set to 50 mA.

Empatica E4. While driving, participants wore a wristband for physiological data acquisition (empatica.com). It collected data on skin conductance (EDA) and heart rate (HR).

2.5 Analysis

The driving task was split into a 5-minute familiarization phase followed by a 10-minute experimental phase. We analyzed these time phases separately, as behavior and physiology in the familiarization (baseline) phase have been found to differ compared to the experimental phase in VR studies (e.g. Meehan, 2000). Furthermore, behavior and physiology have been found to differ at the beginning of an experimental phase compared to the latter half of an experimental phase (Brogni, Vinayagamoorthy, Steed & Slater, 2006). Thus, we split the experimental phase in two. This led to three time phases for statistical analysis with an identical length of 5 minutes each: Familiarization phase (minute 1 to 5), experimental phase 1 (minute 6 to 10) and experimental phase 2 (minute 11 to 15).

The assumption of normality of distributions for all dependent variables (displayed in *Table 1*) was firstly determined by means of the Kolmogorov-Smirnoff test. When the assumption held, analyses assessed differences between the groups by means of repeated measures ANOVAs, MANOVAs, and independent sample t-tests, as described below. For those variables where the assumption of normality of distribution was violated, group differences were determined with the non-parametric Kruskal-Wallis and Man-Whitney tests. Bonferroni corrections were applied when necessary.

A faulty recording of simulator data resulted in missing behavioral data for one participant of the threat group, and only partly recorded driving data for one participant of the control group, which was hence not used in the analysis. One participant of the threat group and one participant of the control group did not fill out the trait-related items of the STAI. All other data was complete.

Table 1

Listing and categories of dependent variables

Dependent Variables (DPs)				
SELF-REPORT				
ITC-SOPI	Sense of physical space	Engagement	Ecological Validity	Negative Effects
STAI	State anxiety before	State anxiety after		
NASA-TLX	Effort	Performance	Frustration	Overall Score
	Mental Demand	Physical Demand	Temporal Demand	
PHYSIOLOGY				
EDA	Familiarization phase	Experimental phase 1	Experimental phase 2	Change of phases
HR	Familiarization phase	Experimental phase 1	Experimental phase 2	Change of phases
BEHAVIOR				
Speed	Familiarization phase	Experimental phase 1	Experimental phase 2	Change of phases
Steering Reversal Rate				
Brake Pedal Use				

3. Results

The post hoc power analysis was performed with G*Power (Faul et al., 2007). Based on the results of a meta-analysis from Horrey and Wickens (2006) on simulator studies, we estimated the effect size of our study to be medium (*ES* = .25; Cohen, 1969). For $N = 32$, *ES* = .25 and $\alpha = .05$, the projected power was $\pi = .28$. A participant count of 32 can be considered rather high in VR research when compared to other highly cited studies (Schuemie et al., 2001). However, this power can be considered low when compared to the desirable power of $\pi = .80$ (Cohen, 1969).

3.1 Self report

Means, standard deviations and p-values for self-report measures are displayed in *Table 2*. Three statistical tests on self-report have been performed. To maintain an alpha level of .05 for those three tests combined, we accepted significance at $p < .017$ after Bonferroni correction. For the ITC-Sense of Presence Inventory, a one-way multivariate analysis of variance (MANOVA) was conducted with the four subscales as dependent variables (DVs) and with group as independent variable (IV). No significant effect was found (Wilk's $\Lambda = .93$, $F(4, 27) = .48$, $p = .751$). Results for both groups were nearly identical on all subscales. Ratings for the NASA-TLX for perceived workload were expected to be higher in the threat group. As visible in *Table 2*, a trend in the expected direction is visible in all subscales except physical demand. However, a MANOVA with the six subscales as DVs and group as IV revealed that those did not reach significance (Wilk's $\Lambda = .83$, $F(6, 25) = .84$, $p = .554$). For the STAI, the anxiety level in the threat group rose slightly after the calibration procedure (*Figure 4*). However, a two-way 2 x 2 repeated measures ANOVA revealed no significant effect of group and time (IVs) on anxiety scores (DVs) ($F(1, 5.10) = .082$, $p = .776$).

Table 2

Means and standard deviations of self-report measures.

	Threat Group		Control Group	
	Mean	SD	Mean	SD
ITC-SOPI (Likert scale 1 – 5)				
Physical Space	2.93	.66	3.12	.63
Engagement	3.52	.51	3.53	.69
Ecological Validity	3.11	.86	3.19	.70
Negative Effects	2.52	1.09	2.43	.94
NASA-TLX (scale 0 – 100)				
Mental Demand	68.75	19.62	60.63	21.28
Temporal Demand	42.19	23.52	37.81	27.14
Physical Demand	54.06	28.70	56.56	17.49
Performance	49.38	32.09	40.63	20.65
Effort	59.31	26.60	53.13	17.88
Frustration	45.00	31.68	32.50	21.37
STAI (scale 20 – 80)				
Beginning of Experiment	32.06	8.65	33.24	8.14
After Calibration	36.63	11.1	n/a	n/a
After Driving	34.13	10.37	34.58	9.32
Trait Anxiety	35.80	11.74	38.44	7.81

Note. Nothing in this table reached significance.

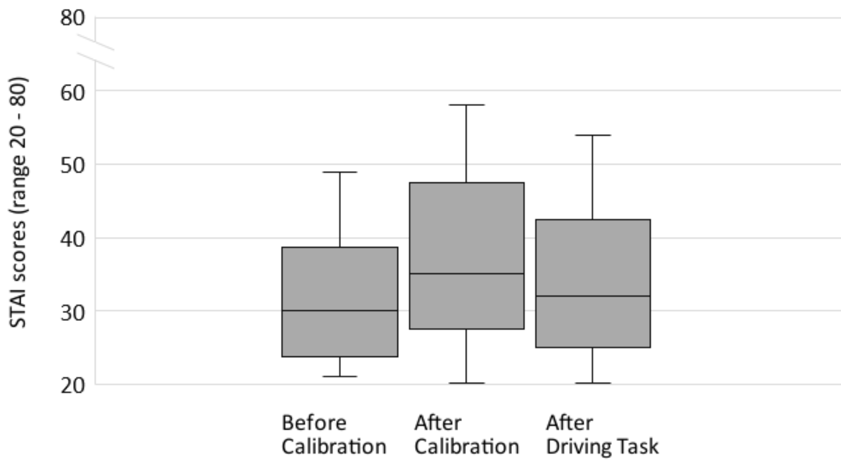

Figure 4. STAI scores within the threat group at the three measured points in time. Scores slightly increased after the shock calibration and dropped again after the driving task. Scores were generally rather low. The change over measured time points was not significant.

3.2 Behavior

Means and standard deviations for behavior can be found in *Table 3*. Five statistical tests of behavioral measures have been performed. To maintain an alpha level of .05 for those fie tests combined, we accepted significance at $p < .01$ after Bonferroni correction. The threat group was expected to drive slower in the experimental phase. Indeed, within the experimental phase, the mean speed of the threat group was lower than in the control group. However, a MANOVA (DVs = speed in km/h per phase; IV = group) revealed that this difference was not significant (Wilk's $\Lambda = .94$, $F(3, 26) = .568$, $p = .641$). We were further interested if the increase of speed from the familiarization to the experimental phase differed between the groups (see *Table 3*). Therefore, we calculated the increase in km/h from the familiarization phase to experimental phase in percent. While the threat group on average increased speed less than the control group, an independent samples t-test revealed no significance ($t(27.165) = -1.119$, $p = .273$).

Table 3

Means and standard deviations of behavioral measures.

	Threat Group		Control Group	
	Mean	SD	Mean	SD
SPEED in km/h				
Familiarization Phase	31.42	8.11	31.00	7.75
Experimental Phase 1	28.93	7.46	31.27	6.17
Experimental Phase 2	31.42	6.49	33.86	6.14
Transition of Phases in %	94.44	25.30	105.82	30.20
STEERING				
Mean Steering Reversal Rate	.2476	.0346	.2745	.0239
BRAKE PEDAL				
Mean Brake Pedal Use	.11	.06	.10	.04

Notes. Steering reversal rate was calculated by dividing the number of steering reversals through the time period driven. Brake pedal use ranged from 0 (no use) to 5 (full pressure on the pedal). The recorded use was averaged over the time period driven. Nothing in this table reached significance.

We expected the threat group to show more steering movement. An independent samples t-test revealed marginal significance between groups ($t(24.858) = 2.473$, $p = .021$). However, increased steering movement was shown by the control group, hence the data laid out in the opposite manner as expected (*Figure 5*).

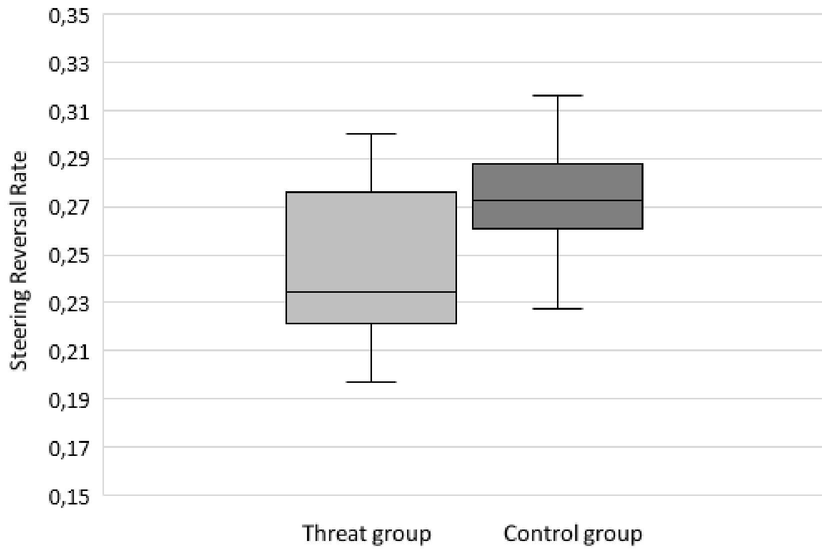

Figure 5. Steering reversal rate, calculated as the number of steering wheel reversals divided by the time period of driving. The control group showed more steering reversals and less variance between participants.

We expected increased brake pedal use in the threat group due to increased cautious behavior. Brake pedal use was not normally distributed according to the Kolmogorov-Smirnoff test, and a Mann-Whitney U test revealed no significant group differences ($U = 110$, $p = .917$). Indeed, the means were strikingly similar for both groups (*Table 3*).

The data of steering reversal rate was lower in the threat group to a marginally significant degree. This outcome was unexpected. To further analyze this, we post-hoc decided to perform an extra analysis. We were interested in a possible positive correlation of steering reversal rate and speed measurement, the behavioral measure that revealed differing group means in the expected manner. This might aid a later interpretation of the data. Indeed, we observed a marginally significant positive one-tailed Pearson correlation between steering wheel reversal rate and mean speed ($r = .382$, $n = 30$, $p = .019$).

3.3 Physiology

Means and standard deviations for physiology can be found in *Table 4*. Three statistical tests on physiological measures have been performed. To maintain an alpha level of .05 for those three tests combined, we accepted significance at $p < .017$ after Bonferroni correction. Due to the high interindividual variability of skin conductance, these scores were z-transformed for every individual before performing the statistical analysis (Ben-Shakhar, 1985; 1987). Skin conductance was expected to be higher in the threat group for the experimental phase. A MANOVA with EDA as the dependent variable and group and phase as independent variables revealed that this was not the case (Wilk's $\Lambda = .901$, $F(3, 28) = 1.02$, $p = .398$). As for the behavioral data, the increase in percentage was calculated from the familiarization phase to the experimental phase. An independent samples t-test revealed no significant results ($t(15.128) = -.187$, $p = .854$). We expected HR to be higher in the threat group than in the control group. A MANOVA (DV = HR, IV = group and phase) revealed no significance (Wilk's $\Lambda = .998$, $F(3, 28) = .018$, $p = .997$).

Table 4

Means and standard deviations and physiological measures.

	Threat Group		Control Group	
	Mean	SD	Mean	SD
EDA (z-transformed)				
Familiarization Phase	-.59	.70	-.85	.48
Experimental Phase 1	.32	.37	.17	.39
Experimental Phase 2	.37	.70	.66	.54
Transition of Phases in %	287.88	570.90	218.11	141.64
HERT RATE (BPM)				
Familiarization Phase	89.9	8.40	90.30	8.30
Experimental Phase 1	90.29	10.40	90.53	7.76
Experimental Phase 2	92.17	11.82	91.61	7.89

Notes. EDA was recorded in microsiemens (µS), and was then z-standardized. HR was recorded in beats per minute (BPM). Nothing in this table reached significance.

3.4 Sorting according to calibration ratings

The mean rating for the shock on a Likert scale from 1 (no sensation) to 5 (very strong) was only moderately aversive (*M* = 3,25, *SD* = 1.0). We wanted to explore possible differences between the control group and those participants of the threat group who rated the shock as more aversive (shock rating of 4 or 5, called the sorted threat group). Therefore, we post-hoc excluded the eight cases of the threat group where shock intensity was rated three or less (8 participants out of *N* = 16). All analyses were performed again. The trends observed above became more pronounced and partly reached significance (*Table 5*). Due to the decreased number of participants, power of this analysis was reduced.

Table 5.

Means and standard deviations of marginally significant results.

	Sorted threat group		Control Group	
	Mean	SD	Mean	SD
SPEED in km/h				
Familiarization Phase	27.24	4.80	31.00	7.75
Experimental Phase 1	23.89	6.43	31.27	6.17
Experimental Phase 2	27.40	3.21	33.86	6.14
STEERING				
Mean Steering Reversal Rate	.2419	.0321	.2745	.0239

Notes. Steering reversal rate was calculated by dividing the number of steering reversals through the time period driven.

A MANOVA revealed that the difference in speed between the groups was only marginally significant (DV = speed in km/h per phase, IV = group) (Wilk's $\Lambda = .69$, $F(3, 18) = 2.74$, $p = .074$, $\eta_p^2 = .313$) (*Figure 6)*. Looking at the single ANOVAs of this analysis, the difference clearly emerged not from the familiarization phase ($p = .26$), but from the experimental phase 1 ($p = .017$) and 2 ($p = .018$). As expected, the participants of the threat group drove slower in the experimental phase. For steering reversal rate, an independent-samples t-test revealed marginal group differences ($t(20) = -2.670$, $p = .015$, Cohen's $d = 1.15$). Contrary to our expectations, the threat group showed a decrease in steering movement. Even though we observed slight differences in the behavior of the two groups, there were still no differences in terms of sense of presence.

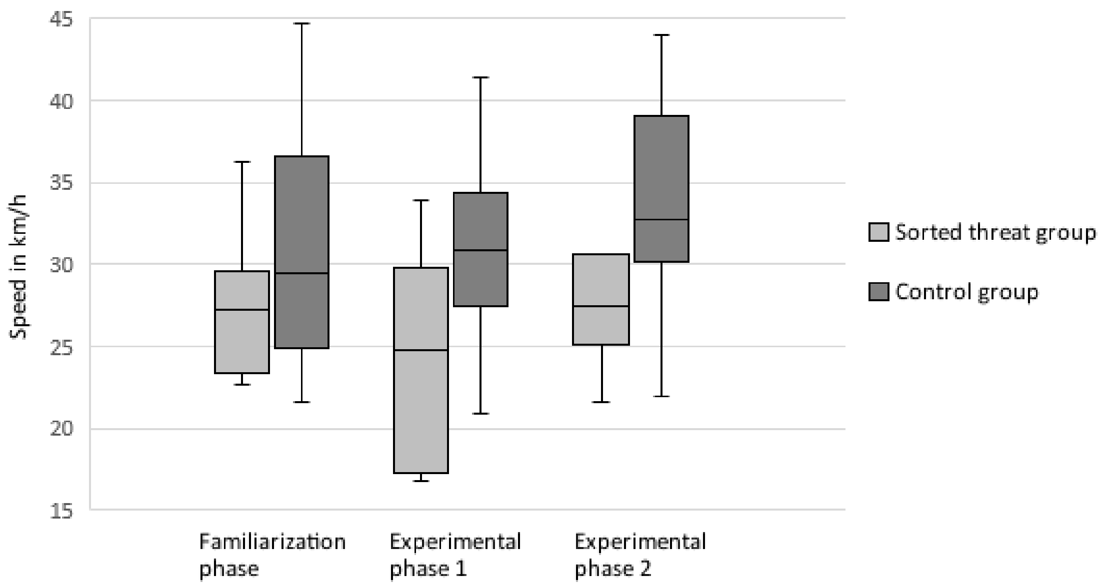

Figure 6. Boxplot of speed in km/h for the control group and sorted threat group per phase. Especially in the experimental phases, the control group drove faster than the sorted threat group. However, this difference did not reach clear significance.

3.5 Crash

A crash was defined as a collision with another object in the simulation, e.g. with a pedestrian, building or other vehicle. Failure of lane keeping or merging onto the sidewalk was not included in that definition. No participant in the threat group crashed during the experimental phase, when the shock machine was supposedly active. In the control group, two participants crashed within this critical time frame.

4 Discussion

The aim of this study was to increase sense of presence in simulator driving. Furthermore, we wanted to clarify the partly inconclusive relationship between sense of presence, anxiety and physiological markers. We added electric shock as an anxiety-inducing risk factor to the driving task. The threat group expected this electric shock in case of collision, while the control group drove in the simulator normally. There were no group differences in self-reports and physiological measurements.

Sense of presence ratings were identical and notably high in both groups. Even though the aversive stimulus has been effective in half of the participants of the threat group, anxiety measures across groups were relatively low and no significant group differences could be found. Marginally significant differences between the groups have been found in steering reversal rate, but in the opposite direction as expected. When excluding participants who rated the shock as low, marginal differences could be found in speed and in steering reversal rate. These, however, were not relevant for most dependent variables, amongst them the most crucial one: Sense of presence.

Assuming that the theoretical background holds and anxiety would increase sense of presence if induced, the lack of results could be due to the following reasons. Firstly, not sufficient anxiety was induced in the threat group, as indicated by low to average STAI scores compared to the norm scores acquired in the development of the questionnaire (Spielberger et al., 1983). In the calibration procedure, even though successfully pilot tested, no level of severe annoyance was reached for each participant of the threat group. The short duration of the shock might be responsible for that (26 ms). Hence, any discomfort was experienced for a brief period of time only. Therefore, the shock might not have induced anxiety. Secondly, arousal might have been induced in the control group as well, overshadowing possible shock effects. Most participants voiced their excitement about driving in a simulator for the first time, a novelty effect that could create arousal indistinguishable from anxiety (Stern, Ray, & Quigley, 2001). Furthermore, the researcher was in the room while the subjects were driving. This could have altered participants behavior and experience as well, e.g. make them more tense (Yoshie, Nagai, Critchley & Harrison, 2016).

Another possible reason why we did not observe group differences in sense of presence measures is that a threat might not increase sense of presence in VR. The analysis of the sorted threat group supports this idea: Even though some differences in behavior due to the anticipation of a shock emerged, this did not influence sense of presence. The significant result for steering reversal rate laid out the opposite way as expected, namely that the control group made more steering corrections. However, it indicates a group difference in behavior that is not reflected in the sense of presence ratings. Notably, sense of presence measures were strikingly high in both groups as compared to the highest achieved score norms in the development of the questionnaire (Lessiter et al., 2001).

Overall, our results suggest that mid-level simulators like ours, typically found in University research, elicit high sense of presence in participants. Increasing sense of presence in highly immersive VR systems, such as the University of Twente driving simulator, might therefore be a hard task. This is the case even though simulated driving cannot lead to physical damage, as might happen during actual driving. Many simulator studies do not measure sense of presence (e.g. Ahlström et al., 2018; Heikoop, de Winter, van Arem, & Stanton, 2017). Though it is advisable to measure sense of presence (Slater, Usoh, & Steed, 1994), our results suggest that this is not a severe limitation of these studies, as it can be assumed to be high in an immersive system like a mid-level simulator. In several simulator studies, authors argue that a lack of expected results or participants acting contrary to the expected manner is due to low sense of presence (even without measuring it) or due to the participants not being afraid of the driving outcome. Authors of these simulator studies use statements like “lack of perceived danger” (Helland et al., 2013, p.5), “lack of physical crash risk” (Melman, Abbink, van Paassen, Boer, & de Winter, 2018, p. 980), “because driving errors in the simulator do not have serious consequences” (Young, Regan, & Lee, 2009, p. 89) or that participants “knew that a potential crash would not cause them physical harm” (Heikoop et al., 2017, p. 124). Our results suggest that statements like these might not be indicators of major limitations of simulator studies. A lack of finding the expected results is likely not linked to a lack of risk and low sense presence.

The relationships between the constructs of sense of presence, anxiety and their measurements are still stimulating a debate in research (Felnhofer et al., 2014). The results by Baños and colleagues (2004), who found that sense of presence is influenced by immersion as well as emotion, do only hold for VR systems low in immersiveness. A study by Jang and colleagues (2002) found differing EDA patterns in groups of low and high arousing VEs, but not in presence ratings. Busscher, de Vliegher, Ling & Brinkman (2011) found differences in arousal, but not in presence ratings. In line with Jang et al. (2002) and Busscher et al. (2011), we found no difference in presence ratings, even though other group differences (speed and steering reversal rate) were found. Frequently, authors claim that the effect of increased sense of presence existed nonetheless, but was not measurable. The described reasons for this are, e.g., participants focusing on the steering wheel (Jang et al., 2002).

Further assumptions that relations exist but have not been found in VR studies hold, for example, for the null-results for HR: In many studies (e.g. Meehan, 2000; Jang et al., 2002), no link between VR and HR have been found. Explanations include the inconclusiveness of results due to artifacts (Meehan, 2000) or small sample sizes (Jang et al., 2002). Our study shows that such conclusions, made so often in the field of VR, might need a closer look.

The main goal of this study was to increase sense of presence by creating an anxiety inducing environment. STAI scores revealed that no anxiety was induced. While conclusions on the relations of sense of presence and several behavioral, physiological and self-report measures have been drawn, the main hypothesis was not supported and all further interpretation of null-results has to be taken with caution.

Additionally, we want to draw attention to the unexpected result of the increased steering reversal rate in the control group. The hypothesis was that the threat group, in an effort to stay in the middle of the lane, would make more steering corrections, quantifiable as steering reversal rate. This expectation was not only rejected, but the data suggested the opposite result. Markkula and Engström (2006), point toward the difficulty of interpreting steering activity, and stress that steering activity “may be indicative of both increased and reduced lane position variance” (p. 1). As our data surprised us, we wondered if it is correlated with another marginally significant behavioral measure which produced means in the manner that we had expected, namely speed. Hence, we post-hoc decided to perform a Pearson correlation with those two variables, and found that they indeed were positively correlated. We concluded from this that increased speed and reduced steering reversals were caused by a shared factor in our study, which we assume to be the risk of receiving an electric shock. More correcting steering movements might be an indicator of less need for attention rather than an effort to stay in the middle of the lane. Supporting findings come from Zhao et al. (2012), who found that higher steering reversal rates were correlated with increased lapses and throttle accelerations, which they attributed to “more failures of attention” (p. 683). This would be in line with the significant correlation with increased speed. Possibly, increased speed also simply demanded more steering correction. However, more research is needed to clarify this. Future studies might also define a certain minimum angular value for their analysis of steering reversal rate, which is recommended by Markkula and Engström (2006), but which we did not include.

For future research, a higher number of participants would increase the power of the analysis and therefore the relevancy of obtained results. It would be advisable to design a calibration procedure with a stimulus that is similarly and clearly aversive to all participants. Conclusions drawn on the relationship between sense of presence, anxiety and the increased validity of a driving simulator will be more solid then. Considering that sense of presence was strikingly high already, it is questionable if inducing severe anxiety is worth the little, if at all, increase of sense of presence in a driving simulator.

Conclusion

The debate around the relationship between anxiety and sense of presence is still ongoing (Felnhofer et al., 2014). Our data supports the notion that sense of presence is, firstly, not so easily measured with physiological and behavioral indicators and, secondly, not so easily influenced. Even drastic measures, like introducing an electric shock, did not enhance sense of presence ratings. For simulator driving in a mid-level simulator, sense of presence ratings were already satisfyingly high even in the control group.

Humanity is increasingly concerned with the downsides of blurring virtual and actual reality. Considering the high presence ratings for simulators, which make them an excellent tool for training and research, realistic virtual reality is undoubtedly a virtue.

References

Ahlström, C., Anund, A., Fors, C., & Åkerstedt, T. (2018). The effect of daylight versus darkness on driver sleepiness: a driving simulator study. *Journal of sleep research*, *27*(3), 1-9.

Alsina-Jurnet, I., & Gutierrez-Maldonado, J. (2010). Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments. *International Journal of Human Computer Studies, 68*, 788–801.

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and emotion: Their impact on the sense of presence. *CyberPsychology & Behavior, 7*(6), 734-741.

Bassett, J. R., Marshall, P. M., & Spillane, R. (1987). The physiological measurement of acute stress (public speaking) in bank employees. *International Journal of Psychophysiology*, *5*(4), 265-273.

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, *22*(3), 292-299.

Ben-Shakhar, G. (1987). The correction of psychophysiological measures for individual differences in responsivity should be based on typical response parameters: A reply to Stemmler. *Psychophysiology*, *24*(2), 247-249.

Brogni, A., Vinayagamoorthy, V., Steed, A., & Slater, M. (2006, November). Variations in physiological responses of participants during different stages of an immersive virtual environment experiment. In *Proceedings of the ACM symposium on Virtual reality software and technology* (376-382). New York: ACM.

Busscher, B., de Vliegher, D., Ling, Y., & Brinkman, W. P. (2011). Physiological measures and self-report to evaluate neutral virtual reality worlds. *Journal of CyberTherapy and Rehabilitation*, *4*(1), 15-25.

Bystrom, K. E., Barfield, W., & Hendrix, C. (1999). A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *8*(2), 241-244.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. *Handbook of Emotions*, *2*, 173-191.

Casutt, G., Theill, N., Martin, M., Keller, M., & Jäncke, L. (2014). The drive-wise project: driving simulator training increases real driving performance in healthy older drivers. *Frontierts in Aging Neuroscience, 6,* 85.

Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences.* NY: Academic Press

Diemer, J. E., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology, 6*(26), 1-9.

Espié, S., Gauriat, P., & Duraz, M. (2005). Driving simulators validation: The issue of transferability of results acquired on simulator. In *Driving Simulation Conference North-America (DSC-NA 2005).* Orlando, FL.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Faure, V., Lobjois, R., & Benguigui, N. (2016). The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior. *Transportation research part F: traffic psychology and behaviour*, *40*, 78-90.

Felnhofer, A., Kothgassner, O. D., Hetterle, T., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2014). Afraid to be there? Evaluating the relation between presence, self-reported anxiety, and heart rate in a virtual public speaking task. *Cyberpsychology, Behavior, and Social Networking*, *17*(5), 310-316.

Feng, R., Zhang, G., & Cheng, B. (2009). An on-board system for detecting driver drowsiness based on multi-sensor data fusion using Dempster-Shafer theory. In *Networking, Sensing and Control, 2009. ICNSC'09. International Conference,* 897-902.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident analysis & prevention, 34*(5), 589-600.

Green, P. (2013). Standard definitions for driving measures and statistics: Overview and status of recommended practice J2944. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications* (184-191). New York: ACM.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139-183.

Heikoop, D. D., de Winter, J. C., van Arem, B., & Stanton, N. A. (2017). Effects of platooning on signal-detection performance, workload, and stress: a driving simulator study. *Applied Ergonomics, 60*, 116-127.

Helland, A., Jenssen, G. D., Lervåg, L. E., Westin, A. A., Moen, T., Sakshaug, K., ... & Slørdal, L. (2013). Comparison of driving simulator performance with real driving after alcohol intake: A randomised, single blind, placebo-controlled, cross-over trial. *Accident Analysis & Prevention, 53*, 9-16.

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology*, *92*(1-2), 84-89.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors*, *48*(1), 196-205.

Jang, D. P., Kim, I. Y., Nam, S. W., Wiederhold, B. K., Wiederhold, M. D., & Kim, S. I. (2002). Analysis of physiological response to two virtual environments: driving and flying simulation. *CyberPsychology & Behavior*, *5*(1), 11-18.

Jorna, P. G. A. M. (1993). Heart rate and workload variations in actual and simulated flight. Ergonomics, *36*(9), 1043-1054.

Kaptein, N., Theeuwes, J., & Van Der Horst, R. (1996). Driving simulator validity: Some considerations. *Transportation Research Record, 1550*(1), 30-36.

Keith, K., Trentacoste, M., Depue, L., Granda, T., Huckaby E., Ibarguen, B., Kantowitz, B., Lum, W., & Wilson, T. (2005). *Roadway human factors and behavioral safety in Europe*. Report No.: FHWA-PL-05-005.

Krajewski, J., Sommer, D., Trutschel, U., Edwards, D., & Golz, M. (2009). Steering wheel behavior based estimation of fatigue. *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 118-124.

Lande, S. D. (1982). Physiological and subjective measures of anxiety during flooding. *Behaviour research and therapy*, *20*(1), 81-88.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. Presence: *Teleoperators and virtual environments*, *10*(3), 282-297.

Markkula, G., & Engström, J. (2006). A steering wheel reversal rate metric for assessing effects of visual and cognitive secondary task load. In *Proceedings of the 13th ITS World Congress*. Leeds.

Meehan, M. (2000, March). An objective surrogate for presence: Physiological response. Paper presented at the *Presence 2000 Workshop*. Delft.

Melman, T., Abbink, D. A., van Paassen, M. M., Boer, E. R., & de Winter, J. C. (2018). What determines drivers' speed? A replication of three behavioural adaptation experiments in a single driving simulator study. *Ergonomics, 61*(7), 966-987.

Nakayama, O., Futami, T., Nakamura, T., & Boer, E. R. (1999). Development of a steering entropy method for evaluating driver workload. *SAE transactions*, 1686-1695.

Nilsson, L. (1993). *Behavioral research in an advanced driving simulator – experiences of the VTI system.* Proceedings of the Human Factors an Ergonomics Society 37th Annual Meeting, 612-616. Seattle.

Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., & Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. *Nature neuroscience, 4*(4), 437-441.

Sang, C. N., Max, M. B., & Gracely, R. H. (2003). Stability and reliability of detection thresholds for human A-Beta and A-delta sensory afferents determined by cutaneous electrical stimulation. *Journal of pain and symptom management*, *25*(1), 64-73.

Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, *4*(2), 183-201.

Schmidt-Daffy, M. (2013). Fear and anxiety while driving: Differential impact of task demands, speed and motivation. *Transportation research part F: traffic psychology and behaviour*, *16*, 14-28.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 3*(2), 130-144.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). State-trait anxiety inventory (form Y). Redwood City. CA: Mind Garden, 77.

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological recording*. Oxford University Press, USA.

Tursky, B. (1974). Physical, physiological, and psychological factors that affect pain reaction to electric shock. *Psychophysiology*, *11*(2), 95-112.

Vrijkotte, T. G., Van Doornen, L. J., & De Geus, E. J. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, *35*(4), 880-886.

Yoshie, M., Nagai, Y., Critchley, H. D., & Harrison, N. A. (2016). Why I tense up when you watch me: Inferior parietal cortex mediates an audience's influence on motor performance. *Scientific reports*, *6*, 19305.

Young, K. L., Regan, M. A., & Lee, J. D. (2009). *Measuring the effects of driver distraction: Direct driving performance methods and measures*, 85-105. Boca Raton, FL: CRC Press.

Zhao, N., Mehler, B., Reimer, B., D'Ambrosio, L. A., Mehler, A., & Coughlin, J. F. (2012). An investigation of the relationship between the driving behavior questionnaire and objective measures of highway driving behavior. *Transportation research part F: traffic psychology and behaviour*, *15*(6), 676-685.

Appendix

Calibration Procedure Guideline

<table>
<tr><th colspan="5">Start at 1 V
Subjects rating
(tick one, change for next level)</th><th>Participant # _____</th></tr>
<tr><td>Nothing</td><td>Perceptible</td><td>Mildly annoying</td><td>Annoying</td><td>Very strong</td><td>Comment</td></tr>
<tr><td>1
(+0.5)</td><td>2
(+0.3)</td><td>3
(+0.1)</td><td>4
(=)</td><td>5
(-0.2)</td><td></td></tr>
<tr><td>1
(+0.5)</td><td>2
(+0.3)</td><td>3
(+0.1)</td><td>4
(=)</td><td>5
(-0.2)</td><td></td></tr>
<tr><td>1
(+0.5)</td><td>2
(+0.3)</td><td>3
(+0.1)</td><td>4
(=)</td><td>5
(-0.2)</td><td></td></tr>
<tr><td>1
(+0.5)</td><td>2
(+0.3)</td><td>3
(+0.1)</td><td>4
(=)</td><td>5
(-0.2)</td><td></td></tr>
<tr><td>1
(+0.5)</td><td>2
(+0.3)</td><td>3
(+0.1)</td><td>4
(=)</td><td>5
(-0.2)</td><td></td></tr>
<tr><td>1
(+0.5)</td><td>2
(+0.3)</td><td>3
(+0.1)</td><td>4
(=)</td><td>5
(-0.2)</td><td></td></tr>
<tr><td colspan="6">Final Intensity: _______ Final Rating: _________</td></tr>
</table>