

Psychological Feasibility of a Virtual Cycling Environment for Human-in-the-Loop Experiments

Lukas Stratmann,¹ Dominik S. Buse,¹ Julian Heinovski,¹ Florian Klingler,¹
Christoph Sommer,¹ Jan Tünnermann,² Ingrid Scharlau,¹ and Falko Dressler¹

Abstract: Vulnerable Road Users (VRUs) are moving into the focus of road safety solutions, be it Advanced Driver Assistance Systems (ADAS) deployed in cars or enhanced guidance systems in smart cities. We concentrate on bicyclists and study means for assessing the impact of such safety solutions on the bicyclists' behavior. We just recently developed the Virtual Cycling Environment (VCE), which allows a bicyclist to ride in a virtual reality environment while sitting on a real bike connected with a simulated environment. In this paper, we present a method for measuring the cyclists' visual attention. This is a fundamental requirement for advanced studies of VRU safety systems including their performance as well as acceptance. In a small empirical study, we assess the feasibility of our approach for conducting psychological experiments in a human-in-the-loop setup.

Keywords: Vulnerable road users; bicyclists; safety system; integrated simulation; virtual reality

1 Introduction

The safety of cars has seen constant improvements in the last years: new technologies, mandatory safety features, and an improved understanding of vehicles have led to fewer fatal accidents. The same cannot be said for Vulnerable Road Users (VRUs) such as bicyclists. While their safety has partly improved, developments compared to other road users are lagging far behind [EC18]. Other than VRUs, cars can be (and are) equipped with passive safety equipment (that is, equipment that mitigates the consequences of an accident) such as seatbelts and airbags. For bicycles, active safety equipment such as Advanced Driver Assistance Systems (ADAS) for the prevention of accidents are a promising solution. Typically, such systems require the cooperation of the cyclists; they must react to warnings about imminent collisions. This means the cyclists need to notice and react to such signals. Consequently, it is of great importance to understand how cyclists react to warning systems, how much time they need to process them, and which kinds of signals are easily understood.

Performing experiments for investigating ADAS for VRUs in real traffic is often not a feasible option, as many types of experiments would expose cyclists to danger of physical harm. In addition, such experiments commonly suffer from low reproducibility. We therefore

¹ Paderborn University, Paderborn, Germany {lukas.stratmann,buse,heinovski,klingler,sommer,dressler}@ccs-labs.org,scharlau@mail.upb.de

² University of Marburg, Marburg, Germany jan.tuennermann@uni-marburg.de

propose to use highly realistic interactive simulations in which a cyclist rides a real bicycle through a simulated Virtual Cycling Environment (VCE). Taking advantage of such human-in-the-loop simulations, the behavior of cyclists and their interaction with future ADAS can be researched under safe and reproducible conditions. In previous work [He19], we described the potential of such a VCE for investigating the potential of improved ADAS purely for cars, i.e., without interacting with the bicyclist. We employed the VCE to record (and make available) movement traces of bicyclists behaving naturally in traffic.

In this paper, we specifically investigate changes in the *visual attention* of participants riding in the VCE under high or low traffic density conditions. In doing so, we explore the potential of the VCE as a tool for future psychological studies pertaining to the safety of VRUs, their perception of the environment, as well as their interaction with traffic or with ADAS. We assess the perceived realism of the VCE to identify potential areas for future improvement.

In essence, our main contributions can be summarized as follows:

- We present a method for measuring cyclists' visual attention using the extended VCE based on a formal theory of visual attention.
- We provide first insights into changes in the visual attention of cyclists when exposed to varying amounts of simulated traffic.
- We identify important aspects for participants' acceptance of such a VCE, in particular, the accuracy of steering controls and the predictability of how simulated cars will react to the presence of the cyclist.

2 Related Work

In the literature, a number of different interactive bicycle simulators have already been developed, e. g., [He19; Kw01]. However, there have not been many efforts to use such simulators for conducting research on the safety of VRUs. Notable exceptions are presented by Matviienko et al. [Ma18], who use a non-steerable bicycle simulator to evaluate the effectiveness of a variety of warning signals for children, as well as the work presented by Plumert et al. [PKC04], who use a virtual cycling environment with simulated traffic to study the behavior of children on bicycles. Experiments on people's attention in simulated traffic, however, so far appear to be confined to the realm of motorized vehicles.

The focus of the present study is on the visual attention of cyclists. Measuring attention in traffic so far has been done with a wide variety of metrics. For example, Strayer et al. [SDJ03] attempted to quantify attention, among other metrics, by collecting drivers' following distance to the car in front and the delays when performing braking maneuvers. These metrics are interesting as they are directly concerned with effects of distracted driving that can result in people being harmed. On the downside, this particular method is tailored to the situation of one car equipped with brake lights being followed by another, which is

not easily transferable to bicycles. As another metric, the authors employed a memory task in which participants had to remember billboards shown during the experiment. Such a task would be easy to adapt for cyclists, but as a between-subjects design, this paradigm would require a substantially larger number of participants the more conditions are to be tested.

Some literature on the effect of traffic density on road users' attention already exists. Strayer et al. [SDJ03] manipulated the traffic density between subjects and found that talking on the phone impaired people's car driving performance more strongly when the traffic density was high. Vlakveld et al. [V115] conducted a field experiment in which they observed cyclists' mental workload and cycling speed in low and high complexity traffic situations while also comparing conventional bicycles to e-bikes. For measuring mental workload, the authors instructed their participants to react to blinking LEDs in their peripheral vision and used a combined metric of reaction time and hit rate. As the authors concede, one difficulty with such field experiments is the selection of appropriate traffic situations for VRUs that are neither too simple nor too dangerous: *The only safe way to investigate cyclists['] behaviour in real complex traffic situations probably is in a bicycling simulator.*

It is exactly this gap in research which we are focusing on by presenting a methodology for measuring cyclists' visual attention in simulated traffic. This method is based on established research in the domain of visual attention and allows for the examination of arbitrarily complex traffic situations without the risk of physical harm.

3 Experimental Study

We conduct an experimental study using our VCE [He19] to measure cyclists' visual attention in the context of different traffic densities. In the following, we outline details about this study and elaborate the research methods and setup of the experiment.

Design We measure participants' visual attention depending on traffic density as a within-subjects independent variable. The *low traffic* condition is defined as an absence of other road users and *high traffic* as an average of 3.6 cars appearing in the cyclist's visual field in every street of this condition. For measuring visual attention, we incorporated a Temporal Order Judgment (TOJ) experiment similar to those described by Tünnermann et al. [TPS15]. Typically, such experiments involve abstract synthetic displays of stimuli in the form of shapes or letters. First attempts have been made to incorporate such experiments in a dynamic environment [Tü16, Sec. 6.4.3]. We take this idea one step further to make predictions about visual attention in our VCE geared towards realism. Participants were asked to navigate a course lined with TOJ trials for which we chose pairs of floating gem stones, as shown in Fig. 1c. After a pair of gem stones appears, first one and then the other flicker in quick succession. The task of the cyclist then is to collect that gem stone which they think flickered first. According to Bundesen's Theory of Visual Attention (TVA) [Bu90], the information that an arbitrary stimulus x in the visual field belongs to some category i is encoded in the capacity-limited visual short-term memory. The processing rate $v(x, i)$

denotes the rate at which this encoding happens for a specific stimulus and category. The sum over all processing rates of all objects and categories is called the overall processing capacity C . As we will show in Sect. 4, these parameters can be estimated from the results of the TOJ task. We hypothesize that an increase in traffic density will cause the allocation of more attentional weight to background stimuli such as the moving cars, which effectively reduces the capacity C available for the order judgment. Given a participant j , we call the processing capacity allotted to the target stimuli of the task in the *high traffic* condition C_j^{high} , and C_j^{low} in *low traffic*. C_μ^{high} and C_μ^{low} are the averages of these processing capacities over all participants. According to our hypothesis, we expect C_μ^{high} to be lower than C_μ^{low} .

After the experiment, we administer a set of questionnaires on the features of the VCE with the aim to estimate the quality of the simulation as a whole: The Flow Short Scale (FSS) by Rheinberg et al. [RVE03], the Igroup Presence Questionnaire (IPQ) by Schubert et al. [SFR01], and a collection of seven custom questions (cf. Tab. 1) specific to our VCE. Each questionnaire targets a different objective. Our intention with measuring flow using the FSS is to get a sense of people’s acceptance of the simulation and of their task regardless of whether they perceive the environment as realistic or not. Rheinberg and Engeser [RE18] characterize the experience of flow as “unselfconscious and complete immersion in a pursuit that, although requiring high levels of skill and concentration, results in a sense of effortless action and control.” Since an important goal of our VCE is to achieve a high degree of realism, we include the IPQ for its realness factor. The custom questions are intended to differentiate between several factors that might influence perceived realism, e.g., whether the behavior of other vehicles felt natural or to what degree the acts of accelerating and braking felt realistic. With the exception of one open-ended question, all custom questions allow responses on a 7-point scale ranging from *strongly disagree* (0) to *strongly agree* (6).

Apparatus We make use of our VCE [He19], which we slightly adapted for our experiments. In order to provide a wide angle of view and to avoid simulator sickness, we choose a triple monitor setup instead of Virtual Reality (VR) glasses. As shown in Fig. 1a, the bicycle is positioned in front of a desk with three 24" 1920 × 1200 monitors positioned 1.5 m from the handle bar, which all operate at a refresh rate of 60 Hz in sync with the constant refresh rate of the visualization. We use an infrared sensor to measure the speed of the rear wheel of the physical bicycle. Further, we obtain the steering angle from the accelerometer and gyroscopic sensors of a smartphone. The sensor readings are fed into the visualization implemented in Unity, which in turn is coupled with our Ego-Vehicle Interface (EVI) [Bu18]. The main task of the EVI is to synchronize bicycle position, visualization, and an urban traffic simulation performed using SUMO.

Stimuli While navigating the virtual road network, cyclists find pairs of red bumps in intervals of 9 m, separated by a 1 m gap, which mark the beginning of a trial. Navigating the bicycle through such a pair will trigger the appearance of two yellow stimuli, called *probe* and *reference*, in the line of sight and slightly to the left and right of the cyclist’s path, i.e., within that part of the visual field that is important for cycling. They are positioned at a

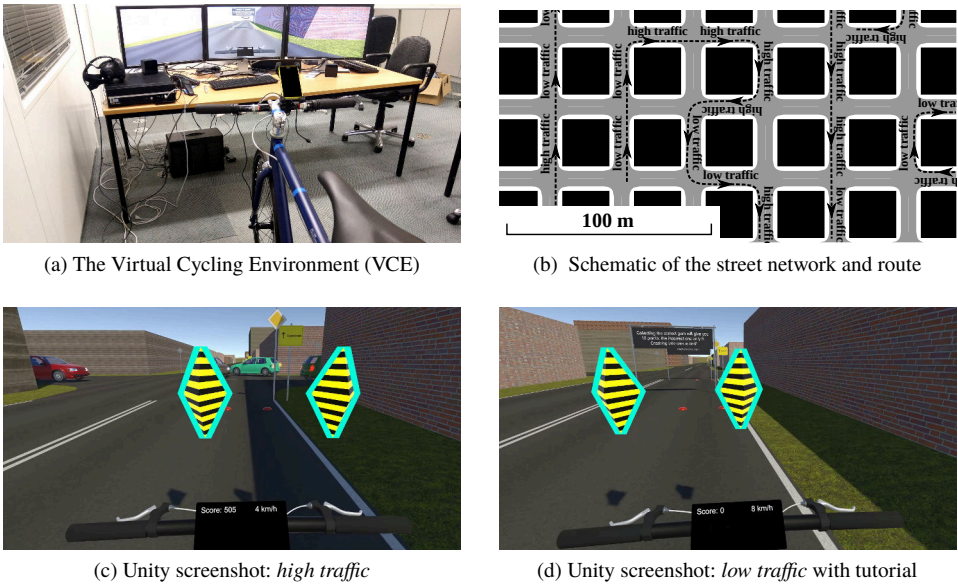


Fig. 1: Experimental setup and screenshots of the scenario for the first block of the experiment

distance of 4.5 m from the pair of red bumps and separated by 0.6 m. In order to minimize potentially confounding effects due to a decrease in contrast between turn indicator lights of cars and the stimuli, we added an additional turquoise outline and black stripes. A screenshot of such a trial is shown in Fig. 1c. After a randomized delay with a duration between 15 and 22 frames, the stimuli of the current trial flicker one after the other with a delay of -5 to 5 frames corresponding to the Stimulus Onset Asynchrony (SOA). A negative SOA indicates that the probe stimulus will flicker first. Excluding the tutorial, each participant encounters at most 456 trials in total, which amounts to approximately 3 trials per street. The traffic density is manipulated within subjects by declaring stretches of between 2 and 5 consecutive streets as either *low traffic* or *high traffic*, as illustrated in Fig. 1b. New traffic is dynamically generated only when the cyclist is about to enter a *high traffic* street. The paths of new cars are carefully chosen to not include any *low traffic* streets and, at the same time, a substantial portion of cars is guaranteed to be visually relevant to the participant by intersecting their predefined routes. On average, a participant encounters 3.6 vehicles between two junctions on every *high traffic* street.

Participants We recruited a total of 19 participants (3 female and 15 male, and one choosing not to answer), whose age ranged from 14 to 30 years. With the exception of 2 persons, all participants were students. The experiment was conducted according to the principles expressed in the *Declaration of Helsinki* and approved by the ethics committee of Paderborn University. Participation was voluntary and without monetary compensation.

Procedure We split the experiment into two blocks of three levels each, hoping to counter fatigue by allowing participants to take breaks. After ensuring participants' informed consent, they were given the chance to familiarize themselves with the simulator in a short four-streets-long tutorial level. The tutorial included banners explaining the rules of the game. Further, the participants were promised the chance to enter themselves into a high score list at the end of each of the two experiment blocks. The Flow Short Scale (FSS) was always filled out immediately after the end of the first three levels. Afterwards, participants continued with the last three levels. Once finished, demographics and survey data for our custom questions and for the IPQ were collected. Each of the blocks took about 20 min to complete. In total, participants typically completed the experiment in less than 60 min.

In order to motivate participants to do their best in the order judgment task (cf. [Tü16, Sec. 6.4.3]), the VCE was extended with a gamification aspect. Players received 10 points for the correct stimulus and 5 points for the incorrect stimulus to incentivize always collecting one of the gem stones. Collisions with cars resulted in the deduction of 20 points to discourage reckless cycling. As shown in Fig. 1b, the scenario for each of the two blocks consisted of a grid road network with 15×15 junctions placed 35 m apart. A random path was constructed visiting all but the outermost junctions in the scenario. 78 street segments of this path were divided into three equally long levels. At the end of each level, the experiment was paused automatically the cyclist was shown the number of points collected in the completed level, the maximum possible number of points that could have been collected in this level, as well as the total number of points so far.

4 Results and Discussion

The experimental setup described in Sect. 3 yields the numbers of times participants collected the correct or incorrect gem stone of a stimulus pair under each condition. In order to estimate the visual processing capacities C_{μ}^{high} and C_{μ}^{low} for the high and low traffic density conditions, respectively, we used the Bayesian model depicted in Fig. 2 based on the work of Tünnermann [Tü16]. The variables n_{ji}^{high} and n_{ji}^{low} denote the number of trials for a given participant j and SOA i . Among these trials, y_{ji}^{high} and y_{ji}^{low} denote the number of times a participant reported the probe stimulus to have flickered first. Correspondingly, the function $P_{p^{\text{1st}}}(C, \text{SOA})$ is the probability that a given participant reports the probe stimulus to have flickered first. Based on the TVA, a concrete function can be derived that links the *probe first* report probability to TVA rate parameters v_{probe} and $v_{\text{reference}}$ [Tü16, p. 76]. In our case, where both stimuli are exactly the same in appearance, we can assume $v_{\text{probe}} = v_{\text{reference}} = \frac{C}{2}$, which leads to

$$P_{p^{\text{1st}}}(C, \text{SOA}) = \begin{cases} 1 - \frac{1}{2} \exp\left(-\frac{C}{2}|\text{SOA}|\right) & , \text{ if } \text{SOA} < 0 \\ \frac{1}{2} \exp\left(-\frac{C}{2}|\text{SOA}|\right) & , \text{ if } \text{SOA} \geq 0. \end{cases} \quad (1)$$

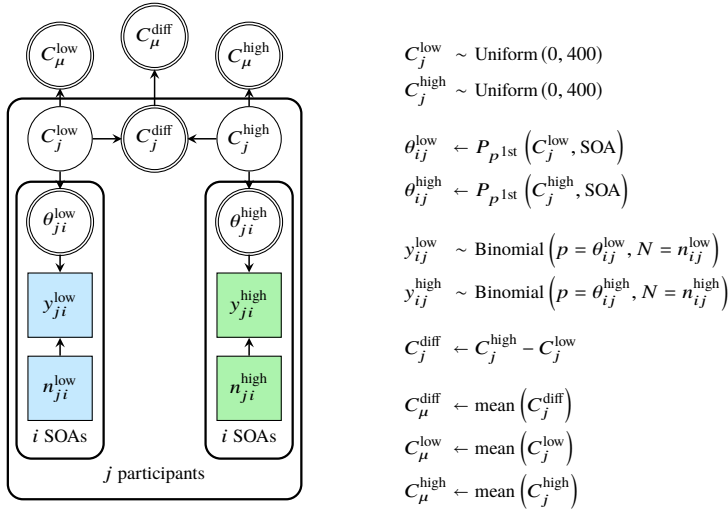


Fig. 2: Bayesian model: Shaded nodes denote observed variables, discrete variables are shown as rectangles, continuous variables as circles, and deterministic variables have a double border.

The model was estimated using the NUTS sampler [HG14] (20 000 iterations in four chains with the default parameters) implemented in PyMC3 [SWF16]. The priors on the C parameters were set to a uniform distribution from 0 to 400, a conservative choice, given that typical values for C range from 50 Hz to 70 Hz [Tü16].

We found the estimated visual processing capacity in the *high traffic* condition to be lower than in the *low traffic* condition, cf. Fig. 3a. This confirms our hypothesis that $C_\mu^{\text{high}} < C_\mu^{\text{low}}$. The individual capacities C_μ^{high} and C_μ^{low} had their modes at 57.2 Hz and 69.0 Hz, respectively, and 95 % Highest Density Intervals (HDIs) of [52.5 Hz, 62.1 Hz] and [63.6 Hz, 75.3 Hz]. The 95 % HDI for the capacity difference C_μ^{diff} did not include a value of 0 Hz difference. Instead, it ranged from -19.4 Hz to -4.4 Hz with a mode of -12.2 Hz, cf. Fig. 3b. These results are in line with findings by Strayer et al. [SDJ03] and Vlakveld et al. [V115], who found an effect of traffic density on mental workload. TVA capacity values can be interpreted as items processed per second. Because this number depends on different contextual factors, the magnitude has no easy or canonical interpretation. However, the results are reasonable because they agree with experimental studies with different stimuli (e.g., [Tü16]). It would be interesting to test how C responds to introducing still more cars, other road users like pedestrians and other cyclists, or further events.

In a next step, we have a closer look on the participants' acceptance of the virtual environment and their gamified task to collect the correct gem stones. The FSS yielded a mean (M) flow score of 4.92 (standard deviation $SD = 0.87$), with a mean *worry* factor of 3.37 ($SD = 1.24$). To put this flow score into context, we can compare it to a number of reference activities

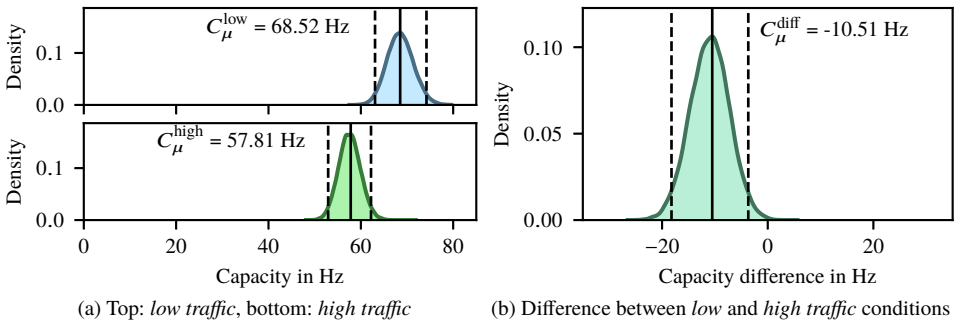


Fig. 3: Posterior distributions of visual processing capacity; the 95 % HDIs are overlaid as dashed lines

Item	0	3	6	M	SD
Using the simulator felt like riding a real bicycle.				2.58	1.46
Steering felt realistic.				2.11	1.52
Accelerating and braking felt realistic.				3.89	1.41
The virtual bicycle behaved according to my expectations from experience with real bicycles.				2.63	1.38
I felt confident in my ability to control the bicycle.				2.11	1.63
Other road users behaved according to my expectations from the real world.				3.47	1.87
I tried to adhere to traffic rules as much as in the real world.				3.74	1.52

Tab. 1: Responses to VCE-related questionnaire (from 0, strongly disagree to 6, strongly agree)

by Rheinberg et al. [RVM05, p. 32]. Of the 9 activities with a lower score, notable ones include watching TV or sleeping, while the 17 higher-rated examples include reading and social communication, but also household chores and doing office work. Results from our evaluations indicate promising opportunities for improvements of our VCE. One potential reason for the disruption of the experience of flow is a lack of control [RE18, p. 601]. Responses to our custom questionnaire items listed with their means and standard deviations in Tab. 1 showed low ratings for participants’ confidence in their control of the bicycle and for their judged realism of the steering mechanism. Written comments in response to *What should be changed in order to improve the realism or the usability of the cycling simulation?* show a similar picture: the aspect of steering was mentioned negatively 11 times. Another frequent comment (10 counts) was in regard to the behavior of other virtual road users, which frequently did not yield to bicyclists although traffic signs suggested otherwise. These results present a possible explanation for the outcome of the IPQ, which produced an average realness factor of 2.20 ($SD = 0.92$). This places the realness score of our VCE in the 58th percentile of valid submissions in the reference data given for the IPQ.³

³ Reference data was obtained from <http://www.igroup.org/pq/ipq/data.php>

5 Conclusion

In this paper, we investigated a methodology for investigating future Advanced Driver Assistance Systems (ADAS) for Vulnerable Road Users (VRUs). We took advantage of novel human-in-the-loop simulation techniques realized in an extension of our Virtual Cycling Environment (VCE). Based on a formal theory of visual attention, we presented a methodology for measuring cyclists' visual attention in simulation. By conducting an empirical study with multiple participants, we evaluated the impact of road traffic density on the visual attention of cyclists. Our results indicate that high road traffic density requires measurably more visual attention for VRUs than traffic with a low density. The results from our purely virtual trials are in line with recent studies regarding the mental workload of road users [SDJ03; V115], thus, suggesting the suitability of the VCE for in-depth psychological studies. In a next step, we will use our VCE to conduct ADAS-testing experiments in a human-in-the-loop setup.

Acknowledgments

Research reported was supported in part by the project *Safety4Bikes* funded by the German Federal Ministry of Education and Research (BMBF) under grant number 16SV7672.

References

- [Bu18] Buse, D. S.; Schettler, M.; Kothe, N.; Reinold, P.; Sommer, C.; Dressler, F.: Bridging Worlds: Integrating Hardware-in-the-Loop Testing with Large-Scale VANET Simulation. In: 14th IEEE/IFIP Conference on Wireless On demand Network Systems and Services (WONS 2018). IEEE, Isola 2000, France, pp. 33–36, Feb. 2018.
- [Bu90] Bundesen, C.: A Theory of Visual Attention. *Psychological Review* 97/4, pp. 523–547, Oct. 1990.
- [EC18] EC: Road safety in the European Union, Report MI-AC-18-001-EN-N, Luxembourg City, Luxembourg: European Commission, Apr. 2018.
- [He19] Heinovski, J.; Stratmann, L.; Buse, D. S.; Klingler, F.; Franke, M.; Oczko, M.-C. H.; Sommer, C.; Scharlau, I.; Dressler, F.: Modeling Cycling Behavior to Improve Bicyclists' Safety at Intersections – A Networking Perspective. In: 20th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2019). IEEE, Washington, D.C., June 2019.
- [HG14] Hoffman, M. D.; Gelman, A.: The No-U-turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15, pp. 1593–1623, Jan. 2014.

- [Kw01] Kwon, D.-S.; Yang, G.-H.; Lee, C.-W.; Shin, J.-C.; Park, Y.; Jung, B.; Lee, D. Y.; Lee, K.; Han, S.-H.; Yoo, B.-H.; Wohn, K.-Y.; Ahn, J.-H.: KAIST Interactive Bicycle Simulator. In: IEEE International Conference on Robotics and Automation (ICRA 2001). IEEE, Seoul, South Korea, pp. 2313–2318, May 2001.
- [Ma18] Matviienko, A.; Ananthanarayan, S.; Borojeni, S. S.; Feld, Y.; Heuten, W.; Boll, S.: Augmenting Bicycles and Helmets with Multimodal Warnings for Children. In: 20th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2018). ACM, Barcelona, Spain, Sept. 2018.
- [PKC04] Plumert, J. M.; Kearney, J. K.; Cremer, J. F.: Children’s Perception of Gap Affordances: Bicycling Across Traffic-Filled Intersections in an Immersive Virtual Environment. *Child Development* 75/4, pp. 1243–1253, 2004.
- [RE18] Rheinberg, F.; Engeser, S.: Intrinsic Motivation and Flow. In (Heckhausen, J.; Heckhausen, H., eds.): *Motivation and Action*. Springer, Cham, pp. 579–622, 2018.
- [RVE03] Rheinberg, F.; Vollmeyer, R.; Engeser, S.: Die Erfassung des Flow-Erlebens. In (Stiensmeier-Pelster, J.; Rheinberg, F., eds.): *Diagnostik von Motivation und Selbstkonzept*. Hogrefe, pp. 261–279, 2003.
- [RVM05] Rheinberg, F.; Vollmeyer, R.; Manig, Y.: Flow-Erleben: Untersuchungen zu einem populären, aber unterspezifizierten Konstrukt, Abschlussbericht, Psychologisches Institut, Oct. 2005.
- [SDJ03] Strayer, D. L.; Drews, F. A.; Johnston, W. A.: Cell Phone-Induced Failures of Visual Attention During Simulated Driving. *Journal of Experimental Psychology: Applied* 9/1, pp. 23–32, 2003.
- [SFR01] Schubert, T.; Friedmann, F.; Regenbrecht, H.: The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10/3, pp. 266–281, 2001.
- [SWF16] Salvatier, J.; Wiecki, T. V.; Fonnesbeck, C.: Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* 2, e55, Apr. 2016.
- [TPS15] Tünnermann, J.; Petersen, A.; Scharlau, I.: Does attention speed up processing? Decreases and increases of processing rates in visual prior entry. *Journal of Vision* 15/1, Mar. 2015.
- [Tü16] Tünnermann, J.: On the Origin of Visual Temporal-Order Perception by Means of Attentional Selection, PhD Thesis, Paderborn, Germany: Faculty of Arts and Humanities, July 2016.
- [V115] Vlakveld, W. P.; Twisk, D.; Christoph, M.; Boele, M.; Sikkema, R.; Remy, R.; Schwab, A. L.: Speed choice and mental workload of elderly cyclists on e-bikes in simple and complex traffic situations: A field experiment. *Accident Analysis & Prevention* 74, pp. 97–106, Jan. 2015.