Do you bike virtually safe? An explorative VR study assessing the safety of bicycle infrastructure

Marc Schwarzkopf, André Dettmann, Jonas Trezl, & Angelika C. Bullinger
Chemnitz University of Technology,

Abstract

Driven by the mobility transition towards a more ecological modal split, bicycles are becoming more popular as a means of transportation in cities. Therefore, bicycle infrastructure should become an increasing focus of urban planners. When designing infrastructure measures for cyclists, user acceptance, especially subjective safety, and comfort experience, are important reasons for the usage. To evaluate such factors in advance and derive the corresponding design requirements for urban planners at an early stage, the use of virtual reality (VR) can help to evaluate planned infrastructure measures. This paper presents an experimental design to evaluate infrastructure measures for cyclists in a VR study with 20 participants in an urban context. Subjects were presented 19 infrastructure measures in VR which were previously evaluated by an expert focus group for objective safety and divided into three safety categories. The images were randomized, and subjects were asked in a structured think-aloud procedure to provide statements about the subjective assessment, as well as reasons for their decision. In this paper, we present the study design and the results regarding reasons for or against specific infrastructure measures and will conclude with a methodological discussion regarding infrastructure assessment via virtual reality to aid urban planners and authorities.

Background

In the wake of societal and technological trends such as demographic change (Buffel & Phillipson, 2012) and advancing digitization (Kramers et al., 2014), almost all areas of daily life are changing. For the future design of cities, this means adjustments in urban development as well as the (re)design of inner-city mobility (Loorbach & Shiroyama 2006; Burns, 2013). In this context, influencing the mobility behaviour of citizens has been the focus of many research projects for some time. This was amongst others done through gamification (Torres-Toukoumidis et al., 2022), financial (cash credits; Thøgersen, 2009; Bamberg & Schmidt, 2001) or political marketing measures to increase the frequency of use of alternative modes of transport (walking, cycling, public transport).

All measures used to increase the attractiveness of alternative means of transportation, can generally be divided into hard and soft policies (Gärling et. al, 2009). Soft policies

In D. de Waard, S.H. Fairclough, K.A. Brookhuis, D. Manzey, L. Onnasch, A. Naumann, R. Wiczorek, F. Di Nocera, S. Röttger, and A. Toffetti (Eds.) (2022). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2022 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

include strategies that focus on increasing information access, increasing motivation and (digital) communication aspects. Hard measures attempt to influence the use of transport modes through physical changes to infrastructure, increased costs for car use, or control of road space. Regarding soft policies, research indicates a mostly short-term change in behaviour (Bamberg & Schmidt, 2001) while hard policies suggest that urban design measures have a higher potential to stimulate a long-term behavioural change. Wardman, Tight and Page (2007) studied factors influencing cycling behaviour and found that cycle lanes separated from the road can increase the frequency of cycling trips by 55% and lead to a slight decrease in car commuting. Here, when designing infrastructure measures for cyclists, user acceptance, especially subjective safety, and comfort experience, are an important reason for the usage (Springer et al., 2021). Besides safety and comfort, a high satisfaction with the implemented infrastructure is another important indicator for the frequency of use. For example, a correlation could be found between satisfaction with the availability of parking spaces for bicycles and actual bicycle use (Martens, 2007).

Such evaluations of bicycle infrastructure and their effectiveness on safety, comfort and overall satisfaction are mostly carried out via onsite field observations of road, traffic junctions and the surrounding areas. Problematic in this regard is the often resource-intensive effort to assess the infrastructure as well as accompanying circumstances such as daytime, weather, season, traffic situation and actual access to participants for interviewing purposes. Moreover, only a momentary observation of the initial structural condition and situation is possible. While this assessment of infrastructure measures in their current state is very important for urban planners, future improvements also need to be assessed with regard to safety and comfort. Here, the difficulties of a beforehand assessment or a longitudinal study design is difficult due to the complex, resource-intensive and lengthy implementation of new measures.

To overcome both difficulties in assessing bicycle infrastructure we propose the usage of virtual reality studies. Based on the findings of and Higuera-Trujillo et al. (2017), who compared different display formats such as photographs, 360° panoramas, and virtual reality to a physical environment, the usage of 3D technologies are the most promising for usage in controlled conditions. The visual evaluation of urban spaces is easily achievable in terms of accessible technology for capturing or creation the necessary 360° images and presenting them to participants under laboratory conditions. Following the studies of Mouratidis & Hassan (2020), who examinded architecture and urban design in virtual environments, this approach seems useful as VR proved to be useful for assessment of public spaces. In addition, cyber sickness appeared to be unproblematic, making this method applicable to a wide range of people. In the following section, we present an explorative study to examine infrastructure measures for cyclists in terms of safety and comfort in a VR environment. We try to answer the question if the study design is suitable for this kind of evaluation and if the subject's responds regarding reasons for or against specific infrastructure measures are valid. We will conclude with a methodological discussion regarding infrastructure assessment via virtual reality to aid urban planners and authorities.

Method

Participants

N = 20 subjects participated in the study. The subjects were randomly recruited via a mailing list and various student groups in messenger services. The only restriction was that subjects had to own a bicycle. Participation was remunerated with 15 €. The mean age was 29.7 years (*SD* = 7.4). 14 subjects reported using a bicycle at least once a week, no subject reported never cycling. Bicycle use was mainly for leisure activities (*n* = 17) and commuting to work (*n* = 14). 11 respondents describe themselves as experienced cyclists, 9 respondents would rather describe themselves as little or not at all experienced cyclists. According to the ATI technology affinity scale (Franke et al., 2019), five respondents are not technology affine, 13 respondents have a low technology affinity and one respondent has a high technology affinity. 14 subjects had already participated in another VR study.

Material

The 360° recordings of the infrastructure measures were made with a GoPro Max in the cities of Chemnitz and Dresden in Germany. The images were then evaluated in a focus group of four experts from the fields of infrastructure planning and bicycle safety and divided into three groups (low, medium and high subjective safety) using 2D normalised pictures. For the evaluation of the 360° images, 19 images were presented (2D normalised examples can be seen in Figure 1). To present the images in a virtual environment a Unity application was programmed. Unity v2019.4.12f1 was used in combination with the SteamVR asset to implement the app on the HTC Vive Pro and the required head tracking. The used 360° photos were declared as skyboxes in Unity, as these are rendered around the entire scene, giving the impression of a complex landscape on the horizon. To control the experimental sequence within the application, a script was written in C#. In addition, a questionnaire was created that included questions about demographics, cycling experience and the ATI (affinity for technology interaction) scale (Franke et al., 2019).

Figure 1. 2D normalised sample images for the three safety categories

Procedure

The method was based on the findings of Mouratidis & Hassan (2020) and Higuera-Trujillo et al. (2017). Each study run had a time window of only 60 minutes for health protection reasons. At the beginning of the study, an initial questionnaire was completed. Afterwards, the VR headset was put on the subject by the experimenter. In a demo image, the subject was able to adjust the headset and familiarise him- or herself with the environment and the controls. Afterwards, a total of 19 further 360° photos were presented in a randomized order. For each photo, subjects were asked by the experimenter to assess whether s/he would consider the presented infrastructure to be subjectively safe for cyclists. In addition, s/he was asked to give reasons for her/his assessment. A prepared think-aloud protocol was used (Smelser & Baltes, 2001). Furthermore, the test person was asked whether s/he knew the infrastructure. At the end of each image, the test person was asked to rate the subjective security on a Likert scale of 1 (low subjective security) and 5 (high subjective security). During the study, subjects were asked about their well-being. All statements were recorded and transcribed. The experimenter was able to follow the subject's actions and movements on his computer (Figure 2). The evaluation of an image was not subject to a time limit. The experimenter was responsible for presenting the next picture.

Figure 2. Setup during the experiment

Results

Since, to the authors' knowledge, there is no comparable study, an essential issue is the validity of the study design. This chapter is divided into two parts. In the first part, the statements and data of the test persons are considered under the aspects of validity and reliability. In the second part, the evaluation of the infrastructure measures by the test persons is analysed.

Validity and reliability

The experimental design presented is a mixed-method approach. However, the qualitative contents predominate. For this reason, no quantitative criteria can be used to assess validity. However, indications for the validity of the design can be derived from the statements of the test persons. Four hypotheses were formed for this purpose:

H1: The interindividual difference in the evaluation of subjectively perceived safety between the subjects is small ($SD \leq 1$).

H2: Subjects use the same criteria to evaluate subjective safety.

H3: The affinity for technology has no significant influence on the evaluation of subjective safety.

H4: Subjective self-assessment of cycling experience has no influence on the evaluation of subjective safety.

The subjects' evaluation of subjective safety has a low standard deviation (Fig. 3). The maximum value of the standard deviation is 1.04 and the minimum value is 0.54.

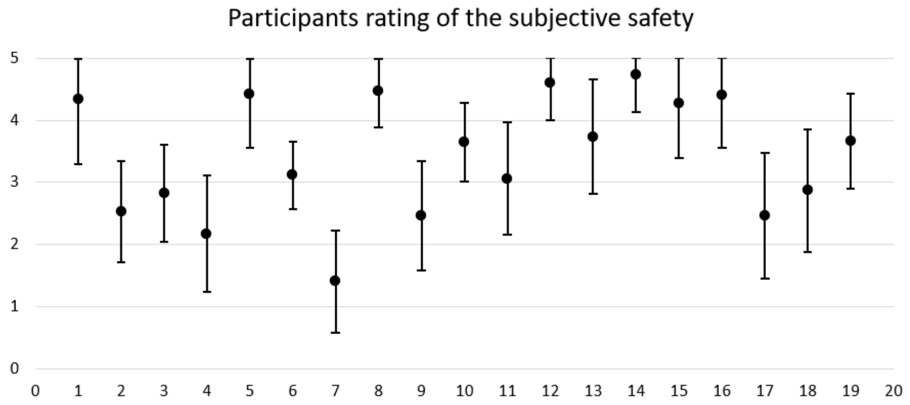

Figure. 3. Evaluation of the subjective safety of infrastructure measures for cyclists; error bars reflect Standard Error (Rating; Scale: 1 – not safe at all; 2 – not safe; 3 – neutral; 4 – safe; 5 – very safe, numbers on the x-axis represent the images shown to the particpants)

The think-aloud protocol is used to identify the strategies or heuristics used by the subjects. (Smelser & Baltes, 2001). Consistent statements by the test persons indicate comparable strategies or approaches to solving a problem (Smelser & Baltes, 2001). Applied to the VR study, this means that the use of the same or comparable criteria for assessing subjective safety across all subjects is an indication of high validity of the test procedure. The assessment in VR would therefore be carried out across all subjects using a comparable set of criteria. To verify this, the think-aloud protocols were evaluated using qualitative content analysis according to Mayring (Mayring & Fenzl, 2019) and the statements made were clustered according to the underlying criteria. In addition, the images were categorised into three categories based on the subjects' subjective safety ratings: high subjective safety (subjective safety rating ≥ 3.8, n_{images} = 6), medium subjective safety (subjective safety rating 2.26 to 3.7, n_{images} = 8) and low subjective safety (subjective safety rating ≤ 2.25, n_{images} = 5). The details are presented in Table 1.

Table 1. Number of frequencies of mentioning the evaluation criteria according to the evaluation of subjective safety. Numbers in brackets indicate the mean value per criteria per user for each image (but just for images where the criteria is visible, e. g. the mean value for parking cars was only calculated for pictures with parking cars).

	High	*Medium*	*Low*
Separation	78 (13)	79 (11,3)	73 (14,6)
Width	41 (6,83)	49 (7)	25 (5)
Signage/marking	51 (8,5)	44 (6.3)	52 (10,4)
Condition of the surface	19 (3,17)	32 (4,6)	34 (6,8)
Clarity	15 (2.5)	27 (4,5)	30 (6)
Parking cars	3 (3)	25 (12,5)	36 (12)
Other	6 (1)	4 (2)	7 (1)

Another potential factor influencing the validity of the experimental design is the affinity for technology. This was determined using ATI (Franke et al., 2019). The sample was divided into two groups (low vs. medium affinity for technology). Four test persons had a low affinity for technology. No difference could be determined for the factor technology affinity on the evaluation of subjective safety (Mann-Whitney U-test, two tailed, $p < 0.05$). The details are presented in figure 4.

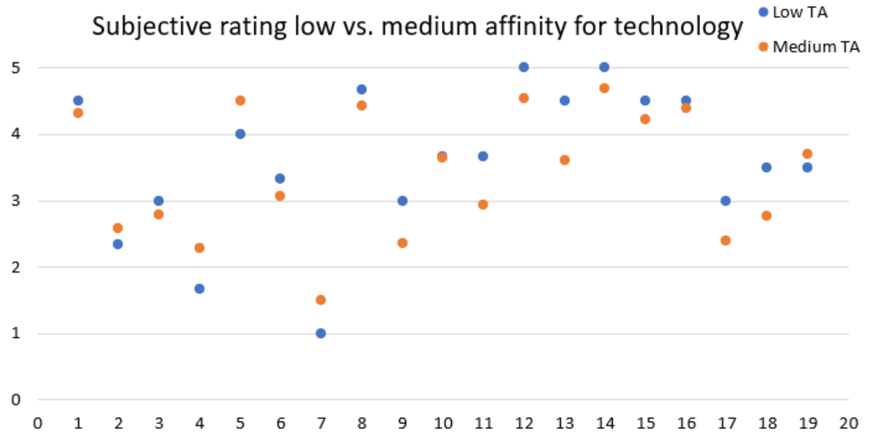

Figure. 4. Evaluation of subjective safety grouped by technology affinity (TA) (Rating; Scale: 1 – not safe at all; 2 – not safe; 3 – neutral; 4 – safe; 5 – very safe)

Another potential factor influencing the validity of the experimental design is the cycling experience of the subjects. The sample was divided into two groups (low vs. high experience). Seven test persons had low experience. No difference could be determined for the influence factor experience on the evaluation of subjective safety (Mann-Whitney U-test, two tailed, $p < 0.05$). The details are presented in figure 5.

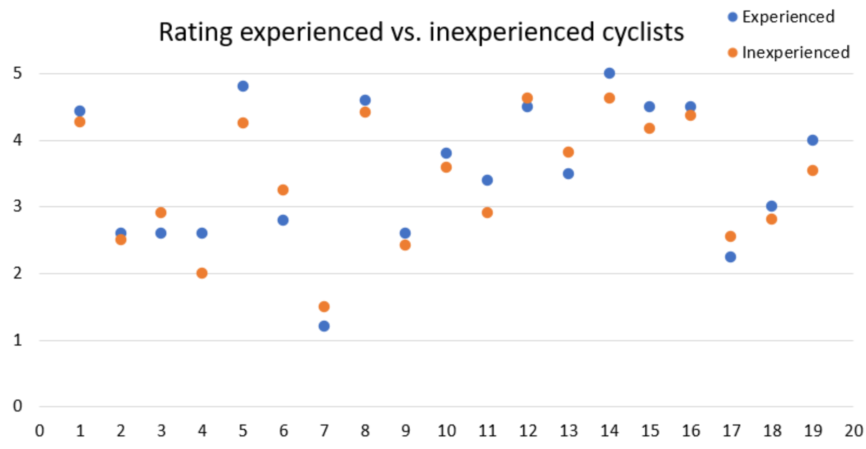

Figure. 5. Evaluation of subjective safety grouped by level of cycling experience (Rating; Scale: 1 – not safe at all; 2 – not safe; 3 – neutral; 4 – safe; 5 – very safe)

Furthermore, the subjects were asked about the ease of use of the VR headset and whether there were factors that had a (perceptible) technical influence on the evaluation of the infrastructure measures. All subjects found the implementation intuitive and stated that there were no perceptible technical factors that interfered with the evaluation. The investigators were also unable to observe any operating errors in the subjects' use of the VR peripherals. In addition, the subjects were also regularly asked about their well-being; no subject complained of nausea or dizziness.

The greatest influence on the evaluation of the subjective safety of the infrastructure measures appears to be the factor of familiarity. Subjects who already knew the infrastructure shown tended to include contextual influencing factors in their evaluation (e. g. density of road traffic, accessibility of the infrastructure measure, disruptive factors on the way to the infrastructure, etc.). The subjects expressed this accordingly in the think-aloud procedure and were asked by the test leaders to omit this information from their evaluation of subjective safety to make sure, that the participants only evaluate what they see on the images. We address this issue in the next iteration of the method.

Evaluation of the infrastructure

The experts' assessment of the subjective safety for cyclists of the infrastructure measures was carried out in a focus group of four people. The experts were from the field of infrastructure planning and road safety. The experts' evaluations for the sample are comparable to the evaluation by the test persons (Tab. 2).

Table 2. Comparison of the results of the expert evaluation and the test persons' evaluation of subjectively perceived safety (ER = Experts Rating; Scale: 1 – not safe at all; 2 – not safe; 3 – neutral; 4 – safe; 5 – very safe)

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
ER	H	L	M	L	H	M	L	H	M	M	M	H	M	H	H	H	L	M	M
M_P	4.33	2.42	2.8	2.15	4.47	3.17	1.4	4.52	2.45	3.72	3.05	4.61	3.77	4.66	4.17	4.47	2.27	2.83	3.66

Discussion

The data collected in the study served exploratory purposes only. The number of test persons as well as the division of the groups (experienced vs. inexperienced cyclists; technology-savvy vs. non-technology-savvy) do not create a basis for consideration from the point of view of inferential statistics. The presented study only served as a proof of concept for the study design and will be expanded in further research projects. Furthermore, due to corona restrictions, the implementation time per respondent was limited to 45 minutes. For this reason, questionnaires on presence in VR or in-depth evaluation of subjective safety and perceived comfort experience could not be used.

Although the presentation of the infrastructure measures was randomised, sequence effects could not be eliminated. The subjects tended to compare infrastructure measures with measures that had already been evaluated.

Nevertheless, the findings presented suggest that the methodology presented is suitable for evaluating different infrastructure measures for cyclists in an experiment. It was shown that the intersubjective evaluation in the presented sample has a low standard deviation. Furthermore, subjects use comparable criteria for the evaluation, regardless of individual cycling expertise, which is in line with previous findings (BMVI, 2017; Götschi et al., 2018; FixMyCity, 2021). This suggests that the use of VR peripherals does not significantly influence the underlying mental models used to evaluate subjective safety. Furthermore, the results have shown that the subjects arrive at comparable assessments of subjective safety as a focus group consisting of experts from the fields of infrastructure planning and cycling safety. The use of an additional group of experts from these fields is therefore not considered necessary for future studies. However, none of the experts were from public planning authorities, so a comparison of the subjective safety of cycling infrastructure with the objective safety defined by national regulations could not be made.

Considerations for future work

The presented methodology for evaluating the subjective safety of infrastructure for cyclists will be further developed in future projects. Besides a larger sample and a group balance (experienced vs. inexperienced, young vs. old), the technique will be further adapted. For example, the method will be expanded with images from a 3D stereoscopic camera. Furthermore, the use of 360° videos as well as auditory stimuli are planned. The methodology could also to be used for the evaluation of other infrastructure areas (e. g. footpaths, bus stops, crossings, etc.). The aim is to provide urban authorities with a methodology for the cost-effective and simple evaluation of implemented infrastructure measures, which can also be used in direct citizen participation.

References

Bamberg, S. & Schmidt, P. (2001). Theory-driven subgroup-specific evaluation of an intervention to reduce private car use. *Journal of Applied Social Psychology, 31*, 1300-1329

Buffel, T, Phillipson, C, & Scharf, T. (2012) Ageing in urban environments: Developing 'age-friendly' cities. *Critical Social Policy, 32*, 597-617. doi:10.1177/0261018311430457

Burns, L. (2013) A vision of our transport future. *Nature 497*, 181-182. https://doi.org/10.1038/497181a

BMVI (2017). Fahrrad-Monitor Deutschland 2017. *Ergebnisse einer repräsentativen Online-Befragung*.https://kreisverbaende.adfc-nrw.de/uploads/media/f ahrradmonitor -2017-ergebnisse.pdf

FixMyCity (2020). *Studie zur subjektiven Sicherheit im Radverkehr - Ergebnisse und Datensatz einer Umfrage mit über 21.000 Teilnehmenden*. https://fixmyberlin.de/research/subjektive-sicherheit

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. *International Journal of Human–Computer Interaction, 35*, 456-467.

Gärling, T., Bamberg, S., Friman, M., Fujii, S., & Richter, J. (2009). Implementation of Soft Transport Policy Measures to Reduce Private Car Use in Urban Areas.

Higuera-Trujillo, J. L., Maldonado, J.L.T., & Millán, C.L. (2017). Psychological and physiological human responses to simulated and real environments: A comparison between Photographs, 360 Panoramas, and Virtual Reality. *Applied Ergonomics, 65*, 398-409.

Kramers, A., Höjer, M., Lövehagen, N., & Wangel, J. (2014). Smart Sustainable Cities – Exploring ICT Solutions for Reduced Energy Use in Cities, *Environmental Modelling & Software*, 52-62.

Krukowicz, T., Firląg, K., Sobota, A., Kołodziej, T., & Novacko, L. (2021). The relationship between bicycle traffic and the development of bicycle infrastructure on the example of Warsaw. *Archives of Transport, 60*. 187-203. 10.5604/01.3001.0015.6930.

Loorbach, D. & Shiroyama, H. (2016): The Challenge of Sustainable Urban Development and Transforming Cities. In: *Governance of Urban Sustainability Transitions*, (pp. 3-12).

Martens, K. (2007) Promoting bike-and-ride: The Dutch experience, *Transportation Research Part A: Policy and Practice, 41* , 326-338, https://doi.org/10.1016/j.tra.2006.09.010.

Mouratidis, K. & Hassan, R. (2020). Contemporary versus traditional styles in architecture and public space: A virtual reality study with 360-degree videos. *Cities, 97*, 102499.

Mayring, P. & Fenzl, T. (2019). Qualitative inhaltsanalyse. *In Handbuch Methoden der empirischen Sozialforschung* (pp. 633-648). Springer VS, Wiesbaden.

Smelser, N. J., & Baltes, P.B. (Eds.). (2001). *International encyclopedia of the social & behavioral sciences* (Vol. 11). Amsterdam: Elsevier.

Springer, S., Kreußlein, M., Krems, J. F. (2021). Shedding Light on the Dark-Field of Cyclists' Safety Critical Events: A Feasibility Study in Germany. *Proceedings of 9th International Cycling Safety Conference ICSC 2021*. Lund, Sweden.

Thøgersen, J. (2009). Promoting public transport as a subscription service: Effects of a free month travel card. *Transport Policy, 16*, 335-343.

Torres-Toukoumidis, A., Vintimilla-Leon, D., De-Santis, A. & López-López, & P.C. (2022). Gamification in Ecology-Oriented Mobile Applications—Typologies and Purposes. *Societies, 12*. 1-12. 10.3390/soc12020042.

Wardman, M., Tight, M., & Page, M. (2007). Factors influencing the propensity to cycle to work. *Transportation Research Part A: Policy and Practice, 41*, 339-350.