Benchmarking the Use of Immersive Virtual Bike Simulators for Understanding Cyclist Behaviors

Xiang Guo, Erin Robartes, Austin Angulo, T. Donna Chen, and Arsalan Heydarian

Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, 22904; e-mail: {xg3kb@virginia.edu, emr4xb@virginia.edu, ava7gw@virginia.edu, tdchen@virginia.edu, ah6rx@virginia.edu}

ABSTRACT

Recent reports indicate that cyclist fatalities are rising. Unlike automobile driver crash and safety studies, there is very limited information and data on how different environmental or design features impact cyclists' behaviors, attention, and awareness. Real world studies evaluating cyclist behavior are limited due to their inherent safety risk; therefore, there is a need for alternate data to better inform the planning and design of roadways for all users. Immersive virtual environments (IVE) have shown to provide a realistic representation of real-world conditions; however, these tools have not been evaluated and validated for vulnerable road users, such as cyclists. The purpose of this study is to assess the use of an IVE bike simulator to study the impact of design and environmental conditions on cyclists' perceived safety and behavioral changes. By benchmarking cyclists' behaviors and perceived safety in real-life settings compared to its representative IVE bike simulation, we can validate whether these IVE simulators are realistic representations of real-world conditions. Furthermore, by connecting these environments with the latest low-cost human sensing devices, we have built a multimodal human sensing data collection system to track participants' gaze, heart rate, and head movement. The preliminary results from a six-participant pilot study indicate that our simulators are capable of replicating cyclists' speed profile, heart rate changes, and most of the head and gaze behaviors and that these measurements are sensitive to environmental changes.

INTRODUCTION

As a result of increasing numbers of single-track roadway travelers (e.g., bicycles, scooters, etc.), roadway design complexities, and varying traffic densities, bicyclist fatalities are rising, with deaths of pedal cyclists in the United States reaching 846 in 2019, close to the highest number of deaths in recorded history (National Highway Traffic Safety Administration 2020). Many contextual factors such as roadway design, physiological states, ambient lighting/noise level, traffic density, and cycling workload (uphill/downhill physical effort) can significantly impact cyclists' safety (Coyle et al. 1991; Zahabi et al. 2011).

However, there exists a lack of robust data sources on the safety and comfort of vulnerable road users, especially for cyclists (Zeile et al. 2016). In traffic safety and specifically for driver-related studies, driving simulators have been widely adopted with a range of benefits compared to on-road studies by creating a safe and controllable environment to simulate different traffic scenarios in a cost-effective way. Simulator-based methods have been applied to study drivers' behaviors, awareness, and psychophysiological states. Driving simulators have been validated with a variety of data (Wynne et al. 2019). One approach to experimental studies relies on the construction of a testbed in which participants can interact with their environment, imitating a naturalistic environment. Virtual Reality (VR) simulation is a promising approach for infrastructure evaluation that avoids the high cost of test bed construction. The benefit of Immersive Virtual Environment (IVE) experiments is to achieve high internal and ecological validity while also being cost effective, and offering complete experimental control to replicate trials (Heydarian and Becerik-Gerber 2017).

Although IVEs have a range of benefits, such tools have been mainly utilized for design improvements in indoor settings and have not been validated for transportation simulation, especially cyclists-safety studies. In the past few years, there has been an increasing number of studies on the application of IVE for transportation simulation. In order for the results of IVE studies to be meaningful, it is essential that the cyclists' behavior between the real world and IVE is consistent to a certain degree (O'Hern et al. 2017). Higher fidelity of the IVE simulator usually can provide a more realistic experience, but this does not necessarily translate to a greater ability to replicate the specific task or behavior of the users (Wynne et al. 2019). Therefore, IVE simulators need to be validated against a set of key performance measures to assess the correlation between results (Wang et al. 2010). Traditionally IVE simulator validations studies have relied on cycling performance such as speed, lane keeping and deviation in lateral position, while the effectiveness of physiological sensing has not been validated. For example, cycling behavior and risk perception using behavior were significantly different between the non-immersive and immersive scenarios (Bogacz et al. 2020). Another study assessed the speed, lane position and speed reduction on approach to intersections of cyclists both in IVE and on-road. Apart from subjective responses and direct cycling performance (e.g., speed, lane position), objective approaches such as gaze variability, head/body movement, and heart rate variability can be used to assess safety of cyclists at different contextual settings. Furthermore, very few studies have collected multimodal physiological data of cyclists in naturalistic experiments (Zeile et al. 2016).

The goal of this study is to benchmark and validate the use of an immersive virtual bike simulator to evaluate cyclist behaviors in a naturalistic on-road experiment with its representative IVE. Through experimental studies we can identify if there exist any significant differences in cyclist performance metrics between the two types of environments. To achieve this, through integrating the latest low-cost human sensing devices, we have built a multimodal human sensing data collection system to track cyclists' gaze, heart rate, pose, and head movement to better contextualize cyclist performance and behaviors. With the preliminary results from a pilot study, we can get insights about which measurement is consistent between IVE and real road environments.

METHODOLOGY

Experimental design. To achieve the identified research objectives, we conducted a pilot study in which we benchmarked participants' cycling behaviors and performance in a real-life environment and its corresponding IVE setting. The benchmark study has a within-subjects design in order to control for variance between subjects. The chosen corridor for this study was Water Street between 2nd St SW and 4th St SE in the city of Charlottesville, Virginia (Figure 1). Water Street is well-trafficked by bicyclists, and has been identified as high risk for vulnerable road users, and is being considered for redesign. The section of the corridor chosen for this experiment consists of two city blocks, with a 4% downhill grade in the eastbound direction, shared lane markings in both directions, a traffic signal at the intersection of East Water Street and 2nd Street SE, and a parking lane in the westbound direction.

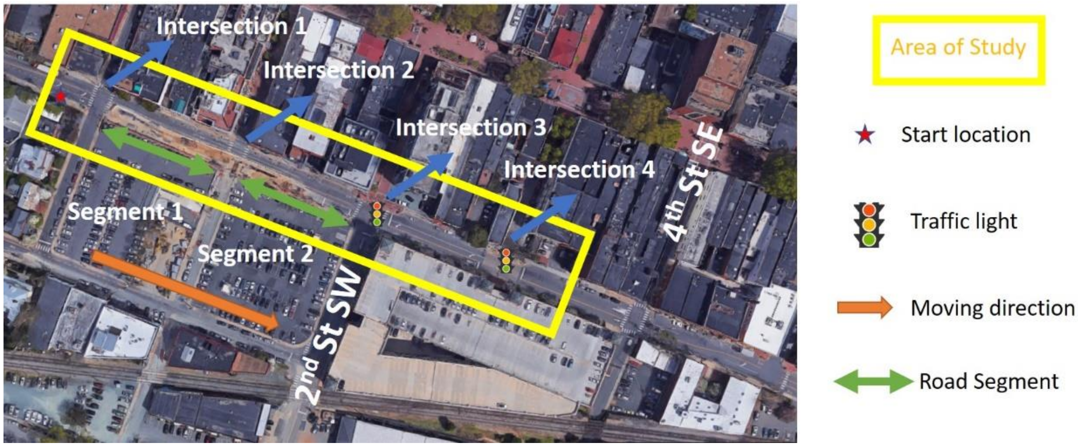

Figure 1. Illustration of experimental area

In order to collect data about existing operating characteristics of the chosen corridor, video footage was collected along the corridor for two weeks. With permission from the City of Charlottesville, four cameras were set up along the selected Water Street corridor and video footage was recorded from August 27th to August 29th and September 3rd to September 5th in 2019, resulting in 144 hours of video recording. Peak traffic hours (7:00 - 9:00 AM and 4:00 - 6:00 PM) of the video footage were reviewed, and corresponding traffic volumes were recorded. These peak traffic volumes were used to determine the traffic flow in the design of the corresponding IVE settings. **Error! Reference source not found.** A-1 shows the google maps view of water street and Figure 2 A-2 is the corresponding IVE setting.

Figure 2. Comparison of real road(A-1) and IVE (A-2) environments; Video collection system of IVE bike simulator(A3-A4); Heart Rate Distribution of experiment(B-1)

Data Collection. The data collection framework and system architecture of measuring cyclists' behaviors and physiological sensing is shown in Figure 3. For the IVE bike simulator, HTC Vive Pro Eye headsets with their accompanying controllers have been selected for tracking interaction in the IVE. This VR headset has an integrated Tobii Pro eye tracker with movement tracing capabilities, which works seamlessly with the headset as compared to traditional eye tracking systems (either screen-based or eye tracking glasses). The spatial location of the controllers (attached to the handlebars) allows the system to detect turning or braking.

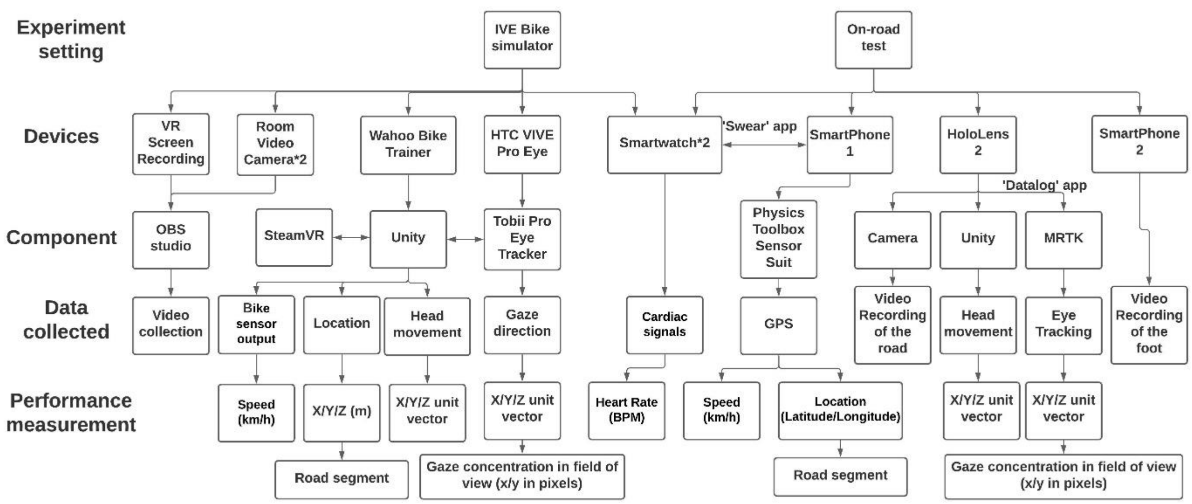

Figure 3. System Architecture and Data Framework

The VR environments were designed and programmed in Unity, and run through the SteamVR platform. With Unity C# script, all the movement of headset and controllers were tracked and extracted. An instrumented Trek bicycle was implemented for both experiments. The Wahoo indoor bicycling training equipment was chosen to connect the physical bicycle to Unity, allowing us to collect the data necessary for this research study (real time speed, instantaneous power, distance travelled) as well as provide haptic feedback to the participant. Additionally, two external video recording devices captured participants' movements during the IVE experiments. These recordings were used to monitor participants and understand their movements and reactions. What participants see in the IVE was also recorded with OBS studio software, which can integrate all videos with a uniform timestamp and frames per second. Furthermore, two android smartwatches equipped with the "SWEAR" app were utilized to track participants' physiological signals such as arm movement and heart rate (Boukhechba and Barnes 2020). More details of IVE can be found on our website wiki (Chen et al. 2021).

For the on-road test, there were two different sensors from the ones used in the IVE experiments: (1) instead of the HTC Vive headsets a Hololens 2 headset was used and (2) an additional Android smartphone app - "Physics Toolbox Sensor Suite" was used to track participants' cycling performance. The Hololens 2 is a pair of mixed reality smart glasses with eye tracking features developed and manufactured by Microsoft. The authors developed a Hololens app based on MRTK SDK called 'datalog' to collect the eye tracking and head movement data. Additionally, the Physics Toolbox Sensor Suite app has the ability to collect GPS, acceleration, sound, lighting and many other environmental factors.

Experiment procedure. As a pilot study, a total number of six participants **(**mean age = 30.3, SD **=** 3.3**)** were recruited for the experiment. All participants are 18 or older, without color blindness, and familiar with the chosen corridor.

For the IVE bike simulator experiment, after arriving at the lab space, the participants were asked to wear the Android smartwatches and fill out a pre-experiment survey about demographic information. Before the formal experiment, the participants were placed into a training scenario to familiarize themselves with the virtual environment, navigation, as well as the calibration for eye tracker and bike simulator. The experimental task was cycling to the end of experimental area, as indicated in Figure 1. After the experiment, the participants were asked to sign up for the on-road study to take place in a few weeks after. The IVE experiments took approximately half an hour to complete.

For on-road experiments, the same bicycle model was utilized, instrumented with sensors to collect a range of variables, as indicated in Figure 3. All on-road tests were conducted on clear weather days, during peak traffic hours. The same steps as the IVE experiments were followed for the on-road study.

Performance Metrics and Analysis. To study the validity of the IVE bike simulator in different contextual settings, all the performance data are calculated separately for different road segments. As shown in Figure 1, road segment 1 is the area between intersection 1 and 2, it includes a 4% downhill grade with a wall on the right to block the parking lot. Road segment 2 is a level road, and the parking lot is visible to the cyclists.

As shown in Figure 3, the following metrics are measured as indicators of participants' performance: speed of the cyclist (km/h), head movements (three-dimensional unit vector), gaze direction (gaze focus on current field of view) and heart rate (beat per minute, BPM). All performance measurements include average and standard deviation (SD) in different road segments. Performance data from the on-road and simulator experiments were compared to assess the relative and absolute validity of the simulator. Validity refers to the ability of a simulator to accurately represent real-world driving (Wynne et al. 2019). There are two major forms of validity: absolute validity (direct value comparison of simulator and on-road testing) and relative validity (same patterns or effects are observed even if the study failed to establish absolute validity). In this study, absolute validity for on-road and simulator data was assessed using paired sample t-tests at a level of significance of 0.05. If absolute validity is failed to establish, Pearson's correlation between the two settings is used to verify the relative validity.

RESULTS AND DISCUSSION

Among the six participants, one participant was missing the eye tracking data and another participant was missing heart rate data for the IVE experiment. The corresponding data was excluded from reporting. Table 1 shows all validity results.

Overall, the on-road experiment has a higher average speed and standard deviation than the IVE experiment, however, there is no significant difference in the average speed across both road segment 1 ($p=0.09$) and segment 2 ($p=0.23$); the difference in standard deviation of speed is significant for both segments ($p1 = 0.04$, $p2 = 0.002$), and Pearson correlations are positive for both, while only segment 1 is significant (0.865, $p = 0.026$), which means relative validity is only achieved for road segment 1. After the downhill road segment, as the speed is higher, participants in real-road tend to have more control on the speed, resulting in a larger variance in the SD of their speed. This can be explained by the difficulty in modeling all real-world elements within the VR environment. The weight of the users is not considered in the IVE bike simulator either, which in the real-world would impact their acceleration and resultant speed after the downhill section. Furthermore, road friction is another important factor which can affect the SD of the speed - in real road segment 2 most participants just keep freewheeling without pedaling, while in the IVE bike simulator, they keep pedaling to maintain the current speed.

Table 1. Performance comparison of IVE simulator and on-road test

Performance	Road Segment 1		Road Segment 2	
	IVE mean (SD)	**On-road mean (SD)**	**IVE mean (SD)**	**On-road mean (SD)**
Speed(km/h)	13.87(1.16*)	18.44(3.97*)	16.31(**0.49***)	20.10(**3.20***)
left/right Head movement	0.061(0.14)	0.089(0.084)	0.066(0.13)	0.066(0.041)
up/down Head movement	**-0.0018*** (0.049)	**-0.19*** (0.050)	0.019(0.056)	-0.10(0.044)
left/right gaze (pixel)	972.06(78.94)	937.38(99.65)	961.58(79.68)	1049.22(240.83)
up/down gaze (pixel)	607.38(**34.27***)	520.61(**88.76***)	**597.97* (40.74*)**	**536.17* (76.52*)**
Mean Heart Rate(bpm)	86.19(2.06)	96.14(2.58)	91.63(1.83)	97.48(1.96)

****the difference between IVE and on-road test are significant,* ***Bold***-neither relative nor absolute validity can be established for the performance*

For head movement, all pairs t-test results are not significantly different except for the up and down head movement in road segment 1 ($p = 0.002$). Pearson correlation for the up/down head movement in road segment 1 is positive but not significant (0.689, $p = 0.20$). After checking the video recording from HoloLens 2, we found that in the real road, there are several manholes on the road (4 in segment 1 and 7 in segment 2) which are not modeled in the IVE. The presence of pedestrians was not modeled in the IVE which may also explain the significant difference between the vertical head movement. This finding indicates that on real roads, cyclists will lower their head due to their concern with manholes or other roadway conditions and contextual settings.

For gaze direction, no significant differences are found across left and right gaze directions. With respect to up and down gaze direction, only the average of up/down gaze direction in road segment 1 is not significant. Participants have significantly more up and down scanning behaviors on-road than in IVE for both segments. The average gaze center is lower in on-road tests in road segment 2. Furthermore, according to Pearson correlations, relative validity is not established for other pairs either: mean up/down gaze direction in segment 2 (-0.276, p = 0.653), the SD of up/down gaze for both segment 1 (-0.660, p = 0.226) and 2 (0.013, p = 0.984). The trend in gaze direction is similar to head movement as participants will scan up and down more frequently due to the real road complexity.

For heart rate data from the smartwatches, there is no significant difference in all comparisons, the overall distribution of HR can be seen in **Error! Reference source not found.** B-1, the distribution of IVE and real road are similar to each other.

CONCLUSION

The goal of this study is to evaluate the use of an IVE bike simulator in understanding cyclists' behavior with multiple low-cost human sensing devices. A pilot study was conducted both in an IVE and on a real road. Various sensors are applied to ensure that similar data output is obtained. Both absolute and relative validity is established across a range of cyclist performance. Results show that most of the performance measurements have absolute validity, some of the features from eye tracking, most of which are in vertical direction, could not establish either absolute or relative validity. This phenomenon may be caused by road geometry change, the appearance of other road users and hardware limitations (especially the headsets used in the study). Overall, the promising results indicate that the IVE bike simulator can be further utilized for understanding cyclists' behaviors.

In addition to the difficulty in modeling all real-world elements as discussed above, the study is also limited in: (1) The small sample size of participants. After the pilot study, an experiment with a larger number will be conducted in the near future. (2) Another limitation is that participants could not change the bike gear within the current IVE bike simulator as the current version does not support gear change; therefore, the gear is consistently in the middle level for all experiments, which might potentially affect the speed change. (3) Lastly, as psychophysiological data are highly event-based, we did not annotate or collect specific events that took place within the real-road condition and in our future studies we will consider collecting and integrating such information within the design of IVE. Further work can be done by including a wider range of age groups, validating more measurements from the sensors and exploring different locations.

ACKNOWLEDGEMENT

We would like to thank the University of Virginia School of Engineering for supporting this research project through the Research Innovation Award Seed grant. Additionally, we would like to thank our participants for participating in this study and the UVA IRB office for providing guidance on participant recruitment. Lastly, we would like to thank Arash Tavakoli for his help in smartwatch set up and data management.

REFERENCES

Bogacz, M., Hess, S., F. Choudhury, C., Calastri, C., Mushtaq, F., Awais, M., Nazemi, M., AB van Eggermond, M., and Erath, A. (2020). "Cycling in virtual reality: modelling behaviour in an immersive environment." *Transportation Letters*, Taylor & Francis, 1–15.

Boukhechba, M., and Barnes, L. E. (2020). "SWear: Sensing using WEARables. Generalized

Human Crowdsensing on Smartwatches." *2019 IEEE 11th International Conference on Applied Human Factors and Ergonomics. IEEE*.

Coyle, E. F., Feltner, M. E., Kautz, S. A., Hamilton, M. T., Montain, S. J., Baylor, A. M., Abraham, L. D., and Petrek, G. W. (1991). "Physiological and biomechanical factors associated with elite endurance cycling performance." *Medicine and science in sports and exercise*, 23(1), 93–107.

Heydarian, A., and Becerik-Gerber, B. (2017). "Use of immersive virtual environments for occupant behaviour monitoring and data collection." *Journal of Building Performance Simulation*, Taylor & Francis, 10(5–6), 484–498.

National Highway Traffic Safety Administration. (2020). "Overview of motor vehicle crashes in 2019." (December).

O'Hern, S., Oxley, J., and Stevenson, M. (2017). "Validation of a bicycle simulator for road safety research." *Accident Analysis & Prevention*, Elsevier, 100, 53–58.

Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., and Coughlin, J. F. (2010). "The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing." *Ergonomics*, Taylor & Francis, 53(3), 404–420.

Wynne, R. A., Beanland, V., and Salmon, P. M. (2019). "Systematic review of driving simulator validation studies." *Safety science*, Elsevier, 117, 138–151.

Zahabi, S. A. H., Strauss, J., Manaugh, K., and Miranda-Moreno, L. F. (2011). "Estimating potential effect of speed limits, built environment, and other factors on severity of pedestrian and cyclist injuries in crashes." *Transportation research record*, SAGE Publications Sage CA: Los Angeles, CA, 2247(1), 81–90.

Zeile, P., Resch, B., Loidl, M., Petutschnig, A., and Dörrzapf, L. (2016). "Urban Emotions and Cycling Experience–enriching traffic planning for cyclists with human sensor data." *GI_Forum*, 1(2013), 204–216.