TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Engineering and Design

Ilja T. Feldstein

Design and Validation Techniques for Highly Immersive Simulator Systems

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung eines Doktors der Ingenieurwissenschaften genehmigten Dissertation.

Vorsitz:

Prof. Dr. Veit Stefan Senner

Prüfende der Dissertation:

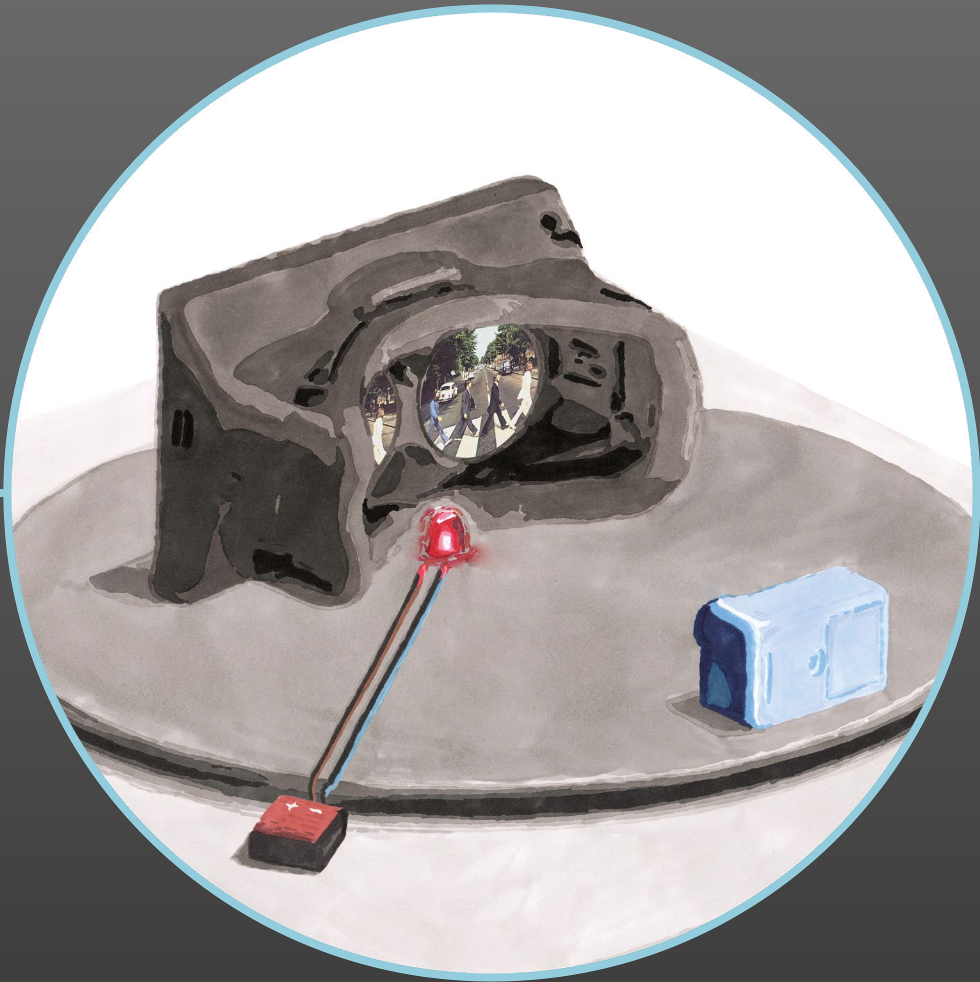
Prof. Stephen R. Ellis, Ph.D.

Prof. Dr. Marc Erich Latoschik

Die Dissertation wurde am 23.06.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 08.03.2021 angenommen.

Ilja T. Feldstein

Doctoral Dissertation in Engineering



Design and Validation Techniques for Highly Immersive Simulator Systems

Ilja T. Feldstein

Design and Validation Techniques for Highly Immersive Simulator Systems

Ilja T. Feldstein

DESIGN AND VALIDATION TECHNIQUES FOR HIGHLY IMMERSIVE SIMULATOR SYSTEMS

– Doctoral Dissertation in Engineering –

with a foreword by
Stephen R. Ellis

Doctoral Dissertation supervised by NASA scientist Stephen R. Ellis and presented at the Technical University of Munich.

The book cover shows artwork created by the author which includes, in parts, the modified Abbey Road album cover by The Beatles (with kind permission by Apple Corps Ltd. for reuse within the scope of the dissertation presented here).

 Note that the reprints presented here were revised and adjusted to the framework document in terms of writing style, format, and wording, in that American English is used and the 7th Publication Manual by the American Psychological Association (APA) is followed.

For my parents

Ulla & Igor

ABSTRACT

The technology surrounding virtual reality (VR) has made tremendous progress in the past two decades, with the recent development also benefitting scientists. In human behavior research, VR systems are steadily gaining popularity, opening up new research opportunities. Such systems allow users to become immersed in a wide range of different virtual scenarios while ensuring user safety and scenario reproducibility.

There are many different technological approaches for designing a VR system that collects and processes the user's actions (e.g., body movements) and provides the user with corresponding sensory feedback (e.g., visual and acoustic stimuli). The complex interplay of the numerous components, however, challenges the design of VR systems in particular. The processing and transition times of the different components and their interfaces may lead to significant end-to-end latencies between the users' captured action and the provided sensory feedback. These delays may interfere with the user's performance regarding the task at hand. Furthermore, excessive delays may affect the user's well-being and even result in nausea. Therefore, it is recommended to track the VR system's latency from an early design stage onward, identify possible sources of delay, and take respective measures that aim at optimizing the end-to-end processing times. The work presented here introduces a simple, yet effective latency measurement technique, managed with the most basic tools without the need to interfere with the VR setup or the simulation process.

After finalizing the VR system's design, a validation process that confirms the validity of the system is indispensable. Numerous studies carried out in the past have demonstrated that human behavior might differ between real and virtual environments. Consequently, virtually acquired data on human performance will not necessarily match corresponding real-world data. A simulator's ecological validity may be examined by comparing how participants behave in the intended virtual scenarios compared to how they behave in similar real-world scenarios. Experimental parameters for the validation study must be carefully chosen, considering that many different factors may have an impact on the human perception and the decision-making processes. The acquisition of real-environment data poses a considerable challenge due to the potentially unavailability of environments of interests, but also the participants' physical well-being must be ensured. The work presented here introduces a possible validation concept that may specifically be applied to simulator systems intended for investigating pedestrians' street-crossing.

ZUSAMMENFASSUNG

Virtuelle Realitäten (VR) und die damit verbundenen Technologien haben in den letzten zwei Jahrzehnten enorme Fortschritte erzielt. Auf dem Gebiet der menschlichen Verhaltensforschung erfreuen sich VR-Systeme stetig wachsender Beliebtheit und eröffnen dabei ein breites Spektrum an neuen Forschungsmöglichkeiten, wobei die Sicherheit der Nutzer und die Reproduzierbarkeit der Szenarien sichergestellt ist.

Entwickler solcher Systeme können auf eine Vielzahl technischer Lösungskonzepte zurückgreifen, um Nutzerhandlungen (z. B. Körperbewegungen) zu erfassen und daraus resultierende virtuelle Rückkopplungen (z. B. visuelle und akustische Reize) zu erzeugen. Das komplexe Zusammenspiel zahlreicher Systemkomponenten stellt dabei eine erhebliche Herausforderung dar: Die Rechen- und Transmissionszeiten der unterschiedlichen Komponenten sowie deren Schnittstellen verursachen unter Umständen eine erhebliche Systemlatenz. Diese Verzögerungen können sich negativ auf die Aufgabendurchführung des Nutzers auswirken und eine übermäßige Latenz kann sogar das Wohlbefinden des Nutzers erheblich beeinträchtigten. Daher sollte Latenz in VR-Systemen bereits von einem frühen Entwicklungsstadium an gemessen, mögliche Übeltäter innerhalb des Systems identifiziert und geeignete Gegenmaßnahmen ergriffen werden. Die hier vorliegende Arbeit beschreibt und erörtert eine simple, jedoch effektive und zuverlässige Latenzmessmethode, welche mit einfachen technischen Mitteln auskommt, ohne dabei direkt in das System oder dessen Simulationsprozesse einzugreifen.

Das VR-System muss nach Fertigstellung abschließend einem Validierungsverfahren unterzogen werden, welches die Eignung des Systems für die beabsichtigte Nutzung verifiziert. Zahlreiche Studien haben in der Vergangenheit anschaulich belegt, dass menschliche Verhaltensmuster in virtuellen Umgebungen von Verhaltensmustern in realen Umgebungen abweichen können. Die ökologische Validität eines Simulators kann durch einen Vergleich von Verhaltensmustern in virtuellen Umgebungen mit Verhaltensmustern in der echten Welt unter vergleichbaren Bedingungen überprüft werden. Dabei müssen experimentelle Parameter sorgfältig gewählt werden, da eine Vielzahl von Faktoren einen erheblichen Einfluss auf die menschliche Wahrnehmung und folglich die Entscheidungsfindung haben können. Die Erhebung von Referenzwerten in realen Umgebungen stellt in Hinblick auf die Verfügbarkeit geeigneter Umgebungen sowie die Sicherheit der Teilnehmer eine erhebliche Herausforderung dar. Die hier vorliegende Arbeit beschreibt und erörtert ein mögliches Validierungskonzept, welches für Validierungsstudien von Simulatoren für Straßenquerungen genutzt werden kann.

RÉSUMÉ

Les réalités virtuelles (VR) ainsi que les technologies associées ont connu d'énormes progrès au cours des deux dernières décennies. Les systèmes VR jouissent d'une popularité toujours croissante pour l'étude des comportements humains car ils permettent de simuler un large éventail de nouveaux scénarios, tout en garantissant la sécurité des participants et en permettant une excellente reproductibilité des situations.

Les développeurs de tels systèmes VR peuvent s'appuyer sur un grand nombre de solutions techniques pour capter les actions de l'utilisateur (par exemple les mouvements du corps) et générer un retour virtuel adéquat (par exemple des stimuli visuels et acoustiques). L'interaction complexe entre les nombreux composants du système représente un défi de taille : les temps de calcul et de transmission entre les différents composants ainsi que leurs interfaces peuvent entraîner une latence considérable. Ces retards peuvent avoir un impact négatif sur l'immersion de l'utilisateur dans le monde virtuel et sur son comportement. De plus, une latence excessive peut avoir un impact significatif sur le confort de l'utilisateur durant l'expérience. Par conséquent, la latence des systèmes VR doit être mesurée dès le début du développement du système, les sources de retards doivent être identifiées et les contre-mesures appropriées doivent être prises. Le présent travail décrit et analyse une méthode de mesure de la latence simple, mais efficace et fiable, qui utilise des moyens techniques faciles à mettre en place sans interagir directement avec le système ou le procédé de simulation.

Après la phase de développement, le système VR doit être soumis à un processus de validation qui vérifie sa capacité à réponde à la question de recherche posée. En particulier, de nombreuses études ont clairement démontré que les comportements humains peuvent être différents dans des environnements virtuels ou réels. Ainsi, la validité d'un simulateur VR peut être vérifiée en comparant les comportements des utilisateurs dans des environnements virtuels et dans le monde réel lorsqu'ils sont soumis à des scénarios identiques. Les paramètres expérimentaux doivent être choisis avec soin pour ce type d'étude. En effet, de nombreux facteurs peuvent avoir une influence significative sur la perception de l'environnement par l'utilisateur et, par conséquent, sur la prise de décision. De plus, la mesure expérimentale dans le monde réel représente un véritable défi pour les chercheurs. En effet, les scénarios étudiés sont souvent difficiles d'accès dans la pratique et peuvent représenter des risques de sécurité pour les participants. Le présent travail de recherche décrit un concept de validation de simulateur VR pour la recherche en sécurité routière permettant d'étudier le comportement de piétons dans une situation de traversée d'une rue.

CONTENTS

FOREWORD

The presocratic Greek philosopher Heraclitus famously claimed: All is change! Through this claim, he was acknowledging that human experience within the natural environment is that of constant flux. For example, it is never possible to step into the same river twice because the river itself was always different. Heraclitus was most likely referring to the continual, concrete variation within the natural environment, but his thoughts also can apply to the change presented within the virtual environments at the heart of virtual reality (VR).

Therefore, the overall design goal for these VR systems is that they provide their users with a changing experience within their corresponding virtual environments akin to the changes that they experience within familiar physical environments. The technologies that make VR possible may therefore be described as "experience machines." The design abstraction through which users experience virtual environments, possibly first described by Turing-Award winner Professor Frederick Brooks, was that they moved within the virtual environment as if they controlled a telerobotic vehicle moving within a real environment.

When users control a directly viewed, nearby physical telerobot, it generally promptly responds to their commands and they are thereafter able to react quickly to make corrections. The robot's reaction time to their commands is generally constant and initiated without significant delay. Under these circumstances, any observed spatial errors in movement can be directly related to visible geometrical misalignments between the operator's viewpoint and the control frame of reference of the robot, describable as combinations of displacements and rotations. In this situation, it is relatively easy for robot operators to understand the geometric adjustments they need to make to control such a nearby telerobot. The robot operator can literally see the rotations and translations to which they may adapt for their particular robot control viewpoint.

If the telerobot is not physically close to its users and is observed by a remote camera from an arbitrary viewpoint not well aligned with a convenient control frame of reference, coordinated control is difficult. If a communication delay is added, successful control may be very difficult, even seemingly impossible. Such control delays with respect to real telerobots are unavoidable due to constants such as the speed of light or sound.

In a virtual environment, a time delay, akin to a robot's communication delay, can also be introduced by the computational modeling of the kinematics and dynamics for the user-interface as well as the rendering of the virtual environment in response to user interaction. This time delay indeed is an aspect of change within the virtual environment but not one likely imagined by Heraclitus. Natural environments may change over time in a wide variety of physical ways, e.g., position, mass, velocity, acceleration, number of objects, colors, temperature, et cetera. But all of these properties have time histories and follow laws of physics, chemistry, biology, and the social customs and habits, all of which follow patterns known by the environment's inhabitants. This environmental knowledge allows users to anticipate future conditions by understanding current physical relationships among the observable physical actors and objects. Furthermore, by integrating observed changes over time, users can predict future environmental conditions.

In contrast to natural environments, synthetic virtual environments are not necessarily constrained to be familiar and therefore can make user interaction much harder since their dynamics may be totally unfamiliar. But, provided the environment's characteristics remain stable over time, users may, through experience, be able to adapt to the novel rules to control the positions and orientations of the objects with which they interact. Their improved interaction can be embodied in what is described as an internal user model of the controlled "virtual telerobot" and the structure and nature of its environment itself. Successful adaptation for unusual interaction dynamics from unusual viewing positions, however, requires that error feedback be prompt and interpretable.

Time delay (i.e., latency) unlike the physical properties of user interaction (e.g., position, velocity, and acceleration) does not itself provide error information for improving a user understanding of the physical model of user interaction, mainly because it is a totally flat signal that does not provide the temporal physical variation useful for developing corrections. Moreover, the presence of latency makes understanding the specific geometric or dynamic nature of a possibly transformed view of the worksite much harder to interpret.

Ilja Feldstein's following dissertation considers a latency within a virtual environment created totally within a digital computer simulation in which the simulation's computation time introduces a delay similar to communication delay. Its unique importance for the development of practical VR systems was most clearly emphasized by Professor Frederick Brooks. Though the problem of system latency in VR can now often be well-managed through efficient programming and predictive techniques (e.g., Kalman filtering), it will inevitably continue to be a concern as VR user interfaces are incorporated into widely physically separated networked systems. In such systems, significant communication delays are unavoidable, potentially substantial, and even possibly variable. But also, the growing resolutions and sampling rates increase the volume of data and required buffering, which will contribute to the overall system latency and will continue to be a challenge to VR designers. Human users needing to interact with objects within such environments will directly benefit from design improvements which can be developed as a consequence of the latency measurements discussed in Dr. Feldstein's dissertation. His dissertation also provides specific examples of how latency in VR or similar systems may be conveniently measured and managed, in particular within VR simulation for road-safety research, such as pedestrians who face a virtual crosswalk in the presence of automobile traffic.

His work provides examples of the disturbing effects of system latency and introduces simple, innovative techniques to measure it. In particular, he shows how conventional mobile phones can be simply adapted to measure latencies within VR systems and to determine their quantitative impact on user performance under various conditions.

Interestingly, difficulties with rendering latency were recognized to be especially problematic in early VR systems in the 1970s and 1980s. The rendering hardware and software at the time were just not sufficient to generate synthetic stereoscopic imagery quickly enough to make rendering latency sufficiently small to avoid a variety of significant unwanted behavioral effects (e.g., control instability, loss of coordination, and nausea). Bioengineers have understood these detrimental effects arising from delays in human-system interaction by the 1980s and consequently applied compensation techniques.

But even if latency during interaction with locally generated VR scene rendering can be finally made imperceptible due to ever faster graphics hardware, latency as a feature of interactive communication systems over long distances is an intrinsic problem in need of continued study: Application-acceptable levels of latency in such systems can be determined and countermeasures can be developed. Measuring and managing the impact of latency on user performance within distant interacting computer systems is likely to remain a continual problem and Dr. Feldstein's work is likely to be consulted well into the future. One particularly interesting aspect of latency is that it can easily change as systems are reconfigured and it consequently becomes a fundamental network characteristic that needs to be continuously tracked.

An additional contribution of Dr. Feldstein's dissertation is his proposed simulator validation technique. As he notes, the comparison of pedestrian behavior in a virtual environment with pedestrian behavior in a matched physical world can be complicated. While there is no physical risk associated, for example, with stepping off a curb into virtual traffic, stepping into physical traffic may very well lead to physical harm. He insightfully suggested that rather than using a road-crossing response to compare user behavior within the two environments, researchers could use a "step-back" response to signal excessive future risk, a measurement that ensures the participants' safety. This is not to suggest real risk should be discounted, but that in the interpretation of the display fidelity of a pedestrian simulator it is necessary to naturally measure separable design components that degrade system usability. Such measurements, in fact, require a common performance measure, equally valid within a real and a corresponding virtual environment. Dr. Feldstein's step-back method is itself an interesting new measure that can be utilized in pedestrian simulators as well as corresponding physical situations and suggests a general approach to investigating a variety of risk judgments that balance apparent risk in corresponding virtual and real environments.

Heraclitus would probably be astonished to learn how latency could place a constraint on communication of change. But he also would likely be satisfied to know that according to his assertion "All is change!", change will remain a fundamental part of the study of both the physical and the virtual universes, and some aspects of change can be easily used to predict future change. Latency unfortunately is not one of these. How can one know what will happen next, if all one knows at the moment is that nothing is happening? This is the puzzle emphatically presented by latency in a virtual world.

Oakland, March 2024

Stephen R. Ellis, Ph.D.,
NASA Head Scientist for Advanced Displays and Spatial Perception (ret.)

ACKNOWLEDGMENTS

Many people in my academic and personal surroundings have supported me in the completion of this engineering dissertation. First of all, I would like to express my most profound appreciation to my Ph.D. supervisor and mentor from NASA, Stephen R. Ellis, who permanently stimulated captivating discussions on a high scientific level and significantly contributed to the refinement of my research skills. I am also grateful for the extensive mentoring that I received from Alex R. Bowers from Harvard University and Gilles Régnier from Arts et Métiers, who never ceased to share their invaluable research experience and knowledge with me. Furthermore, I wish to express my gratitude to Christoph Gehlen and Veit Senner from the Technical University of Munich who paved the stony way of my dissertation. Additionally, my work benefitted from the exchange with many more researchers all around the globe, who kindly lent me their time, shared their collected data, and engaged in fruitful discussions for which I am grateful. I would also like to thank the various scholarship programs, namely the Fulbright Program, the Studienstiftung des Deutschen Volkes, and DAAD, for funding me while providing appealing cultural offerings in parallel.

My sincere appreciation also goes to my long-term academic companions, who also became my dear friends: Most notably, Cécile Boudot supported me immensely and showed the significance of sincere friendship for over 15 years across many different countries and research institutions. Christian Lehsing helped in creating numerous memorable moments on our shared academic paths across Germany, Portugal, France, China, Thailand, Japan, and finally, the United States. I am also thankful to Antonia S. Conti-Kufner, who never grew tired of answering my questions, be it related to linguistic subtleties of the English language or the complex world of statistics. In addition, my work was significantly supported by numerous students, who contributed immensely to the final outcome, with two of them standing out in particular: Georg N. Dyszak and Felix M. Kölsch.

I also wish to thank my family that includes my parents, Ulla and Igor, and also Danielle, Jaime, Tilla, Jonathan, Anya, and Leo for their continued understanding and support in every possible way. Last but not least, I am thankful to the unconditional support that I received over all the years by my loyal friends, Catalina Bursacovschi and Julius Scheifele, who always had my back, so that I could fully concentrate on my research.

The making of a dissertation is not a straightforward path with many obstacles to be overcome. Sometimes you succeed, sometimes you learn, or to close with the poetic words of Jules Verne: "Science, my lad, is made up of mistakes, but they are mistakes which it is useful to make, because they lead little by little to the truth."

Munich, January 2024

Ilja T. Feldstein

PRIOR PUBLICATIONS

Peer-Reviewed Publications Forming the Groundwork of the Dissertation

- in reprint order -

Feldstein, I. T., & Ellis, S. R. (2021). A simple video-based technique for measuring latency in virtual reality or teleoperation. *IEEE Transactions on Visualization and Computer Graphics*, *27*(9), 3611–3625.

- *I. T. Feldstein is the main author.*
- *S. R. Ellis (NASA Ames Research Center) participated in the revision of the manuscript as well as in meaningful discussions related thereto.*

Feldstein, I. T., Lehsing, C., Dietrich, A., & Bengler, K. (2018). Pedestrian simulators for traffic research: State of the art and future of a motion lab. *International Journal of Human Factors Modelling and Simulations*, *6*(4), 250–265.

- *I. T. Feldstein is the main author.*
- *C. Lehsing (Technical University of Munich) participated in the data acquisition and the preparation of the manuscript.*
- *A. Dietrich (Technical University of Munich) participated in the data acquisition and the preparation of the manuscript.*
- *K. Bengler (Technical University of Munich) participated in the funding acquisition.*

Feldstein, I. T. (2019). Impending collision judgment from an egocentric perspective in real and virtual environments: A review. *Perception*, *48*(9), 769–795.

- *I. T. Feldstein is the sole author.*

Feldstein, I. T., & Dyszak, G. N. (2020). Road crossing decisions in real and virtual environments: A comparative study on simulator validity. *Accident Analysis & Prevention*, *137*, Article 105356.

- *Feldstein is the main author.*
- *G. N. Dyszak (Technical University of Munich) participated in the data acquisition, analyses, as well as in meaningful discussions related thereto.*

Additional Topic-Related Peer-Reviewed Publications

- in chronological order -

Feldstein, I. T., Güntner, A., & Bengler, K. (2015). Infrared-based in-vehicle head tracking: A prototype for tracking drivers' head movements in real time. *Procedia Manufacturing*, *3*, 829–836.

Feldstein, I. T., Dietrich, A., Milinkovic, S., & Bengler, K. (2016). A pedestrian simulator for urban crossing scenarios. *IFAC-PapersOnLine*, *49*(19), 239–244.

Feldstein, I. T., & Bengler K. (2016). Validation aspects of virtual reality simulators for traffic research. In *Proceedings of the 2016 Driving Simulation Conference Europe (DSC '16)* (pp. 223–224). Paris, France.

Lehsing, C., & Feldstein, I. T. (2018). Urban interaction – Getting vulnerable road users into driving simulation. In K. Bengler, J. Drüke, S. Hoffmann, D. Manstetten, & A. Neukum (Eds.), *UR:BAN Human factors in traffic* (pp. 347–362). Springer.

Lehsing, C., Ruch, F., Kölsch, F. M., Dyszak, G. N., Haag, C., Feldstein, I. T., Savage, S. W. & Bowers, A. R. (2019). Effects of simulated mild vision loss on gaze, driving and interaction behaviors in pedestrian crossing situations. *Accident Analysis & Prevention*, *125*, 138–151.

Feldstein, I. T., & Peli, E. (2020). Pedestrians accept shorter distances to light vehicles than to dark ones when crossing the street. *Perception, 49*(5), 558–566.

Feldstein, I. T., Kölsch, F. M., & Konrad, R. (2020). Egocentric distance perception: A comparative study investigating differences between real and virtual environments. *Perception*, *49*(9), 940–967.

SUPERVISED STUDENT THESES

- in alphabetical order -

Ahmadi, A. (2017). *Inferential statistical assessment of human kinetic behavior and perception in virtual environments* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Ahmadi, A. (2018). *Human motion patterns in virtual environments: Robust data mining and advanced correlation analysis* [Master's thesis]. Technical University of Munich, Department of Mechanical Engineering, Associate Professorship of Sport Equipment and Sport Materials.

Aydemir, E. (2016). *Investigation of the just noticeable difference regarding the speed perception using a head-mounted display* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Ben Arab, A. (2016). *Integration of a markerless real-time motion capture system into a simulation software* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Bernhardt, B. (2015). *Driver assistance systems for improving visual perception in a poor visibility situation and considering an optimal human-machine interface* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Buchinger, D. (2014). *Obstacle detection and tracking using a stereo night vision system* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Demmelmair, S. (2014). *Study concerning the distraction and driving behavior by presenting information on a head-up display compared to an "emissive projection display"* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Depser, A. (2016). *Elaboration of a sound library for driving and pedestrian simulators* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Dietrich, A. (2014). *Optimization of a skeleton sled for the 2014 Winter Olympics: The sled cover* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Associate Professorship of Sport Equipment and Sport Materials.

Dietrich, A. (2014). *Proof of concept of a pedestrian simulator* [Master's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Dillmann, D. (2015). *Development of a motion capture suit with active LED markers for optical motion-capturing systems based on infrared light* [Master's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Dyszak, G. N. (2016). *Comparison of the perception of local radial motion in virtuality and reality* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Dyszak, G. N. (2018). *An evaluation of immersion and presence in virtual reality systems* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Färber, J. (2016). *Discretization of user inputs to increase the performance of an optimization software* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Fischer, E. (2015). *Traction and impact measurement regarding outdoor use of sports surface*s [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Associate Professorship of Sport Equipment and Sport Materials.

Gawronski, P. (2016). *Numerical optimization of positioning settings for optical motion capture systems* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Gerö, A. (2015). *Conversion of noise vectors for a pedestrian simulator* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Gruber, L. (2016). *Optimization of a skeleton sled for the 2018 Winter Olympics: The head bearing* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Associate Professorship of Sport Equipment and Sport Materials.

Güntner, A. (2013). *Development of a head-tracking method in passenger cars* [Master's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Haag, C. (2017). *Analysis of urban driving behavior in different simulator settings* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Hachmann, J. (2016). *Experimental investigation of human kinetic behavior in virtual reality* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Hammel, C. (2016). *Human performance in virtual road crossing scenarios* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Hänel, F. (2016). *Measurement of latency in virtual reality* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Hönick, S. (2015). *Tracking of walking motion for a pedestrian simulator* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Jeschke, V. (2014). *3D user interfaces – Overview and specific application "pedestrian simulator"* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Kalb, L. (2016). *Visual and acoustic information transfer in virtual reality on the example of a boundary warning system* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Kölsch, F. M. (2016). *Comparison of the egocentric distance perception in virtuality and reality* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Kronsteiner, T. (2013). *Optimization of a skeleton sled for the 2014 Winter Olympics: The saddle* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Associate Professorship of Sport Equipment and Sport Materials.

Lehnhoff, C. (2016). *Requirements analysis, conception, and integration of a Kinect-based motion capture system* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Leroy, A. (2015). *Study of several futsal floors with an innovative device: TrakTester* [Bachelor's thesis]. Munich University of Applied Sciences, Department of Applied Sciences and Mechatronics.

Lichtmannegger, C. (2016). *Software architecture and user interface development for numerical optimizations* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Mayer, C. (2014). *Driving simulator study to distraction by projections on the windshield* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Metzler, O. (2016). *Development of a wireless LED marker for optical motion capture systems* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Milinkovic, S. (2015). *Pedestrian simulator – Presence evaluation and validation* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Milinkovic, S. (2016). *Evaluation of multimodal warning concepts in virtual reality* [Master's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Okoh, K. (2017). *Detection and tracking of medical instruments in minimally invasive procedures using galvanomagnetic effects* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Micro Technology and Medical Device Technology.

Orthen, B. (2017). *Developing a Unity-based driving simulator setup for investigating change blindness* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Rabast, P. (2016). *Optimization of a skeleton sled for the 2018 Winter Olympics: The construction frame* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Associate Professorship of Sport Equipment and Sport Materials.

Rabast, P. (2017). *Development and design of a cost-effective skeleton sled for junior athletes* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Associate Professorship of Sport Equipment and Sport Materials.

Reng, J. (2016). *Evaluation of methods and construction of a measuring system for the data acquisition of vehicle dynamics* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Riedel, G. (2015). *Economic consideration of an energy harvesting approach for a futsal pitch* [Master's thesis]. Technical University of Munich, Department of Mechanical Engineering, Associate Professorship of Sport Equipment and Sport Materials.

Rösch, M. (2016). *Development and validation of a real-time 3D data fusion for a motion capture system using multiple Kinect sensors* [Master's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Sandbichler, C. (2016). *Benchmark on motion capture systems* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Titze, S. (2016). *3D object modeling in the simulation environment of SILAB 4.0* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Tsiamitros, N. (2016). *Interface creation in between a sound engine and a virtual environment* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Unkauf, T. (2016). *CAD and texture mapping for the 3D modeling of complex objects in virtual reality* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Wilczeck, T. (2015). *Virtual sound environment: Creation of synthetic car sound* [Bachelor's thesis]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Wolff, B. (2016). *Haptic information transfer in virtual reality on the example of a boundary warning system* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

Zillig, O. (2015). *Human performance in the virtual pedestrian simulator* [Term project]. Technical University of Munich, Department of Mechanical Engineering, Chair of Ergonomics.

ACRONYMS AND ABBREVIATIONS

Acronyms

3D	three-dimensional
ACM	Association for Computing Machinery
ADAS	advanced driver-assistance system
AI	artificial intelligence
AIAA	American Institute of Aeronautics and Astronautics
ANC	active noise canceling
ANOVA	analysis of variance
APA	American Psychological Association
BGU	Ben-Gurion University of the Negev
CAD	computer-aided design
CAVE	cave automatic virtual environment
CMOS	complementary metal-oxide-semiconductor
CPU	central processing unit
CRT	cathode-ray tube
CUDA	compute unified device architecture
DOF	degree of freedom
DPU	data processing unit
DST	deceleration-to-safety time
FAA	Federal Aviation Administration
FOR	field of regard
FOV	field of view
GPU	graphics processing unit
GtG	gray-to-gray
GUI	graphical user interface

HMD	head-mounted display
HRTF	head-related transfer function
IEEE	Institute of Electrical and Electronics Engineers
IFSTTAR	Institut français des sciences et technologies des transports, de l'aménagement et des réseaux
IPD	interpupillary distance
ITD	interaural time difference
IQR	interquartile range
JND	just-noticeable difference
laser	light amplification by stimulated emission of radiation
LCD	liquid crystal display
LED	light-emitting diode
lidar	light detection and ranging
NASA	National Aeronautics and Space Administration
OEM	original equipment manufacturer
OLED	organic light-emitting diode
PET	post-encroachment time
PQ	presence questionnaire
SPIE	The International Society for Optics and Photonics
TTA	time-to-arrival
TTC	time-to-contact
TUM	Technical University of Munich
UQO	Université de Québec en Outaouais
USB	universal serial bus
VAC	vergence-accommodation conflict
VE	virtual environment
VGA	video graphics array
VOR	vestibulo-ocular reflex

VR	virtual reality
VSync	vertical synchronization
VVM	visuo-vestibular mismatch

Physical Units

cm	centimeter
fps	frames per second
g	gram
GB	gigabyte
GHz	gigahertz
h	hour
Hz	hertz
kHz	kilohertz
km	kilometer
m	meter
mA	milliampere
MHz	megahertz
mm	millimeter
ms	milliseconds
mW	milliwatt
nm	nanometer
ppi	pixels per inch
px	pixel
rpm	rounds per minute
s	second
sr	steradian
V	volt
y/o	years old

Statistical Abbreviations

β	standardized regression coefficient
F	ratio-of-variances value
M	mean value of the sample
n	number in the sample
p	probability value
r	Pearson's correlation coefficient
R^2	coefficient of determination
SD	standard deviation
SE	standard error
t	t-test value
T	Wilcoxon-signed-rank-test value

1 INTRODUCTION

Car manufacturers and original equipment manufacturers (OEMs) have been relying on driving simulators for decades. These useful tools allow for assessing drivers' perception and behavior, which ultimately enables the systematic improvement of vehicle safety and comfort. However, new challenges in the automotive industry (e.g., autonomous vehicles) require understanding all road users and not only drivers. This includes, in particular, pedestrians, who are among the most vulnerable road users. Car manufacturers and OEMs need to design artificial intelligence (AI) for vehicles so that the software piloting vehicles will be able to recognize and process human behavior patterns. For example, pedestrians' intentions may be inferred through the accurate interpretation of their body language and movement. In addition, the question of how pedestrians will react to and interact with highly automated, potentially driverless vehicles has to be addressed. City planners are also facing challenges in managing the ever-increasing traffic volume and, consequently, may need to understand road users more accurately in order to design more efficient and intelligent traffic routing through resourceful traffic management.

The established concept of driving simulators that are used for the performance and behavior assessment of vehicle occupants may be extended to pedestrian research as well: The technology surrounding virtual reality (VR) allows for immersing the user into virtual environments, which form the basis of such performance and behavior assessments. Such systems are rapidly evolving, opening up a wide variety of new opportunities for behavioral research. The performance of VR components has taken a tremendous leap forward in the past few years, while acquisition and running costs have dropped significantly. This trend is linked to the financially strong entertainment industry, which entered the VR market about a decade ago and currently serves as the technological driver for the whole VR industry. The way for the entertainment industry was paved thanks to the emergence of inexpensive but highly advanced sensors and displays that have recently been developed for the large smartphone market, among others. This has allowed for the development of high-performance head-mounted displays (HMDs) and reliable low-latency motion-capture systems at reasonably low prices. In addition, the ever-expanding VR community supports the development of extensive content for powerful game engines that generate highly immersive virtual environments. These are the main components required in an HMD-based VR system that enables the investigation of pedestrian behavior in immersive virtual traffic scenarios.

Studies that are carried out in virtual environments have some significant advantages over field studies that either rely on empirical analyses of observations (e.g., at real urban intersections) or require experiments that are carried out in some sort of real environment. Observational studies and real-world experiments are both limited in the variety of scenario designs since they depend on the existence and accessibility of specific real settings. Observational studies will also depend on the likelihood of the situation since investigators do not control parameters such as speeds, directions, and frequencies of observed road users. In addition, observational studies frequently lack metadata about observed road users, such as age or driving experience, thus restricting possible analyses. Experimental investigations in real environments, on the other hand, are limited by safety concerns involving the participants and often require long, tedious, and costly preparations. For instance, a simple parameter modification, such as a color change of the test vehicle, calls for sourcing an entirely different car. In contrast, many parameter changes in virtual environments just take a few clicks to implement.

Virtual environments are safe for the experimental participant, offer a nearly unlimited number of possible scenarios (within the limits of the technical constraints), and guarantee the repeatability of scenarios at any given time. Furthermore, virtual scenarios can be set up and modified quickly and inexpensively. For a fair comparison though, one must consider that the effort required to design and build the VR laboratory does represent quite a financial investment and time expenditure. Furthermore, virtual environments do entail a relevant downside: User perception and behavior in virtual environments may differ from those in real life (e.g., Bhagavathula et al., 2018; de Kort et al., 2003; Feldstein, 2019; Feldstein & Dyszak, 2020; Feldstein et al., 2018, 2020; Feldstein & Peli, 2020; Fink et al., 2007; Recarte et al., 2005; Slater et al., 2000). As a result, virtually acquired data may lack ecological validity. It is crucial that VR systems undergo a validation process that compares the perception and behavior relevant to the intended studies with the perception and behavior in equivalent real environments. It is recommended to replicate a known real environment in the virtual environment for the validation study since this enables a valid comparison. Investigators have to ask which sensory stimuli are relevant to the intended simulator usage and must consequentially be investigated. Investigating every possible stimulus may be unnecessary and could require an unreasonable amount of time and resources.

The dissertation presented here deals with the development, design, and validation of a simulator that uses VR technology to immerse the user in various traffic environments. The designed system was intended to enable the investigation of perception and behavior from a pedestrian perspective in various street-crossing scenarios. It was an iterative and time-consuming process that included many difficulties, attempts, and setbacks relating to numerous aspects of the simulator's different components and interfaces.

The next section provides an in-depth technological overview of the simulator and its components. The four associated publications are to be found in the annex. The first publication deals with latency in VR systems and suggests a convenient way of tracking the latency performance of simulators throughout the design process. The second publication discusses the general motivation for pedestrian investigations, the newly developed street-crossing simulator (also referred to as pedestrian simulator), and some initial results, which were collected during some pilot studies using this specific simulator paradigm. The third publication reviews factors that may affect the perception of approaching vehicles, explicitly comparing real and virtual findings. The fourth publication suggests a validation technique that can be applied to street-crossing simulators and demonstrates some notable differences in human behavior between real-world environments and the virtual environment presented here.

2 A NEW DESIGN FOR A STREET-CROSSING SIMULATOR

This section discusses the simulator's different components that essentially consist of the virtual environment, the motion-capture system, the visual and auditory displays, as well as the cabling system for power supply and data communication.

2.1 Virtual Environment

For the new street-crossing simulator, the *Silab* software framework (WIVW, Würzburg, Germany) was selected for the implementation of the virtual environment. The decision was made in favor of this driving-simulator software to enable investigation of the interaction between pedestrians and drivers by linking the prospective street-crossing simulator to existing driving simulators that were already running on *Silab*.

Since a proper validation requires the comparison of the virtual behavior with the behavior observed in reality, the campus of the Technical University of Munich at the Department of Mechanical Engineering was replicated in the virtual environment. In a first step, the base area was recreated by laying out essential elements that are available in the software's library, such as streets, sidewalks, grass areas, cars, trees, and bushes, at respective locations, using satellite images from *Google Maps* (Google, Mountain View, CA, U.S.). In a second step, the raw shapes of buildings were recreated using CAD software applications while using *Google Maps* again

Figure 1. The real campus of the Technical University of Munich (left) and the virtual replica (right).

but also independently provided measurements and architectural data from the university. In a third step, the surfaces of respective buildings were photographed, whenever possible from a far distance, using a telephoto lens at high resolution so that artifacts due to optical and perspective distortions were minimized. The images were cropped, remaining distortions were corrected, and pixel and color characteristics were optimized, using raster graphics editors. In a fourth step, the surfaces of the CAD models were texture-mapped with the edited images, using a 3D rendering application. The completed buildings were then imported into the virtual environment. Finally, various details and objects (e.g., lanterns, benches, and fountains) were added to the environment, analogous to the process used to create the buildings, ultimately increasing the overall realism of the given virtual environment. An example of the real environment and the replicated virtual one in *Silab 4.0* is shown in Figure 1.

Here are some examples of tools that may be used to replicate a real environment: *Autodesk Inventor* (Autodesk, San Rafael, CA, U.S.) and *DS CATIA V5* (Dassault Systèmes, Vélizy-Villacoublay, France) are both suitable applications with an extensive range of tools for modeling complex shapes. Photographs may be edited with image editors such as *Adobe Photoshop* (Adobe, San Jose, CA, U.S.) or the *GIMP* freeware (The GIMP Development Team). The *Cinema 4D* software (Maxon Computer, Friedrichsdorf, Germany) and the *Blender* freeware (Blender Foundation, Amsterdam, the Netherlands) are both suitable for texture mapping and the *AnyCAD Exchange3D* freeware (AnyCAD Graphics Solution, Shanghai, China) may be used to convert CAD files.

2.2 Motion Capture

Motion-capture systems may be divided into optical systems that rely on external cameras, inertial systems that rely on miniature inertial sensors attached to the user's body, mechanical systems that rely on measuring the joint angles of exoskeletons, or magnetic systems that rely on measuring the relative magnetic flux between transmitter and receiver. These systems have different characteristics relating to precision, reliability, latency, ergonomics, the number of tracked body segments, the covered tracking space, and acquisition costs. As a result, VR designers must carefully assess and consider the requirements for the motion-capture components.

The street-crossing simulator presented here used gyroscopes, accelerometers, and magnetometers that were integrated into the HMD and which allowed for tracking rotational head movements. Tracking the user's head turns in real time is crucial to avoid visuo-proprioceptive conflicts that may result in simulator sickness (e.g., Allison et al., 2001; DiZio & Lackner, 1992). The Federal Aviation Administration (FAA) calls for maximum latencies of 150 ms for civilian simulators, while most military operators require simulators to have latencies below 100 ms (Allerton, 2009). These standards have been defined for aviation to ensure reliable pilot-in-the-loop performance in time-critical tasks, although typical scenarios in aviation could tolerate significantly higher latencies (Bailey et al., 2004). Most fields that rely on simulators are not regulated as strictly as the aviation industry. The research community that uses VR applications has not yet set any binding standards with regard to system latencies but has only shared observations and made recommendations that do vary widely. Link et al.

(2002) suggest that latencies for HMD systems that exceed 150 ms are likely not acceptable, latencies of 100 ms are marginal, and latencies below 50 ms are desirable. Bailey et al. (2004), who thoroughly analyzed the latency requirements for head-worn displays under consideration of human factors, concluded that the latency for such systems must remain below 20 ms. This conclusion is consistent with the numerous studies that strived to determine the threshold for human-detectable latency by measuring the psychometric function, ultimately deriving an average psychophysical threshold of around 15 to 20 ms (Adelstein et al., 2003; Ellis et al., 1999, 2004; Mania et al., 2004; M. J. Regan et al., 1999). That said, some sensitive participants revealed just-noticeable differences (JNDs) for detecting latencies that were significantly below the average psychophysical threshold just mentioned. Jerald (2010) thus recommended that latencies in head-referenced VR systems should preferably be below 3 ms. With the currently available technologies, such latencies are far from realistically achievable delays in complex virtual environments. Since human sensitivities to latency considerably depend on the virtual task at hand (Bailey et al., 2004), VR designers have to evaluate applications individually in order to define latency requirements. It must be clear that lower latencies often involve trade-offs in terms of lower image resolution and graphical details (Feldstein & Ellis, 2021).

For rotational and translational body movements other than rotational head movements, an optical *Vicon* motion-capture system (Vicon Motion Systems, Yarnton, U.K.) with ten *Vicon T10* cameras covering a tracking area of 4 m × 2.5 m was integrated in the street-crossing simulator presented here. It should be noted that the number of cameras would have been sufficient to cover a larger area, but the laboratory's dimensions restricted the usable area.

Vicon uses triangulation algorithms to determine the exact three-dimensional coordinates of 39 markers that are attached to the user's body. The complementary *Vicon Nexus* software translates marker combinations and their positions into individual body segments that add up to a human body shape, the *Vicon Plug-in-Gait* model. *Vicon* systems are popular in research applications that require motion capture, primarily due to their high precision, high reliability, and low latency. That said, *Vicon* systems have some technological shortcomings in addition to the high purchase costs: *Vicon* systems customarily use small spherical markers that reflect infrared light emitted by LED strobe units that are integrated into the cameras' housings. There is a substantial time needed (approximately 15 minutes per participant) to attach and calibrate the 39 markers. The markers are usually attached with double-sided adhesive tape. In consequence, abrupt and sharp movements may cause the markers to fall off. In addition to that, the markers lose their reflective surface characteristics over time, even when handled with care. Tracking reliability will also depend on camera distance and frame rate.

These drawbacks were addressed by replacing the original reflective markers with custom-made LED modules that ultimately led to improved tracking reliability as well as better usability. Sensors and optical filters implemented in *Vicon T10* cameras are best suited for infrared wavelengths from 720 nm to 940 nm (according to the manufacturer), which makes the widely available 850-nm LEDs suitable for the intended application. A selection of infrared LEDs with different beam angles, radiant intensities, and radiant fluxes in combination with a selection of diffusers comprising different shapes, sizes, and materials was tested on a custom-made test bench for a variety of different tracking angles, tracking distances, camera frame rates, and sensor intensity thresholds. In the tests, the system responded best to an 850-nm LED with a radiation half-angle of ± 65°, a minimum radiant intensity of 6.3 mW/sr, and a total radiant flux of 70 mW while being operated using a forward current of 100 mA. A tear-shaped polystyrene diffuser with a wall thickness of 2.5 mm further improved the trackability. Tests showed that the LED modules could be tracked at least twice as far as the original reflective markers due to higher luminance and hence better visibility.

LED modules also enabled reliable tracking at five times higher frame rates (up to 1 kHz) than reflective markers that were limited to 200 Hz at a four-meter distance. The frame rates are relevant to real-time applications since the frame rate of the motion-capture system will influence the end-to-end system latency (Feldstein & Ellis, 2021). The advantage of LED modules became even more evident when their performance was compared to older reflective markers that had already begun to show some wear. Unlike the reflective markers, LEDs will perform at a constant level for many years, even when used daily.

The LED markers were attached to a custom-made full-body suit. This motion-capture suit was made using a highly elastic material so that users with very different body shapes and sizes could wear it. The markers were attached following the *Vicon Plug-in-Gait* model and were usually already in place when the user put on the one-size-fits-all suit that could be worn over clothing. Hook-and-loop fasteners (also known by the trademark *Velcro*) ensured that markers could be easily repositioned on the suit if adjustments were necessary to conform to unusual body proportions. The LED markers were all connected through sewn-in cables and centrally powered though a belt holding lithium-ion polymer batteries. This motion-capture suit reduced the mounting and calibration time from fifteen minutes to under five. In addition, the hook-and-loop fasteners prevented the markers from falling off, as was the case with the previous system that relied on double-sided adhesive tape.

Another advantage of LED modules over reflective markers is the flexibility in terms of camera placement: With reflective markers, cameras need to be mounted at an elevated position and tilted downward since the built-in strobe units will otherwise interfere with the signal processing of the cameras on the opposite side. This restricts the placement and alignment of the cameras around the area to be tracked. With LED markers, the strobe units in the cameras become purposeless and can be turned off, thus enabling the placement of the cameras in lower positions while facing each other.

The positioning and orientation procedure of the *Vicon* motion-capture cameras is not straightforward and calls for some experience with the system. Except for the manufacturer's vague recommendation to distribute and orient the cameras evenly over the desired area to be tracked, there are no explicit instructions for the placement. For this reason, it was decided to develop a software application that calculates the optimal positioning of the cameras. This application, which runs under the name of *Obstacle*, provides a calculated recommendation for camera distribution and orientation based on the various parameters provided. Before discussing the software's underlying algorithms, the operating principles of the *Vicon* system are outlined in the following.

At least two cameras must be able to see a marker at the same time to obtain a three-dimensional localization of the respective marker. The position may then be calculated by using triangulation algorithms and epipolar geometry, which are applied to the two-dimensional camera images (Figure 2). In the particular case in which two cameras are positioned facing each other and the marker is located precisely on the spatial axis between the two camera sensors, a third camera is necessary for locating the marker. When the number of cameras viewing a marker at the same time increases, the likelihood of detection failing decreases, and the precision of location determination increases due to improved optical resolution (Figure 3).

The developed application for camera placement has a user-friendly graphical user interface (GUI) that allows the user to provide the system with

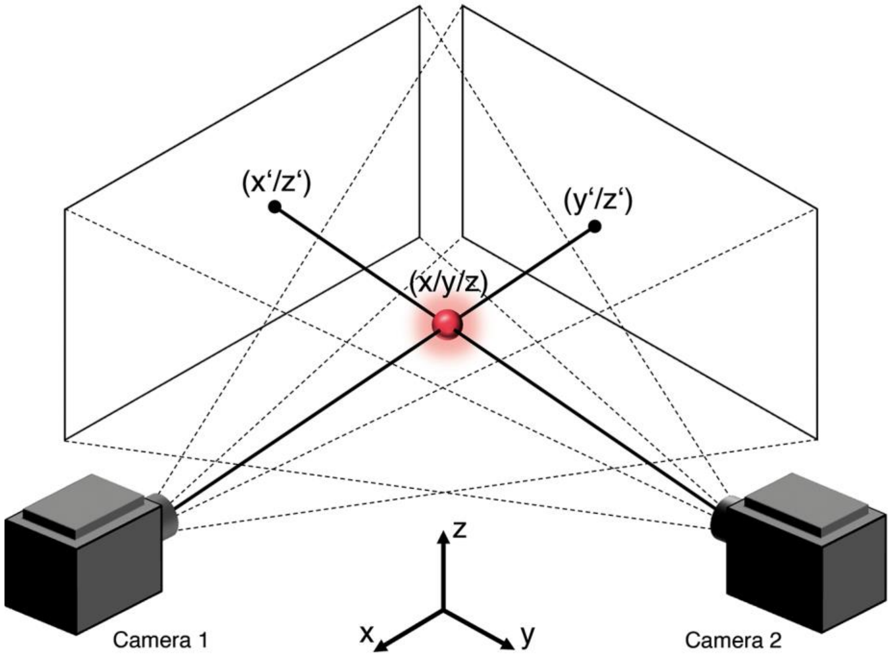


Figure 2. Epipolar geometry and triangulation of the two-dimensional camera images allow for the calculation of a marker's three-dimensional coordinates.

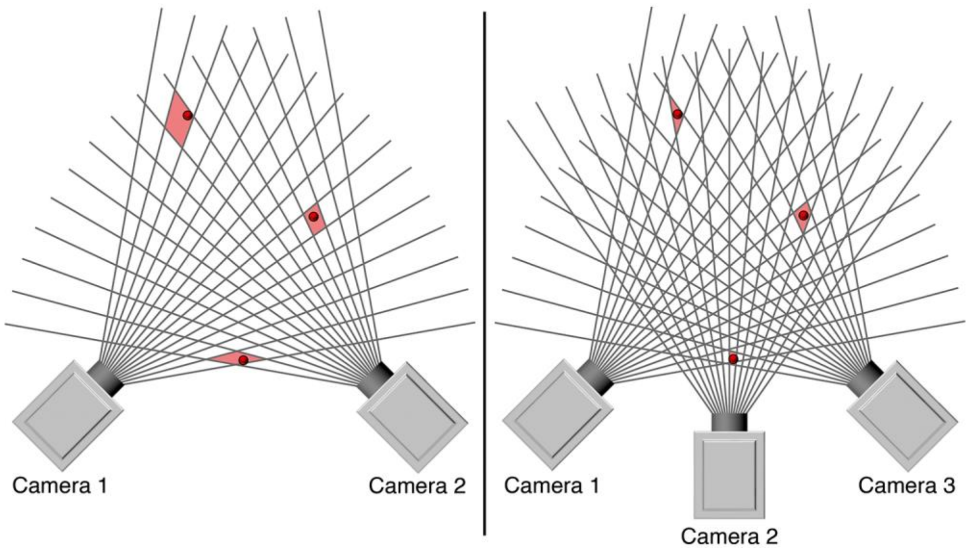


Figure 3. Localization of markers with two cameras (left) and three cameras (right): Because of the discrete camera resolution, an increasing number of cameras increases the accuracy of measurements.

information regarding the laboratory space, available camera equipment, and application parameters:

- Laboratory space: The algorithms require information about the room's height, shape, and dimensions (i.e., floor plan).
- Camera equipment: The number and model of cameras available must be specified. The application takes data from an implemented library based on the selected camera model to retrieve the dimensions and resolution of the camera sensor. Additionally, the focal lengths of the camera lenses mounted must be specified.
- Application parameters: The user must specify the (minimum) diameter of used markers and the maximum tracking height required. Finally, the calculation resolution (i.e., the precision and accuracy of the calculated result) must be selected.

Each camera captures a space that is defined by the field of view (FOV) and the maximum tracking distance of the camera (Figure 4). The application calculates the FOV based on the lens' focal length and the sensor's dimensions. The maximum tracking distance depends on the marker's diameter and the sensor's resolution. Thus, the viewing frustum of a camera corresponds to a pyramid-like space, with the far plane (i.e., the bottom of

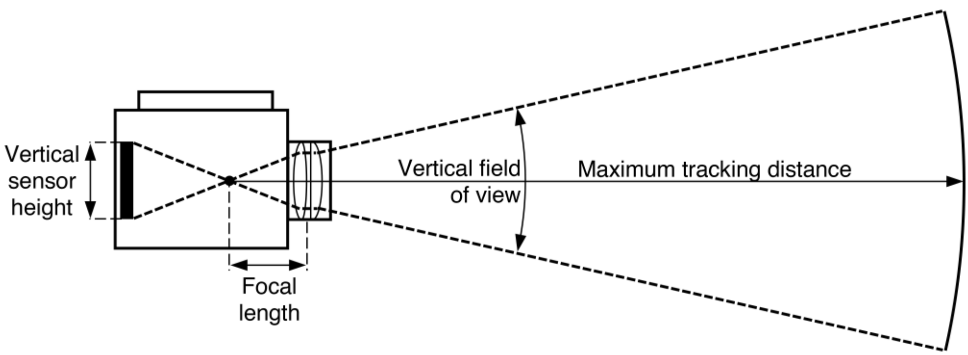


Figure 4. Relation between camera parameters and viewing frustum: The camera's field of view results from the lens' focal length and the sensor's dimensions. The maximum tracking distance depends on the sensor's resolution and the size of the tracked markers.

the pyramid) forming a convex curvature. The radius of this curvature results from the maximum tracking distance.

The application simulates the camera placement within the given laboratory space. First, the cameras are evenly distributed along the room's outline in a best-guess arrangement. The application then simulates the repositioning and realignment of each camera. The cameras' positions vary in translational directions: horizontally along the room's outline and vertically along the room's entire height. The cameras' orientations vary around the pan, tilt, and roll axes. The position and the angles are varied discretely, with only one degree of freedom (DOF) out of the two translational and three rotational ones being altered for one single camera at a time.

The program virtually projects a grid structure on the floor. The algorithms then calculate for each intersection of the grid structure the number of cameras that capture this point in the given camera configuration. Since motion capture is a three-dimensional challenge, the grid structure and the associated calculations are repeated in layers between the room's floor and the previously defined maximum tracking height. This essentially creates a virtual point cloud that fills the entire tracking space, with the algorithms determining the number of cameras that capture each point. Based on this data, the application can determine the total area that is captured by at least three cameras from the floor up to the maximum tracking height. Even if two cameras are usually sufficient, detection may fail and masking may occur, adversely affecting the tracking reliability. Consequently, the preferred objective is at least three cameras per tracking point.

The application carries out the calculation for every single camera configuration with the ultimate objective of finding the camera configuration that allows for the largest area being tracked in the room by

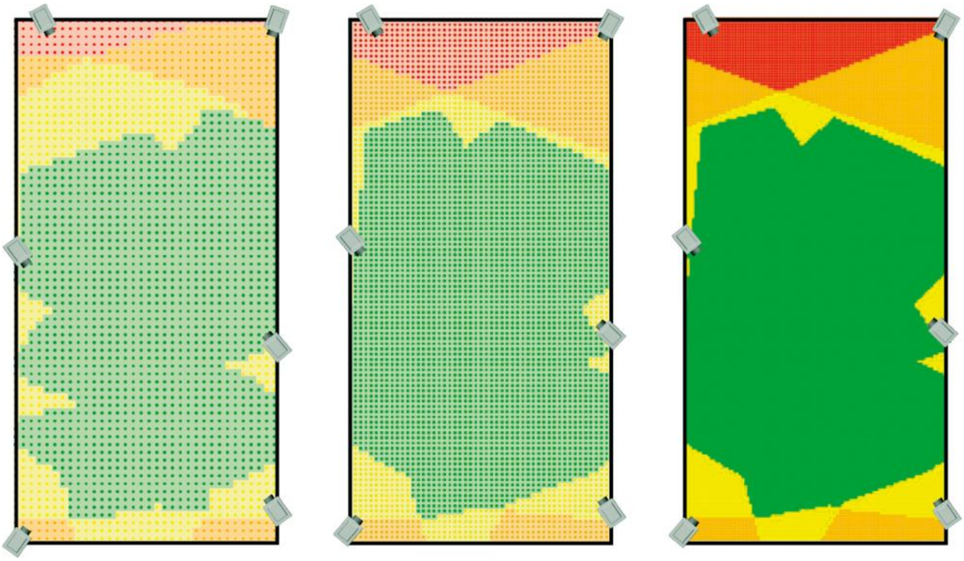


Figure 5. The optimized camera placement for a room (4 m × 8 m) with six tracking cameras was determined at calculative resolutions of 12 cm, 8 cm, and 4 cm (left to right). The results are visualized using heat maps. Red areas are not captured by any camera, orange areas by one camera (which is not sufficient), yellow areas by two cameras (which is sufficient, but not recommended since the view for a camera may be obstructed), and green areas by at least three cameras (which is considered to be sufficient for a reliable usage).

at least three cameras. The afore-selected calculation resolution defines the grid structure's density, the distance between the layers, as well as the variation increment of the cameras. The calculation resolution is, along with the number of cameras and the room size, the deciding factor for the software's calculation time. Depending on these factors, the calculation time may vary between a few seconds and several hours. The discrete approach using a point cloud allows for significantly reducing the calculation load but may also lead to some uncertainty in finding only a "local optimum." A higher resolution will lead to a more accurate result but also increase the calculation time (see Figure 5). In a common room, as shown in Figure 5, the number of possible placements for one single camera consists of several billion possibilities. The total number of possible camera configurations exponentially increases with every added camera. Some optimization algorithms were applied that filtered irrational configurations in advance, significantly reducing the calculation load. Ultimately, the software plots a heat map of the final solution, along with numerical information relating to the positioning and orientation of each camera.

2.3 Visual Display

Vision is the primary source of sensory information for humans (Dahm, 2005), with other senses only playing a marginal role, specifically in environments including automotive traffic (Hills, 1980). VR designers must carefully determine which type of visual stimulus shall be implemented in simulations of such environments. The choice may affect the user's sense of perception and well-being, including the experience of presence, their susceptibility to simulator sickness, as well as their behavior.

It is postulated that various sources cause sickness symptoms in virtual environments, with most of them being associated with aspects of visual and spatial perception relating to sensory conflicts or optical effects. The leading cause of nausea in virtual environments is thought to be the lack of consistency in sensory information. Such discrepancies (e.g., caused by system latency) may lead to a visuo-vestibular mismatch (VVM) and retinal image artifacts related to the vestibulo-ocular reflex (VOR), potentially causing symptoms of discomfort and ataxia effects (Bertolini & Straumann, 2016; DiZio & Lackner, 1992; Frank et al., 1988; Peli, 1999). This underscores the importance of relying the visual stimuli on reliable motion capture and ensuring low end-to-end latencies for immersive human-in-the-loop systems. In addition, vergence-accommodation conflicts (VAC) have been reported to cause discomfort and eye strain (Hoffmann et al., 2008; Peli, 1999), although these restrictions may not be relevant to virtual tasks that do not require interaction within the personal space since distances over two meters do not require active accommodation. Long system latencies have been reported to cause oscillopsia effects (Allison et al., 2001), although the periodic jumping of virtual scenery may also be related to an overloading of the calculation capacity of the simulation software. Furthermore, screen properties such as resolution, refresh rate, luminance, contrast, and FOV are believed to interact in the production of asthenopia-inducing flicker (Kolasinski, 1995; Pausch et al., 1992). Failing to adjust technical parameters to the user's physiological individuality—for example, an incorrectly set interpupillary distance (IPD) in HMDs—may also affect the user's well-being (Peli, 1999).

Sickness in virtual environments has been studied extensively and discussed controversially. Interested readers and in particular VR designers may find reports that have reviewed numerous potential factors that have been postulated to be linked to nausea and discomfort in virtual environments in Davis et al. (2014), Kolasinski (1995), Kolasinski and Gilson (1998), and Peli (1999).

The different types of display formats and the associated characteristics and requirements have been discussed since the advent of highly immersive virtual environments (see Ellis, 1994). Visual displays for these environments can be essentially categorized into two major display formats: head-worn display systems and floor-mounted display systems. Head-worn display systems (e.g., HMDs) isolate users from the actual environment entirely. Surrounding display systems, such as flight simulators, driving simulators, or CAVEs (Cave Automatic Virtual Environments), immerse users "indirectly," often involving some physically mocked up components that are used with the displays. HMD-based and CAVE-like systems have been both successfully implemented for street-crossing experiments (Feldstein et al., 2018). The different advantages and disadvantages of the two concepts are briefly outlined in the following.

Previous experiments have shown that CAVE-like systems overall performed better in terms of task performance (e.g., Cavallo, Dang, et al., 2019; Mallaro et al., 2017; Riecke et al., 2005) or experience of presence (e.g., Juan & Pérez, 2009; K. Kim et al., 2014). However, the technological development of HMDs has taken a tremendous leap forward over the past two decades and has especially gained momentum since entering the mass market a few years ago. Existing deficits are continuously tackled on all fronts, improving parameters such as display resolution, FOV, refresh rate, latency, data bandwidth, and ergonomics (Feldstein et al., 2020).

Unlike CAVEs, HMDs allow for complete isolation from the real world, which enhances the user's sense of presence (Slater et al., 1996). This comes, however, with the trade-off that HMD users cannot visually perceive their real body, requiring the implementation of an avatar in HMD-based systems. No avatar or a poorly implemented one may negatively affect the sense of presence (Leyrer et al., 2011; Slater & Usoh, 1994), but also the visual perception of space may suffer (see Mohler et al., 2008, 2010; Phillips et al., 2010; Ries et al., 2009). Investigators should, therefore, be aware that scenarios requiring a substantial visual awareness of the own body and limbs may not work well in HMD-based systems.

HMD-based virtual environments offer an unobstructed field of regard (FOR) with no physical limitations on the user's viewing direction. The FOV, on the other hand, is most commonly significantly reduced. The quantification of the FOV of an HMD model is ambiguous when variable factors, such as the user's IPD, viewing optics, and eye relief, are not taken into account, all of which determine the effective clipping planes that define the user's viewing frustum (see Figure 6). Consequently, the nominal FOV of an HMD should be viewed as a guideline. For instance, changing the eye relief by only one centimeter can cause the available FOV to be altered by up to 30° (see Valve Corporation, 2017).

Inconsistent documentation can also lead to confusion: The FOV of HMDs is often indicated in literature without specifying whether the FOV

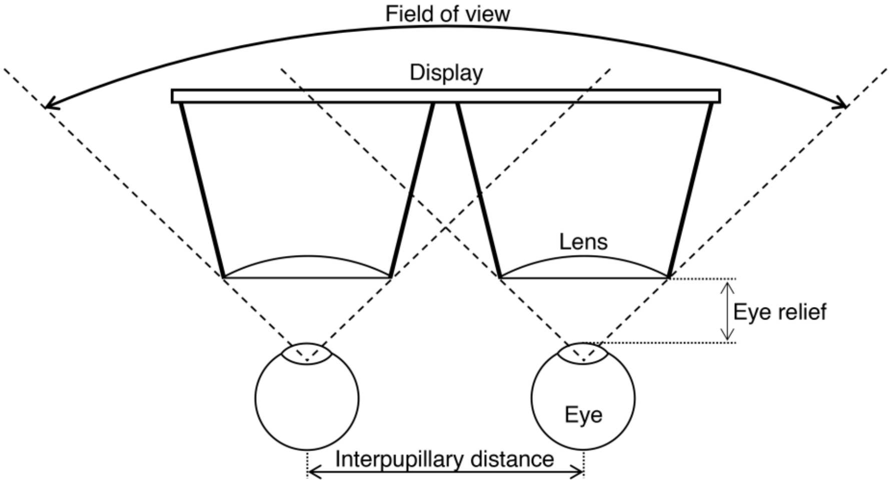


Figure 6. A typical relation between interpupillary distance, eye relief, and field of view in head-mounted displays.

was measured diagonally, vertically, or horizontally. Current state-of-the-art HMDs commonly offer a diagonal FOV of around 110°, such as the *HTC Vive Pro* (HTC, New Taipei, Taiwan), *Oculus Rift S* (Oculus VR, Menlo Park, CA, U.S.), *Samsung Odyssey+* (Samsung, Seoul, South Korea), or *Valve Index* (Valve, Bellevue, WA, U.S.), while a few exceptions like the *VRgineers XTAL* (VRgineers, Dover, DE, U.S.) and *Primax 8K* (Primax, Shanghai, China) provide nominal FOVs of up to 200°. HMD designers currently still have to deal with certain trade-offs: Larger FOVs come with a loss of visual clarity since the available screen technologies have a limited resolution. The deficits are not only caused by the physical resolution of the display but also by bandwidth limitations when handling the data stream of high-resolution content between the operating system and the HMD in real time. For example, the *Primax 8K* has two 4K UHD displays—one for each eye—but can only process a data input of 1,440 px × 2,560 px per eye, which is less than half of the display's physical capacity, requiring the resolution of the scene to be upsampled to the display's resolution post hoc, employing interpolation (Lai, 2017).

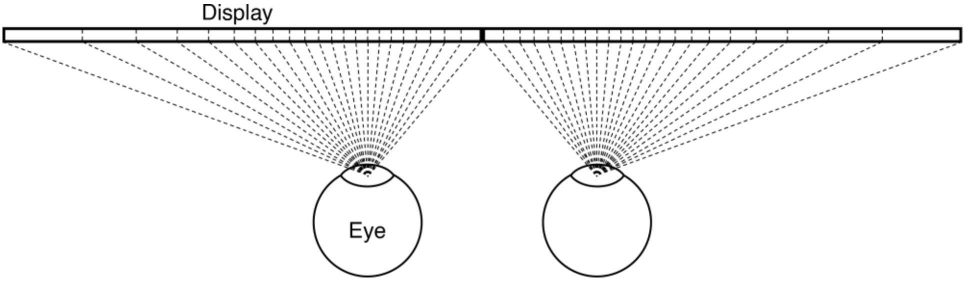


Figure 7. The flatness of the screens would cause the user to perceive the virtual scene increasingly warped with increasing visual eccentricity. Thus, the image rendering process requires carefully implemented algorithms that compensate for such distortion effects.

The human viewing frustum may conveniently be described by the use of polar coordinates due to the nature of how light is projected onto the curved surface of the retina, while the physical characteristics of two-dimensional screens require Cartesian coordinates. As a result, the virtual scene must be projected onto the clipping plane on which the screen is located, requiring the virtual image to be precisely rendered in accordance with the user's point of view (Figure 7). Small rendering errors—that may also be related to the users' anatomical individuality—will have a greater impact on the perceived environment, when the FOV is wide. Consequently, such distortions may affect the user's immersion (Kreylos, 2017). Furthermore, the peripheral retina's increased sensitivity to flicker may cause discomfort for users of HMD systems with wide FOV (Boff & Lincoln, 1988; Peli, 1999).

In CAVE-based virtual environments, the implementation of large FOVs that correspond to the human horizontal visual field of up to 200° (Harrington, 1971) may be easily achieved by surrounding the user with corresponding large screens. On the downside, an omnidirectional FOR, such as available in HMDs, is a difficult endeavor for CAVE systems, due to the complexity and costs of installing such screens above and below the user, as well as technical problems in handling edge effects where orthogonal projection screens meet. Setting up multiple projections and projection

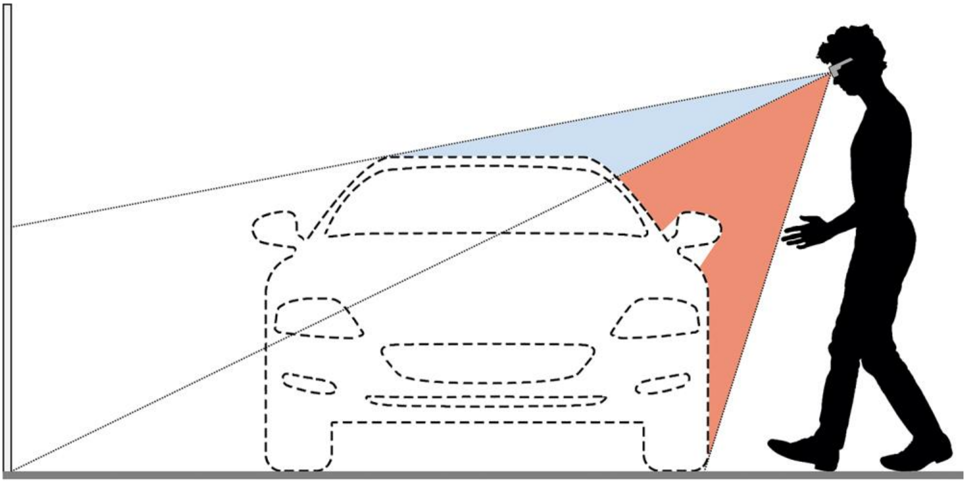


Figure 8. Objects in the lower viewing frustum cannot be simulated when the CAVE possesses only wall projections and no ground projection. This limits the possibility of rendering virtual objects near the user.

surfaces may create visible transitions at the corners and edges and thus disrupt the immersion. Investigators must be aware that virtual objects in close proximity to the user cannot be simulated in CAVEs when the system lacks a projection to the ground. For instance, when standing in a CAVE that is limited to wall projections, vehicles may only be simulated at a certain distance from the user and not within the user's lower viewing frustum (see Figure 8). A wall-only CAVE may be sufficient when pedestrian behavior is being investigated, where the variable of interest is being measured while the vehicle is not (yet) in the immediate vicinity of the participant (e.g., Dommes et al., 2014). However, it is clear that the range of possible studies is consequently limited.

The display's proximity to the user's eyes in HMDs poses a major challenge to HMD manufacturers since the currently available displays are still unable to provide a completely satisfactory resolution and pixel density. Curcio et al. (1990) determined a mean acuity of the human foveal cone density at 66.3 cycles per degree, which corresponds to 0.9 cycles per arcminute. Thus, a display would require a pixel spacing of fewer than 0.45

arcminutes per pixel to match human foveal acuity, which is still far from what current HMDs can offer. That said, in light of the rapid development that can currently be seen taking place in display technology, the necessary screen resolution does not seem unattainable in the medium term. CAVEs do not share the image-resolution difficulty that HMDs have because ultra-high-definition projections are displayed at a greater distance from the user, providing the human eye with absolute clarity of the scenario.

HMDs have evolved quite significantly in terms of comfort when compared to models sold a decade ago: The weight is only a fraction of what it used to be—current models weigh between 400 g and 800 g—and the balance point has also been improved, reducing the load on the user's head. That said, they are still less comfortable to wear than the lightweight 3D glasses that are used in CAVEs. Furthermore, HMD-based systems require additional body equipment in case the motion-capture system—necessary for creating the avatar—uses on-body markers or sensors. On the other hand, researchers may use the readily available motion data post hoc for advanced analyses of the participants' motion behavior. The relatively close and fixed accommodation distance in current HMDs that typically remains in the personal space within less than two meters may cause discomfort to the eyes over time. However, the focus point in HMD—unlike in CAVEs—can be changed by the power of the HMD lenses. The development of small, dynamically focusable lenses is in full swing. Such lenses involve techniques for dynamically changing the refractive index of the optical substrate and are expected to be introduced in HMDs in the medium term.

VR engineers are currently making great strides in developing omnidirectional treadmills that shall introduce a new level of range of motion in VR systems. Marketable and affordable products are expected to be launched in the next few years. Currently, the lengths of the HMD cables and the area that is covered by the motion-capture system dictate the user's range of motion within the physical setup. The range of motion in CAVEs, on the other hand, is dictated by the size of the CAVE system itself, with the screens physically limiting the users' freedom of movement. Designers who build a CAVE need to be aware that a larger CAVE will increase the users' range of motion but will also substantially increase the purchase costs. When comparing the costs of CAVE and HMD systems, costs for CAVE systems are several orders of magnitude higher, with a single projection module usually exceeding the costs for a mainstream HMD. Furthermore, the compactness of HMDs offers flexibility and mobility for the use of HMD-based systems, while the size and inflexibility of CAVEs restrict the ability to relocate CAVE-based systems.

In summary, researchers need to consider the intended virtual tasks when choosing between an HMD and a CAVE system and thus take into account parameters such as the isolation from the real world, self-perception ability, FOV, FOR, visual clarity, ergonomics, range of motion, and budget. The studies associated with the thesis presented here used an *Oculus Rift Development Kit 2* (Oculus VR, Menlo Park, CA, U.S.) that has a resolution of 960 px × 1,080 px (per eye), a pixel density of 386 ppi, a refresh rate of 75 Hz, a low-persistence mode of 2 ms, a nominal FOV of 100° (diagonally), and a weight of 440 g. The original HDMI and USB cables that were three meters long were replaced with ten-meter cables with a built-in amplifier (which is necessary due to the extended range).

2.4 Auditory Display

Auditory display formats in virtual environments can be divided into loudspeakers and headphones. There are fundamental differences between the two approaches, analogous to CAVEs and HMDs for visual displays: Headphones—similar to HMDs—isolate the user from the real environment and require a reliable and accurate motion tracking of the user's head with low latency when spatially directed sound is implemented. Some studies that investigated latencies in virtual acoustic environments provide insight into detectable latency thresholds and allow for deriving system requirements (see Brungart et al., 2005; Lindau, 2009; Mackensen, 2004; Yairi et al., 2007). Headphones require the implementation of head-related transfer functions (HRTF) and interaural time differences (ITD) that allow for the acoustic simulation of sound sources in three-dimensional space. Loudspeakers, on the other hand, are generally independent of the rotations of the user's head since they maintain their orientation relative to the projected virtual environment. Prima facie, this makes loudspeakers the more convenient solution in VR systems, especially when multiuser applications are intended (Gröhn et al., 2007). However, the use of loudspeakers poses a challenge for VR designers with regard to control over perceived spatial imagery (Begault, 2000; Blauert, 1997; Pelzer et al., 2014; Wightman & Kistler, 1989). Realistic simulation of spatial sound through loudspeakers would require complex vector-base amplitude panning, dynamic crosstalk cancellation, and full-sphere surround sound formats to be applied upon the sound processing based on accurately tracked user positions (Fellgett, 1974; Gerzon, 1977; Guastavino et al., 2007; Kuhlen et al., 2007; Lentz, 2006; Masiero & Vorländer, 2014; Pelzer et al., 2014).

Furthermore, multiuser settings that require users to share the same visual and auditory display (e.g., CAVEs) complicate the presentation of several simultaneous stimuli that accurately reflect the users' individual positions, although technically not impossible (see Jiang et al., 2018; Y. Kim et al., 2006; Kurabayashi et la., 2015). The downsides of headphones, on the other hand, may involve ergonomic issues in terms of comfort and the internalization of perceived sound (Brimijoin et al., 2013; Durlach et al., 1992; Hartmann & Wittenberg, 1996; Pelzer et al., 2014). In most cases, CAVE-like systems rely on loudspeaker arrangements, while HMD-based systems use headphones for reasons of similarity with regard to respective technical requirements.

A pair of *Bose* headphones, model *QuietComfort 25* (Bose, Framingham, MA, U.S.), were chosen to be implemented in the street-crossing simulator presented here. These headphones have an integrated active noise canceling (ANC), which cancels out laboratory noise, ultimately supporting the user's isolation from the outside world and thus enhancing the feeling of presence.

White noise or nondirectional ambient noise requires little effort to be implemented in a virtual environment. The integration of spatially directed sound, however, faces two challenges in particular: the availability of manipulable basic sound components and the real-time sampling of binaural sound output in accordance with the user's position and orientation as well as reflection characteristics of surrounding acoustic environment.

The original sound library that was provided with *Silab* was not suitable for the street-crossing simulator's needs, since the original sound was designed for driving simulators that require dull ambient sound, simulating the interior acoustics of a vehicle. Furthermore, the sound output in *Silab* was not designed for use with headphones but with loudspeakers. Within the scope of the thesis presented here, a new sound library was created that contained ambient sounds but also the elemental sound files necessary for the creation of spatially directed sound from the perspective of a pedestrian.

Ambient sounds were recorded at various locations inside and outside the city, using a handheld digital audio recorder, a *Zoom H5* (Zoom, Tokyo, Japan), with two matched unidirectional condenser microphones set at an angle of 90° and recording at a sampling rate of 96,000 Hz. The raw sound was edited, filtered, and mixed using the open-source *Audacity* digital audio editor (The Audacity Team), resulting in a variety of different ambiances (e.g., varying traffic and weather conditions).

Spatially directed sound, on the other hand, cannot be simply prerecorded, since there is an infinite number of possible scenarios that depend on situational parameters. For example, the sound that one single vehicle generates will depend on the vehicle's characteristics (engine type, aerodynamics, rolling noise), vehicle's driving parameters (gear, engine speed, driving speed), environment (wind speed, wind direction, road surface), and relative position to the observer (distance, relative orientation, relative direction of movement). For this reason, a sound library with basic vehicle sound elements was created from a pedestrian's perspective as a first step. The engine sounds of gasoline and diesel vehicles were recorded at different engine speeds. The engine noise was isolated by recording the respective vehicle on a chassis dynamometer, avoiding wind and rolling noises. The engines were recorded at incremental engine speeds in steps of 1,000 rpm.

The recording of the tire sounds was carried out on a real road surface to ensure realistic sound characteristics. This led, however, to airflow around the microphones that were mounted next to the tires, corrupting the recording and making sound collection from microphones on the moving vehicle unusable. Stationary microphones, on the other hand, may include distortions due to Doppler effects when no constant distance between the vehicle and the microphones is maintained. For these reasons, the recording took place at a small traffic circle with the microphone placed in the middle and the car driving around it. By doing so, the microphone maintained the same distance to the tires at all times. An additional inconvenience was the disturbing noise coming from the vehicle's engine. The solution to this problem was the use of an electric car, which eliminated the noise emitted by a combustion engine. The sound was recorded at incremental vehicle speeds of 5-km/h steps.

All soundtracks consisted of only a few seconds, which is essential for real-time applications since files must be kept small in order to avoid long loading times. Therefore, the tracks must be played in loops during the simulation. This required the soundtracks to be edited so that the beginning and the end of each track ultimately matched, ensuring a smooth subliminal transition as left the loop was cycled.

The recorded and edited soundtracks did not yet create any binaural, spatialized sound effects. This required a real-time programming interface that sampled the audio tracks for the virtual vehicle in accordance with the instant simulation parameters, involving the observer, the vehicle, and the environment. Such a sample player was programmed using the *Pure Data* programming language, with widely available open-source content facilitating the programming task significantly. The sample player selected the soundtracks in accordance with the vehicle type, vehicle speed, and engine speed. Since the soundtracks for the engine were only available in 1,000-rpm increments and tire and wind noise in 5-km/h increments, parameters in between were interpolated by gradually mixing two consecutive engine speeds and vehicle speeds, respectively. The created sound additionally took into account the observer's relative position and orientation to the sound source, as well as environmental factors (e.g., wind), to mix a realistic three-dimensional, binaural acoustic output.

Finally, a data processing unit (DPU) was necessary that connected the sound engine with the motion-capture system and virtual-environment software. This interface was optimized to reduce the end-to-end delay and, thus, avoid audio-visual or audio-proprioceptive conflict.

2.5 Cabling System

In the setup described in this dissertation, body-mounted components such as the visual display (i.e., the HMD), the auditory display (i.e., the headphones), and the motion-capture unit for rotational head movements (integrated into the HMD) must communicate with the external stationary control center. The control center receives data from the on-body motion-capture unit and the external camera-based motion-capture system, simulates the virtual environment based on the received data, and sends corresponding visual and acoustic images to the HMD and headphones. Additionally, the HMD, the headphones' ANC, and the on-body LED markers require power. The HMD has to be supplied with a current of 500 mA at a voltage of 5 V through the USB port, while the ANC was simply powered using a standard AAA battery. The LED markers were supplied with power from a custom battery belt around the user's waist. There was a total of three cables between the VR user and the stationary control center: An HDMI cable provided the visual image to the HMD, a standard stereo audio cable delivered the audio to the headphones, and a USB cable handled the HMD's power supply as well as the data transfer of the HMD's motion-capture unit.

When the user was walking through the virtual environment while wearing the HMD, the risk of tripping over or getting caught in the cables was a major ever-present problem. In the early stages of the VR system, the investigator was required to walk with the participant and keep the cables out of the participant's way in order to avoid any accident or disrupt the immersion. For this reason, it was necessary to develop a solution that would keep the cables out of the user's way at all times. The idea was to install a dynamic cable-routing system above the user so that the cable could follow the user's movements. The cable-routing system can consist of a combination of translational and rotational DOFs to cover a specific area, such as shown in Figure 9. A rotation-based double-hinged arm has several advantages over solutions involving linear guides: The components are significantly less expensive, the construction is less complex, and the area of interest is easier to cover since the arm can swing out and does not require a rail system covering the entire ceiling above the area of interest.

A significant problem with cable-routing systems is the mass of their components and, thus, the associated inertia: If the system passively follows the user's movements, the weight of its components creates a sluggishness

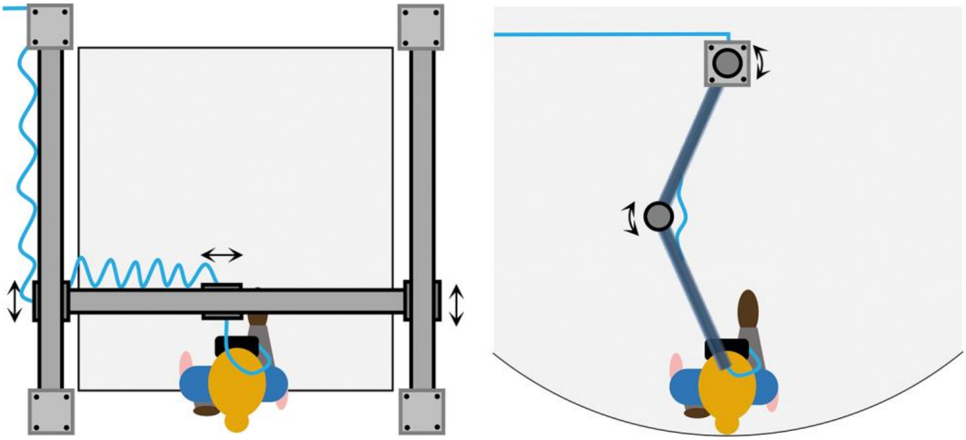


Figure 9. Examples of ceiling-mounted cable-routing concepts: Linear guides (left) and double-hinged arm (right).

that pulls on the user's head, especially when the user accelerates or decelerates. This will interfere with the user's immersion and can even cause the HMD to be pulled off the user's head. This concern applies in particular to the massive linear guides but also to rotation-based systems: Even if the double-hinged arm is built using lightweight carbon-fiber tubes, the outer hinge may add a significant weight that will produce substantial inertia in combination with the arm's leverage effect. A possible solution to this problem may involve adding an active drive. A double-hinged arm with a cable-pull system that was manually operated by the investigator was successfully used in preliminary trials within the scope of the thesis presented here. Future solutions may consider adding two electric actuators (one for each hinge) that will control the cable-routing system automatically based on the VR user's direction of movement. The directions of movement are readily available thanks to the motion-capture system.

The mechanical difficulties and challenges with cabling suggest the use of wireless systems. However, system designers must be aware that these bear different difficulties and challenges in terms of system latency that cannot be easily overcome with the currently available technology.

3 ARTICLE I – EXTENDED ABSTRACT

A Simple Video-Based Technique for Measuring Latency in Virtual Reality or Teleoperation

Published in Feldstein, I. T., & Ellis, S. R. (2021). A simple video-based technique for measuring latency in virtual reality or teleoperation. IEEE Transactions on Visualization and Computer Graphics, 27(9), 3611-3625.

Purpose: VR systems rely on the complex interplay of many different components that produce sensory and computational delays. Designers of VR systems are challenged to minimize end-to-end latencies since long end-to-end latencies and related visuo-proprioceptive conflicts are known to affect the user's experience of presence, task performance, and well-being. VR designers need to track, identify, and reduce latencies within the system from an early design stage.

Method: A simple, yet effective method for measuring latency may consist of recording the VR system with an ordinary consumer camera that is capable of recording with an adequate image resolution at a high frame rate. The motion-captured physical movement within the VR system and the resulting virtual movement that appears on the VR display are recorded with the camera in one sequence. Identifying these events in the footage allows for calculating the end-to-end latency through frame counting.

Method Testing: It was expected that the measuring tool (i.e., the camera) and the human evaluator analyzing the recorded footage would both introduce uncertainties in the determined latency value. The latency measurement method was tested using a camera that recorded at a frame rate of 240 fps and a resolution of 848 px × 480 px, with the ultimate goal of quantifying occurring uncertainties. Twenty untrained latency evaluators analyzed eight latency recordings. The evaluators repeated the analysis five times, resulting in a total number of forty latency values per evaluator

Results: The results confirmed that the VR system partially contains an inevitable underlying latency fluctuation that is due to unsynchronized processes in the system and unaffected by the latency measurement method. Additional uncertainties arising from the measurement method itself were linked to technological factors and human error. There were two technological factors identified that added uncertainty to the determined latency value: the camera sensor's sampling rate due to the sample-and-hold nature of the video footage and the camera sensor's inherent pixel delay due to the rolling shutter of the pixel exposure circuitry. With regard to the analytical frame-counting procedure that relied on human evaluators, the results of the determined latency values indicated a small variance (0.017 s) between different evaluators. The variance of determined latency values within evaluators (i.e., same evaluators repeated the measurements) was much smaller (0.004 s). A more in-depth analysis revealed that the human-induced variance was almost exclusively due to the difficulties of judging the physical motion and not the virtual one.

Conclusion: The suggested latency measurement technique is easy to apply, and the determined values have reasonably good accuracy for most purposes. A number of simple measures have the potential to reduce the combined technological and human-induced uncertainty to one millisecond. Such measures may include an increase in the frame rate, the use of a turntable, and the implementation of a motion-indicating LED.

4 ARTICLE II – EXTENDED ABSTRACT

Pedestrian Simulators for Traffic Research: State of the Art and Future of a Motion Lab

Published in Feldstein, I. T., Lehsing, C., Dietrich, A., & Bengler, K. (2018). Pedestrian simulators for traffic research: State of the art and future of a motion lab. International Journal of Human Factors Modelling and Simulations, 6(4), 250–265.

Purpose: In the past, traffic researchers primarily focused on investigating behavior of drivers. However, the modern challenges surrounding autonomous vehicles and complex urban traffic management call for understanding all road users, including pedestrians. There have been various approaches introduced for experimental assessments and the collection of data that reflect pedestrian behavior on real streets, such as the pretend-road method or the shout task. These methods were compromise solutions and had various shortcomings with regard to pedestrian safety, repeatability, and scenario variability. Technological advances over the past two decades led to the advent of simulators based on VR technology that allow for a safe, repeatable, and reproducible assessment of pedestrians while offering a wide range of possible scenarios.

Method: A newly built VR setting that immerses the user into the scenario of a pedestrian is introduced. It consists of a video-based motion-capture system, a state-of-the-art HMD, and an adjusted driving-simulator software. In a series of pilot studies with virtual street crossings, the system was investigated for its performance as well as user acceptance of the given setting. In addition, the simulator was used in a number of experiments with

linked simulators: Participants in driving simulators had to react and interact with pedestrians who were either programmed bots or human-controlled avatars that were operated using the new pedestrian simulator.

Results: The results of the exploratory research suggest that users of the pedestrian simulator widely accepted the setting and quickly acclimatized to the new environment. Apart from one case out of 75 users, there were no signs of simulator sickness. The participants indicated a satisfactorily high experience of presence in the virtual setting (measured using the Witmer-and-Singer Presence Questionnaire). Recorded walking patterns show a familiarization process taking place, suggesting that participants get gradually used to the virtual environment. However, the walking pace in the virtual environment remained significantly lower than in the real world even after some period of familiarization with the virtual environment, revealing a maximum walking pace that was 12.5% lower in the virtual environment on average. In studies with linked simulators, a lack of interaction between the driver and the pedestrian willing to cross the street was observable whenever the pedestrian was programmed, which resulted in unrealistic and reckless pedestrian crossings. Whenever the pedestrian avatar was human-controlled, a natural interaction behavior between the two road users produced considerate and prudent behavior.

Discussion: The initial results of the new system show that the development of such simulators is promising and heading in the right direction. Individual deficits of the components used are apparent and offer room for future optimization (e.g., the technical characteristics of the HMD's display). Social interaction between road users remains an essential component in lifelike street-crossing scenarios and investigations. The studies with linked simulators show a broader application potential for such pedestrian simulators, which allowed for more realistic encounters between road users than scenarios with programmed bots.

5 ARTICLE III – EXTENDED ABSTRACT

Impending Collision Judgment from an Egocentric Perspective in Real and Virtual Environments: A Review

Published in Feldstein, I. T. (2019). Impending collision judgment from an egocentric perspective in real and virtual environments: A review. Perception, 48(9), 769–795.

Purpose: There are numerous experimental parameters and factors that may affect how pedestrians judge an approaching vehicle. These factors need to be taken into consideration when investigating road-crossing decisions by pedestrians and interpreting the respective findings. Many experiments over the past decades strived to single out such factors, ultimately contributing to a more comprehensive understanding of human perception and decision-making processes involving approaching objects.

Measurement Methodologies: Investigators need to be aware that the measurement method itself may affect the outcome with different methods employing different cognitive abilities. Such methods may be essentially divided into estimation tasks and discrimination tasks. Estimation tasks contain coincidence anticipations or interceptive actions, whereas discrimination tasks commonly employ comparisons in pairs, although within-group comparisons have also been reported.

Influencing Factors: The numerous factors that have been shown to affect one's judgment of approaching objects may be divided into human factors, compositional factors, and technological factors. Human factors include parameters such as gender, age, driving experience, risk tolerance, and physiological characteristics but most notably the tendency of people in road-crossing scenarios to err on the safe side. Compositional factors include classical depth cues as well as approaching modalities such as angle and speed of approach, spatial and temporal distance, and observation time. In addition, surrounding circumstances, such as light and weather conditions, may also play a role. Technological factors include shortcomings in the VR settings that may involve monoscopic viewing conditions, a fixed focal plane, a reduced FOV, insufficient display brightness and contrast, and unrealistic virtual events.

Environment Comparison: A comparison between studies that have been conducted in real environments and studies that used virtual environments suggest that various observed factors and phenomena depend on the experimental environment, indicating the questionable adequacy of virtual environments for road-crossing experiments. This is further supported by studies that explicitly compared road crossings in real and virtual environments, revealing some substantial differences between participant behavior in the virtual environment and behavior on actual streets.

Conclusion: Researchers intending to investigate road-crossing scenarios need to consider a number of factors that may influence the participants' decisions. Many of the observed factors are intertwined. For example, technological factors may depend on psychophysical effects that are associated with the environment's compositional characteristics. In virtual environments, investigators must examine the suitability of specific VR settings for the intended research question since virtual environments have been shown to alter the perception and behavior of human users.

6 ARTICLE IV – EXTENDED ABSTRACT

Road Crossing Decisions in Real and Virtual Environments: A Comparative Study on Simulator Validity

Published in Feldstein, I. T., & Dyszak, G. N. (2020). Road crossing decisions in real and virtual environments: A comparative study on simulator validity. Accident Analysis & Prevention, 137, Article 105356.

Purpose: Newly developed simulators require a validation process that confirms the legitimacy of substituting a real environment for a virtual one when aiming to answer specific research questions. In the past, validation attempts for street-crossing simulators have shown that pedestrian behavior on real streets differed from that on virtual streets. For example, pedestrians showed a significantly riskier behavior when confronted with virtual cars as opposed to confrontations with actual cars. In the study presented here, an attempt was made to validate a new type of street-crossing simulator. The intention of pedestrians to cross a street was to be investigated within a virtual environment and compared with the intentions of pedestrians crossing a corresponding real street.

Method: Thirty participants standing on the edge of a one-way street faced vehicles that approached at a speed of 30 km/h, 35 km/h, or 40 km/h. The participants were instructed to choose the moment when they felt that the approaching vehicle was too close to permit a safe crossing. The participants indicated their decisions by taking a step back that led them off the street and onto the sidewalk, thus ensuring a realistic road-crossing assessment without exposing the participants to actual danger. Each participant made fifteen encounters in the virtual environment and fifteen in the real one.

Results: The time-to-contact (TTC) with the approaching vehicle at the moment of the participants' road-crossing decision turned out to be significantly lower in the virtual environment than on the real street. It was also shown that participants primarily based their crossing decisions in the virtual environment on the spatial distance of the vehicle, neglecting the vehicle's speed. In the real environment, however, participants based their crossing decision on the temporal distance (i.e., TTC), thus incorporating the vehicle's speed in their crossing decisions.

Discussion: Despite the high functional similarity between the real and virtual scenarios, the participants behaved differently in the virtual environment than in the real one. A riskier behavior has once again been observed in a simulator setting. Furthermore, participants showed difficulties in judging the vehicle's speed within the virtual environment and thus neglected this parameter in their decision-making process during the virtual experiment. This shortcoming may be attributed to the inadequacy of the VR components, more specifically, the HMD's display characteristics. VR technology needs to evolve further to eventually remove those existing differences between real and virtual environments that modify users' perception and behavior. However, failing to observe identical behaviors in both environments does not disqualify the VR system directly from being a useful research tool since a downscaling of the measurement scale may unravel the validity issue.

7 DISCUSSION

7.1 Simulator Design

7.1.1 Virtual Environment

Researchers need to be aware that adding excessive details and complexity to the virtual environment will potentially increase the system latency due to the processing units' limited resources and data bandwidth (Feldstein & Ellis, 2021). It is therefore recommended to carefully take into account what details may be necessary within the scenery and which ones the VR user may not notice anyway.

The *Silab* traffic simulation software was initially developed for use with driving simulators in research applications. The software is a well-thought-out solution for simulating traffic scenarios and collecting a wide range of data relating to driving behavior. However, the software's source code is proprietary, which impedes user modifications other than those foreseen by the producer. This fact becomes particularly problematic when it is necessary to make alterations to the software. In collaboration with the software producer, an avatar was created that responded to data provided by the motion-capture system, mimicking tracked movements. Unfortunately, the avatar was poorly and unrealistically designed, the shape and agility were unnatural and lacking details, while the individual body segments were limited in their number and controllability (see Feldstein et al., 2018). It was apparent that pedestrians and their design were not a focal point of the software producer, whose core business is driving simulators, and with the source code being protected, user improvements and modifications to the avatar and its controls were not possible. For egocentric pedestrian investigations, a poorly designed avatar may be acceptable since the participants focus on the environment rather than on their avatar. In the experiments, participants judged the self-representation to be unrealistic, but this did not seem to affect the given task (Feldstein et al., 2018). However, when interaction studies are carried out, such as experiments on driver-pedestrian interaction (see Lehsing, 2019; Lehsing et al., 2015, 2019), implementing the required body language, facial expressions, and verbal communication is not possible with given restrictions in the software environment.

In conclusion, specialized software applications such as *Silab* are quite costly, and users are limited in making modifications. There are some powerful source-available game engines, such as *Unity* (Unity Technologies, San Francisco, CA, U.S.) or *Unreal Engine* (Epic Games Inc., Cary, NC, U.S.), that may be genuine alternatives for researchers seeking to develop a simulator for behavioral research. Such game engines enjoy great popularity due to their broad application range and source-code-available characteristics, with a large developer community creating massive online-available content. Most classical simulator applications are still lagging behind the captivating realism and light effects that may be achieved with such game engines that are simply evolving at a much faster pace. Researchers who seek a budget-friendly solution without sacrificing the quality of the virtual environment may consider using such engines. It should be noted that a considerable amount of time and effort will be needed to make applications such as *Unity* or *Unreal Engine* run on the desired hardware platform, unlike commercial solutions that, apart from some initial configuration, are plug-and-play. Tutorials and support in numerous online forums are widely available. Some initial traffic simulation trials using Unity within the scope of the thesis presented here turned out to be a promising approach, but were not part of the experiments found in the annex, which were all carried out using *Silab 4.0*.

7.1.2 Motion Capture

The new motion-capture suit significantly improved the system's tracking reliability. However, the elastic suit with sewn-in cables was prone to malfunctions due to cable breaks, even though the cables had sufficient excess lengths. This problem may be resolved by implementing coiled cables in the suit or sewing the cables into the fabric in sinuous patterns. An alternative solution consists of using wireless LED markers that are *Bluetooth* controlled and wirelessly charged. Some promising prototypes of such wireless markers were designed and built within the scope of the thesis presented here, but not yet implemented in the simulator system.

Another advantage of LED markers over reflective markers may be the possibility of operating LEDs with different frequencies, which in turn would allow for automatically identifying individual markers. This would further enhance the tracking reliability and shorten the calibration process in the beginning. Markers that keep disappearing from the sensors' visual fields will benefit in particular from such an approach: The system would be able to instantly re-identify the markers at any time as a result of the marker's specific frequency.

An essential downside of working with markers is the limitation of trackable body segments and details. The classic *Vicon Plug-in-Gait* model breaks the human body down into seventeen different body segments (Feldstein et al., 2018). This prevents details such as finger movements or facial expressions from being reproduced, although they may be relevant to human-interaction studies. Some optical motion-capture systems forego the usage of markers and rely on complex image-processing algorithms (see Feldstein et al., 2015). Such image processing tools made a huge leap forward in recent years, with reliable and inexpensive hardware and software solutions being introduced by the gaming industry, such as the *Microsoft Kinect* sensor (Microsoft, Redmond, WA, U.S.). Some initial trials within the scope of the thesis presented here showed a promising approach when replacing the costly and complex *Vicon* system with an array of multiple *Kinect* sensors that allow for a real-time data fusion. Such markerless systems are inevitably the future, although it will still take a few years until the same reliability that is provided by a marker-based system can ultimately be reached.

7.1.3 Visual Display

Although technological progress is clearly observable in modern HMDs (Feldstein et al., 2020), experimental assessments with the simulator system presented here still showed deficiencies regarding the technology behind the visual display (Feldstein et al., 2016, 2018; Feldstein & Dyszak, 2020; Feldstein & Peli, 2020). A key technological factor that needs to be addressed is the HMD's insufficient display resolution that is far from the capacity of the human retina. While the HMD used in the experiments presented here provided images at one megapixel per eye, a display resolution of approximately 324 megapixels per eye would be necessary to match human foveal acuity at a $90° \times 90°$ FOV (R. N. Clark, 2017). Although the visual acuity across the human retina varies radically, it is economically unlikely that there will be displays with nonuniform resolutions that would allow for lowering screen-resolution requirements. Nevertheless, a continuous increase in display resolution is expected to take place over the next few years, which presents VR designers with another problem: the computational expenditure for rendering images at respectively high resolutions and the related increase in system latency (Feldstein & Ellis, 2021). A solution to this problem may consist of nonuniform image rendering by providing the foveal view with a high-resolution image while the periphery is rendered at a lower resolution, thus taking advantage of the eccentric acuity degradation of the human eye (Albert et al., 2017; Hitchner & McGreevy, 1993; Kiyokawa, 2006; Luebke & Hallen, 2001; Reddy, 1995). Furthermore, future research may further investigate the shortcomings regarding the reduced FOV and the fixed focal plane in HMDs, although these factors are expected to be of minor relevance for road crossings on one-way streets (Feldstein, 2019; Feldstein & Dyszak, 2020).

In conclusion, rendering a visual scenery remains a challenge since excessive virtual details and image resolution affect latency and thus create a visuo-proprioceptive conflict (Feldstein & Ellis, 2021). Sparse environmental details and a low image resolution, on the other hand, will negatively impact the experience of presence. VR designers are therefore challenged to strike a balance between acceptable system latency, implemented level of content detail, and rendered visual resolution. This challenge will persist at least in the medium term since the VR market's trends to implement wireless data transfer will further aggravate the limitation of the data flow rate.

7.1.4 Auditory Display

Numerous studies have shown that the auditory display within the virtual environment may affect the experience of presence (see Bergström et al., 2017; Dinh et al., 1999; Hendrix & Barfield, 1996; Kobayashi et al., 2015; Larsen & Pilgaard, 2015; Larsson et al., 2007; Nordahl & Nilsson, 2014; Patel as cited in Slater & Wilbur, 1997; Poeschl et al., 2013). That said, acoustic feedback may not necessarily be crucial to the investigated task (e.g., Cavallo, Dommes, et al., 2019; Oxley et al., 2005; Simpson et al., 2003; Soares et al., 2020). Investigators need to decide what level of acoustic stimuli is required in their simulation: no acoustic stimuli, white noise, nondirectional ambient sound, or binaural spatially directed sound. Each level drastically increases the technical expenditure necessary for implementing the type of stimulus. Patel (as cited in Slater & Wilbur, 1997) observed that the step from no acoustic stimuli (i.e., the exposure to laboratory noise) to white noise led to the highest increase in the experienced presence. The benefit of implementing spatially directed sound is debatable: Larsen and Pilgaard (2015) and Larsson et al. (2007), unlike Hendrix and Barfield (1996), did not measure a higher presence experience when nondirectional ambient sound was upgraded with spatially directed sound.

These findings underline the importance of isolating the VR user from the real environment but suggest that the quality of the auditory display may play a marginal role. Obviously, the relevance of the auditory perception will largely depend on the task facing the user. For instance, the perception of directional sound may be relevant when a pedestrian intends to cross a road with two-way traffic, containing multidirectional threats that may not be fully visually monitored. Future studies that will investigate the impact of auditory perception in a wide range of different traffic scenarios are warranted.

7.1.5 Cabling System

The hardware interface, in particular the cable management, is an element that is often neglected during the design process of sophisticated technologies that rely on the interaction of several elaborate components. This negligence is not limited to VR systems but may also be observed in other challenging technologies. For example, the continuous development and modification of the *Mir* space station led over time to the growth of tangled and unorganized cabling all over the station, ultimately contributing to the power shortage crisis in 1997 that substantially threatened the station's safety (Ellis, 2000; Harland, 2005). Another example is the design process of the world's largest passenger aircraft, the *Airbus A380*, in which cable management was neglected throughout the change control processes, resulting in significant production delays in the mid-2000s and thus a loss of several billion euros (N. Clark, 2006). These examples illustrate the importance of including the hardware interface early in the design process, even if the focus itself is placed on the central components.

Wired interfaces in VR are an inconvenient necessity, with the best cable management being the one that goes unnoticed by the user. It is technologically unlikely that wireless data transmission will entirely replace wired solutions in the medium term since wireless connections cause a significant latency and are thus not suitable for current VR applications. This problem is likely to persist since technological progress is being challenged by the ever-increasing amount of data that needs to be transmitted (Feldstein & Ellis, 2021). Future VR systems will likely evolve toward the miniaturization of the control center, allowing the control center to be mounted on the user's body (e.g., integrated into the HMD hardware). However, current battery life and heat dissipation are major challenges for VR designers, which in turn further impede the design of ergonomic body-worn systems that do not require wired connections to external systems.

Until VR technology reaches wireless maturity, it is advised to implement a cable-routing solution, such as discussed in the thesis presented here. An additional recommendation is installing actuators on the routing system so that the system may automatically follow the user's captured movements, instead of being passively dragged. Such an intelligent cable-routing system is an essential component that contributes to the users' experience of presence and mobility within the system.

7.2 Validation Approach

There is no doubt about the necessity to validate novel simulator systems for behavioral research since differences between behavior in simulators and behavior in the real world have been reported on numerous occasions (see Section 1). VR designers must carefully consider the validation approach. The validation study must be application-oriented, with the validation process covering the intended research scenarios. The goal is to ensure that the stimuli provided in the virtual environment produce the same human behavior as similar scenarios in the real world. The validation process must investigate all relevant stimuli while ensuring the viability of necessary reference measures in the real world, with the participants' safety being ensured at all times.

The simulator that was developed within the scope of the thesis presented here was primarily intended for researching pedestrian behavior at road-crossing scenarios. Thus, for proper validation, the behavior of pedestrians in the virtual environment needed to be compared with their behavior at real road crossings. The validation study required a design that made it possible to measure genuine road-crossing behavior but ensure the pedestrians' safety while acquiring the reference values on the real street. Actual road crossings always pose a risk and should be avoided in experiments on real streets. A different approach was therefore required, which also allows for measuring pedestrians' crossing intentions. Another significant challenge was the manipulation and control of experimental parameters in the real environment. For example, a classic gap-acceptance study that requires a motorcade with precisely controlled vehicle speeds and gaps would have been nearly impossible to set up on a real street.

For these reasons, a new approach was introduced: the step-back method. This method only required one approaching vehicle with digital cruise control that allowed for a precise and constant speed. No participant was at risk at any time since participants expressed their acceptance threshold for crossing the street by stepping back. The experiment by Feldstein and Dyszak (2020) demonstrated that the step-back method is an adequate substitute for actual road-crossing assessments. This study also showed significant behavioral differences in real and virtual environments regarding the participants' decision-making processes, demonstrating once again the importance and necessity of validating simulators for behavioral research.

Future studies are needed that address specific technological shortcomings and attempt to determine dose-effect relations as well as predict performance differences by explicitly manipulating the respective shortcoming. This may help manufacturers of VR components focus on relevant technological shortcomings when improving their products and help investigators accurately interpret their research findings when using VR settings.

7.3 Conclusion

All in all, considerable effort is required to design and build a VR system for behavioral research. Many different components need to be carefully thought through in accordance with the research purposes. In addition, the complex interplay of the components poses a considerable challenge. Once this effort is accomplished, however, a wide range of different virtual scenarios is possible that allows for easy manipulation of experimental parameters while ensuring the participants' safety and the experiment's repeatability.

The challenge in the validation process entails the acquisition of reference values from the real world. The difficulty consists of setting up an appropriate experiment in a real environment that allows for manipulating the parameters of interest and measuring the data of interest while ensuring the participants' safety. The complexity of acquiring the reference values in a real-world experiment explains why proper validation processes are often neglected and illustrates why virtual environments are such a convenient solution for behavioral studies. Researchers must be aware that the ecological validity of experimental findings using VR settings may only be given once the setting has undergone a successful and comprehensive validation process.

ANNEX

A Article I – Full Reprint

A Simple Video-Based Technique for Measuring Latency in Virtual Reality or Teleoperation

Revised reprint from Feldstein, I. T., & Ellis, S. R. (2021). A simple video-based technique for measuring latency in virtual reality or teleoperation. IEEE Transactions on Visualization and Computer Graphics, 27(9), 3611-3625.

Abstract. Designers of virtual reality (VR) systems are aware of the need to minimize delays between the user's tracked physical actions and the consequent displayed actions in the virtual environment. Such delays, also referred to as end-to-end latency, are known to degrade user performance and even cause simulator sickness. Though a wide variety of hardware and software design strategies have been used to reduce delays, techniques for measuring and minimizing latency continue to be needed since transmission and switching delays are likely to continue to introduce new sources of latency, especially in wireless mobile environments. This article describes a convenient, low-cost technique for measuring end-to-end latencies using a human evaluator and an ordinary consumer camera (e.g., cell phone camera). Since the technique does not depend upon the use of specialized hardware and software, it differs from other methods in that it can easily be used to measure latencies of systems in the specific hardware and software configuration and the relevant performance environments. The achievable measurement accuracy was assessed in an experimental trial. Results indicate a measurement uncertainty below 10 ms. Some refinements to the

technique are discussed, which may further reduce the measurement uncertainty to approximately 1 ms.

A.1 Introduction

Virtual environments, also referred to as virtual reality (VR), are display systems that are designed to immerse users in computer-generated virtual worlds, providing them with the sensation of presence in the virtual environment. These display systems ideally offer a vivid, surrounding, extensive, inclusive experience that matches aspects of the real world and thus leads to a strong sense of *presence* (Slater et al., 1996; Steuer, 1992). This means that the system offers rich virtual content (*vividness*) from multiple directions (*surrounding*), stimulating a variety of different human sensory modalities (*extensiveness*) while isolating users from the real environment (*inclusiveness*). In this somewhat isolated state, all the simultaneously presented sensory stimuli are consistent with each other in the sense of content and timing (*matching*), much as they are in a real environment. The role that the sense of presence plays in the success of such virtual environments is a matter of considerable discussion (Cummings & Bailenson, 2014; Slater, 2009; Slater & Wilbur, 1997) and especially debated in terms of how to measure presence (Ellis, 1996; Insko, 2003; Lessiter et al., 2001; Singer & Witmer, 2000; Slater, 1999; Slater & Steed, 2000; Usoh et al., 2000; Witmer et al., 2005; Witmer & Singer, 1998). However, it is clear that a delay between the user's movement and the response to the movement in the virtual environment may negatively affect the user's performance and feeling of presence (Blissing et al., 2016; Ellis et al., 1997, 2002; Meehan et al., 2003; B. Watson et al., 2003; Welch et al., 1996). Additionally, the decorrelation of the perceived visual motion from the concurrently sensed vestibular motion can cause a form of motion sickness, also referred to as simulator sickness (Allison et al., 2001; Bertolini & Straumann, 2016; Davis et al., 2014; DiZio & Lackner, 1992; Frank et al., 1988; Peli, 1999).

Some latency in VR is unavoidable and a design concern for all systems that require direct user interaction. Those that involve position signals being transmitted over significant distances and passed through communication systems with inherent switching delays are especially susceptible (see Peñín, 2000). This susceptibility arises because the control of some of the latency is external to the hardware and software that generate the virtual environment itself. Nevertheless, a general design goal of such systems is the minimization of overall delays.

The formal psychophysical threshold (see Green & Swets, 1966) for visual delays that are detected by average human observers has been determined to be around 15 to 20 ms (Adelstein et al., 2003; Ellis et al., 1999, 2004; M. J. Regan et al., 1999). Notably, not all studies investigating users' abilities to discern latencies in virtual environments employ formal psychophysical methods used in the cited studies. It has been shown that full psychometric functions for latency discrimination provide a basis for predicting the frequencies of correct user detections regarding ranges of short to long latencies. Carmack (2013) recently confirmed some of these findings. There is, however, significant individual variation evident when user detection of latency is studied with standardized psychophysical techniques aiming at measuring the full psychometric function. Some participants who have had practice at discerning slight movements of visual displays (e.g., airline pilots) exhibited just-noticeable differences (JNDs) at far below 10 ms. For this reason, Jerald (2010) suggests that system designers should aim as low as 3 ms for the latency requirement of future systems. Interestingly, latency discrimination does not appear to follow Weber's law (see Fechner, 1860): The JND for latency discrimination of the psychometric function remains unaffected by the latency baseline value (Ellis et al., 1999). In fact, this finding allows the inference of the JND for a system with a theoretical zero latency from systems with several measurable latencies by simply extrapolating their results back to a zero-latency condition.

It should be noted that a perceptible latency does not just by itself cause simulator sickness or other performance disturbances: Interaction between the user's task and the virtual environment's scene content can play a determinative role. For example, scenarios that require real locomotion and especially rapid head movements will be more likely to cause motion sickness by disrupting the habitual correlation between visual and proprioceptive motion (Allison et al., 2001; Bertolini & Straumann, 2016). There are some indicators that virtual-environment users may be more sensitive to the artifactual effects arising from head movements than those occurring from delayed virtual objects related to hand movements (Ellis et al., 2004). In general, the threshold of *detectable* latency may not be equal to the level of *motion-sickness-inducing* latency (see Draper et al., 2001).

Our article discusses the latency measurement of VR and teleoperation systems. We discuss one approach in particular as well as factors that may affect measurement accuracy for that specific method.

A.2 Technical Background

Achieving latencies that drop below a user's JND remains a significant challenge. There is a long pipeline of potentially unsynchronized processes that contribute to the full end-to-end latency (see Jacoby et al., 1996; Jung et al., 2000): from the moment the user initiates a movement until the moment the action is visualized on the screen, as shown in the simplified breakdown in Figure A-1. Each step has an individual, potentially

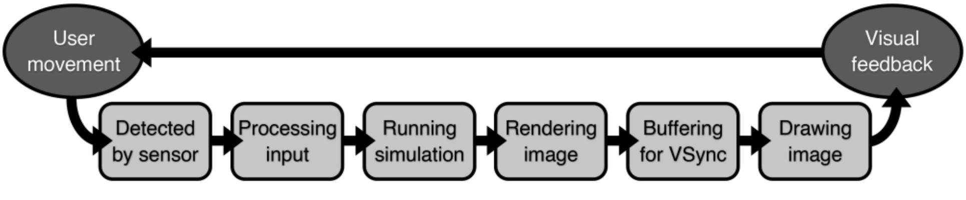


Figure A-1. Simplified system pipeline showing the path from the user's head movement to the visual feedback.

unsynchronized clock rate and intrinsic delay, contributing respective latency and frame-rate effects, much as various sinusoidal frequency components of vibration waves add to produce complex waveforms and beat frequencies. Additionally, if the effects of buffers introduced between the steps are not synchronized, they can increase the average end-to-end latency as well as introduce some variation over time.

The sensor delay depends on the physical sensing processes within the sensor itself as well as associated electronics and software routines that process the sensor information and transmit the output to the virtual-environment application. After receiving the motion signal, the application runs its simulation based on collected motion data and delivers the output to the graphics processing unit (GPU) that renders the image. The image waits in a frame buffer for the signal of the *vertical synchronization* (VSync) that synchronizes the read-out of the rendered image with the monitor's refresh and finally draws the resulting image on the screen of the VR display. It should be noted that *vertical synchronization* is a fixed term, although the synchronization in a head-mounted display (HMD) usually happens in a *horizontal* direction: State-of-the-art HMDs typically employ screens that were developed for the smartphone market with an upright orientation. In HMDs, these screens are implemented horizontally, leading to a rotated pixel refresh direction.

All steps, apart from the VSync, are necessary steps and present in all VR systems using an HMD. VSync prevents a refreshing of the image while an image is still being drawn on the screen (usually row by row, or—in the case of the rotated orientation in HMDs—column by column), buffering the newly rendered image. Disabling this feature would lower the end-to-end latency but also create a visual artifact known as *screen tearing*. The dynamic performance gain from omitting the VSync feature is controversially discussed in the VR and gaming community (see Bedikian, 2013; Blatt, 2018; Boxer, 2017; Wawro, 2011) since for some applications the negative impact of added latency and judder due to vertical synchronization outweigh the negative impact of screen tearing. Hence, designers of VR systems should pay attention to specific technical settings, such as frame rate, refresh rate, and buffering techniques, since those influence the extent of delay produced by the vertical synchronization. Bedikian (2013), Soomro (2015), and Tang (2015) experimented with different types of vertical-synchronization modes while varying system configurations and observed a latency increase of one to six frames (i.e., up to 100 ms) when VSync was enabled. For this reason, vertical synchronization should not be thoughtlessly included in the rendering software.

The last step in the processing chain involves drawing the image on the screen. The *response time* of the screen's physical pixel elements is pivotal for this step and is usually indicated as the time a single pixel on the screen requires to change from gray to white and back to gray again, consequently referred to as gray-to-gray (GtG) response time. Apart from causing additional latency, a long-lasting GtG transition is also one of the sources of a phenomenon called *motion blur*, which reduces the feeling of presence in virtual environments. The issue of slow pixel response times meanwhile belongs to the past, especially when taking modern OLED displays into account (Morrison, 2014). However, motion blur may occur even in displays with an instant pixel response due to a long pixel persistence, which will eventually disrupt the illusion of smooth motions within the virtual environment (Rejhon, 2018; Udiljak, 2016). A possible way to overcome or, at least, to reduce this sample-and-hold issue is through control of image persistence. Modern HMDs consequently use *low-persistence* screens that display the image only part-time: Pixels remain black between two consecutive frames. This short black interval is not perceived by the user in real time but provides higher motion clarity even at fast-paced virtual scenarios, reducing motion blur. Figure A-2 compares the effect when scenarios that contain fast head movements are displayed on an HMD at full-persistence mode and low-persistence mode. Although the persistence mode

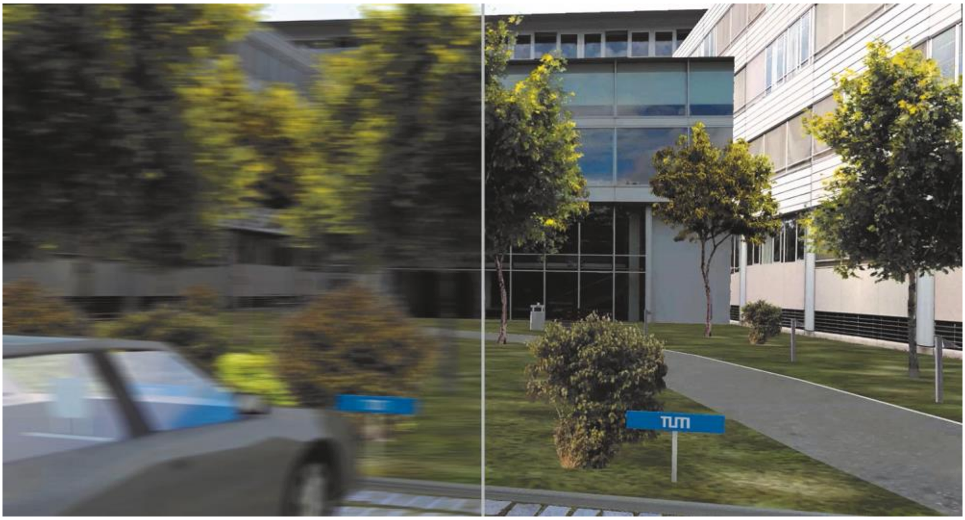


Figure A-2. Comparison of motion blur effects occurring in virtual scenarios with fast head movements: display configurations at full-persistence mode (left) and low-persistence mode (right).

is not affecting the system latency directly, it is relevant to the latency measurement technique introduced below.

Some VR system designers attempted to reduce latency by implementing predictive algorithms, such as extended Kalman filters that evaluate the tracked movement in real time and predict the position shortly in advance by using smart extrapolation (Adelstein et al., 2001; Jung et al., 2000; Merriken et al., 1988). Jung et al. (2000) demonstrated that the participants' abilities to detect some motion artifacts due to latencies of about 16 ms could be reduced to chance levels even by predictors that do not incorporate human-movement dynamics. Presumably, including such dynamics could extend the performance of such predictors. That said, these algorithms are not appropriate for scenarios in cases of turbulent and rapidly changing movements that may be difficult to predict and thus be prone to error (Allerton, 2009; Azuma & Bishop, 1995).

Despite recent technological advances in widely available systems, latencies within VR systems are not expected to disappear in the near future. Although all the technological components of these systems have become steadily faster and more powerful, the range and graphic detail of handled tasks have also expanded. For example, the improvements of the visual display quality for head-referenced systems, such as the visual resolution, the color gamut, the binocular field of view, and specialized visual rendering effects, can increase the time to finish and scan out an individual video frame, and thus add latency to a system that otherwise might be limited by the geometric complexity of a rendered scene. Additionally, the shift from wired data transfer to wireless, mobile systems can add delay components that are essentially exogenous to the rendering process (e.g., transmission and switching delays), thus increasing the end-to-end latency. Contemporary VR applications may also include communication technology, and thus they are also likely to be introduced into spatially widely distributed networked multiuser systems. Such systems have the potential to introduce significant latencies since these delays are like roundtrip delays experienced in teleoperations. Consequently, developers and, ultimately, government regulators will need the ability to conveniently assess and manage full system latency experienced by end users. From a VR designer's point of view, virtual-environment systems may be modified in ways specifically designed to decrease end-to-end latency (see Wilson, 2009). However, even modifications not directly related to latency improvement may yet affect the overall delay. Latency measurements for setups that undergo varying configurations are consequently needed, helping designers to systematically isolate elements that increase or decrease the end-to-end latency.

A.3 Related Work

A.3.1 Overview of Related Work

A variety of methods on how to quantify end-to-end latency and how to measure partial latencies have been introduced and debated in the past. In the following, some of these techniques for VR systems are outlined.

Liang et al. (1991) used a pendulum in combination with a video camera to determine the latency of a motion-tracking sensor. The camera monitored the pendulum, which was equipped with a tracking sensor. The sensor transmitted time-stamped measurements of the pendulum's position to the computer system while displaying the time stamps simultaneously on a screen within the camera's visual field. Subsequent playback of the videotape allowed for determining tracker latency.

Mine (1993) also attached an HMD motion-tracking sensor to a pendulum. The pendulum's low point of the swinging arc was tracked using an independent photodiode. The sensor's position signal was juxtaposed to the photodiode's indication of the low point on a digitizing oscilloscope, visualizing the delay of the tracking sensor and its associated signal processing software. An additional photodiode fixed on a monitor that replaced the HMD's screen allowed for determining the end-to-end latency.

Adelstein et al. (1996) initially focused on position-tracking-sensor delay by using a horizontally oscillating mechanical arm equipped with a rotary encoder that provided a very precise reference signal that was compared to the sensor signal. Ultimately, their laboratory was able to internally tap the VGA signal within the rendering computer to accurately measure full system latency, independently from specific display hardware, excluding the few milliseconds of CRT phosphor rise time and video frame scan (Hill et al., 2004). These last two components of full system latency are very stable and can be physically accurately known so that a corrected latency value can be achieved by simply adding them to the latency measurement.

He et al. (2000) strived to determine the end-to-end latency using a video camera. They recorded a cross-shaped controller being moved in the real world, with a virtual cross being displayed simultaneously on a screen behind the controller. By examining corresponding inflection points of the real and virtual motions, the full system latency could be determined, though the low frame rate of the video made identification difficult and error-prone.

Steed (2008) used an ordinary 25-Hz video camera to record a motion tracker and an LED, both attached to a pendulum. An algorithm determined a sinusoidal signal by fitting a sine function to corresponding successive positions on frames of the recorded footage. This signal was then compared to the signal obtained by the tracking sensor displayed in the camera's visual field, with the phase difference revealing the sensor latency.

Di Luca (2010) used two light-sensing devices in order to measure the end-to-end latency of an HMD. One was attached to the screen of the HMD and another one on top of the HMD, facing a desktop monitor. The desktop monitor and HMD screen both displayed a gray-gradient image that altered the gray shade sideways. The two light-sensing devices captured the current gray shade simultaneously on the respective display while the HMD was being moved in front of the monitor. An analysis carried out afterward allowed for comparing the timed alteration of captured gray shades with one another, thus determining latency by identifying the offset.

Friston and Steed (2014) strived for a fully automated measurement approach. They captured the real movement of a marker-equipped computer mouse and the consequent cursor motion on a desktop display using a camera and applied an automated frame-counting algorithm to determine the delay between the two actions.

A.3.2 Discussion of Related Work

Although the techniques described above provide useful historical reference points regarding latency measurement in VR systems, they raise several issues: Most of the described latency measurement techniques required tedious preparations, specialized equipment, and a detailed technical understanding of the system that surrounds the virtual environment. Some of the techniques even require disassembling the VR system. For example, it may be necessary to remove the motion-tracking sensor from the HMD so that it may be mounted on a pendulum. Researchers should also be aware that some of the techniques only permit for measuring the delay of the motion-tracking sensor and associated software, not the VR system's full end-to-end latency that also involves other elements (see Section A.2).

Furthermore, employing a pendulum or a mechanical arm restricts the variety of movements that the VR designer can examine. A thorough latency assessment requires, however, the measurement of a variety of movements that may consist of rotational and translational motions. A VR system may employ several sensor types for the different frames of reference, which may produce different latencies. In addition, human users are more likely to be sensitive to delayed egocentric head movements than to the effects of delayed limb movements (e.g., hand movements) or delayed motions of virtual objects. These aspects should be considered when aiming at a complete assessment of latency and its impact on the user.

The sensors that are used to measure latency in VR systems by detecting the real and the subsequent virtual movement are subject to measurement delays themselves. Hence, sensors must be selected so that their internal latencies or noise do not significantly contribute to the overall measurement. Fortunately, since they are usually used to measure the onset and subsequent offset of motion, their inherent latencies may be assumed to cancel, provided that measurement noise can be ignored.

When measuring an end-to-end latency for system design purposes, the used hardware and software components must be the same as those intended to be actually employed and not just convenient stand-ins. For example, latencies that are measured on a CRT desktop monitor as a substitute for an LCD screen of an HMD (see Mine, 1993) may lead to different delays because of physical differences in the display technology.

An application-oriented measurement of latency with realistic virtual scenarios is needed in order to measure a behaviorally relevant end-to-end latency: Neither a flashing square (Mine, 1993), nor a two-dimensional cross (He et al., 2000), nor a simple gray-gradient texture (Di Luca, 2010) is necessarily an adequate substitute for a complex virtual environment in which detailed rendering of self-interacting content introduces specific latency components. In retrospect, it is not surprising that many of the older evaluation techniques did not test latency with the highly detailed virtual environment that could be of contemporary interest. The digital rendering and display performances of older systems were markedly inferior in many aspects to present rendering, motion capture, and user position sensing. In fact, they were orders of magnitude inferior to present capabilities. Furthermore, in the past, the ultimate applications for VR were not well enough defined for developers to design or even understand the specifically required system performance. However, considerable variability of the end-to-end latency may be observed, depending on factors like resolution, the complexity of the virtual representation, and communication delays. With rendering applications partially accounting for high computational expenditure, some VR designers work on developing systems with a high-quality image for the user's foveal vision while the peripheral vision is provided with a lower-quality image, thus taking advantage of the degradation of peripheral visual acuity and effects of saccadic suppression (Albert et al., 2017; Hitchner & McGreevy, 1993; Kiyokawa, 2006; Luebke & Hallen, 2001; Reddy, 1995).

These observations emphasize the necessity of running a complex virtual environment during the latency assessment, with the applied complexity being representative of the VR system's intended use. The different measurement methods described above involved disassembling the setup, restricting investigable motions, substituting hardware components, and replacing a potentially complex virtual environment with simple graphics that were experienced in a controlled laboratory environment. Consequently, the conditions under which latency has been measured may lead to values that are not necessarily representative of fielded systems.

In summary, a latency measurement technique for VR systems should ideally measure the genuine end-to-end latency, incorporating all actual hardware and software components, using a representative virtual environment while enabling an assessment of a variety of rotational and translational motions. Disassembling the VR setup for the latency assessment should generally be avoided, unless the goal of the latency measurement is to determine partial latencies within the system pipeline.

A.4 A Simple Latency Measurement Technique

The concept of the latency measurement method that we suggest is reasonably accessible and straightforward: A camera that is capable of recording at a high frame rate (e.g., 240 fps) is directed at the screen of an HMD displaying a virtual environment (see Figure A-3). The camera is positioned to include the surrounding HMD housing, so the movement of its contours is clearly visible in the video frame. While a specialized high-speed camera may be utilized for such applications, smartphones and action cameras nowadays are also capable of recording at sufficiently high frame

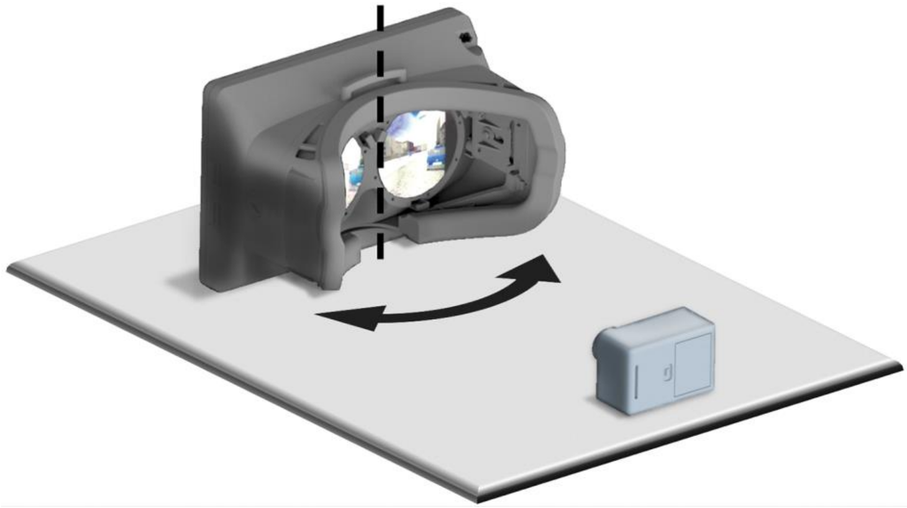


Figure A-3. Latency assessment configuration.

rates. The HMD is then in some way moved (e.g., given a manual pulse of motion) to produce correlated motion of the housing and corresponding contour motion within a concurrently visible virtual environment. For demonstration purposes, a rotational motion was chosen. Thereafter, a frame-by-frame analysis of the recorded footage will allow determining at what video frame the movement of the HMD hardware initiates, and at what frame the virtual environment on the HMD's screen begins to reflect its motion. This probing motion for the given demonstration is most conveniently produced by rotating the HMD so as to rotate the virtual environment around a corresponding rotation axis. Knowledge of the footage frame rate allows the investigator to determine the end-to-end latency between the initiation of the HMD's movement and the reaction on the screen by counting frames between these two events, using conventional video editing applications.

For the presented method, a human evaluator identifying and counting the frames is a convenient and easy solution but may be prone to human error. However, designing an evaluation software that identifies and counts the frames automated is of considerable difficulty, especially since the presented method is supposed to be applicable universally for a wide range of different scenarios. The design of such software will require a considerable amount of time and potentially time-consuming adjustments for each new scenario, all of which themselves could be prone to error. A human evaluator, on the other hand, may assess the recording without further ado.

Although the discussed demonstration specifies the usage of an HMD while assessing rotational motions, the same concept may be applied for a wide range of systems, including CAVE-based systems, and a wide range of motions, including translational movements. With appropriate arrangements, the method can also be applied to teleoperation systems, provided that the sending and receiving environment are both within the camera's visual field. Alternatively, if the sender and receiver are not available in the same physical space, the round-trip latency may still be measured from an image transmitted from the remote site. Note that VR systems are essentially a simulation of teleoperation systems in which the imagery, usually coming from a remote camera, is replaced by a computer graphics simulation.

A.5 Method Demonstration

A.5.1 Equipment and Participants

A.5.1.1 Experimental Setting

The HMD used for this demonstration of latency measurement was an *Oculus Rift Development Kit 2* (Oculus VR, Menlo Park, CA, U.S.). The HMD's display runs at 75 Hz and provides a resolution of 1,920 px × 1,080 px. The HMD's tracking sensors run at 1 kHz and transmit the collected data approximately every 2 ms. For the latency measurement procedure, the Fresnel lenses mounted on the HMD's screen were removed to provide the camera with an unobstructed view of the screen. The HMD is equipped with the VSync feature as well as the low-persistence mode (see section A.2). Both features were activated during the experimental assessment. In the low-persistence mode, the display's pixels are illuminated for approximately 2 ms and go then off until the next image is loaded. Since the image is loaded column by column, the low persistence never affects the full screen at the same time: The image is loaded from right to left (from the user's point of view) and also continuously disappears in the same direction. As a result, part of the screen appears black when taking an external snapshot of the screen, as done during our experiment. Figure A-4 shows such a snapshot and also exposes the rolling shutter of the camera's sensor: The image loading progress on the HMD's screen appears skewed instead of vertical.

The system did not use a virtual environment that was provided by the HMD's manufacturer. Instead, the HMD displayed a customized virtual environment that was generated by the *Silab* simulation software (WIVW,

Figure A-4. View on the HMD's screen without the Fresnel lenses: The snapshot captures the dark pixels in the middle due to the low-persistence mode. The virtual loading progress appears skewed, which is due to the recording camera's rolling shutter. Note that this is a cropped image since the full image also displays the HMD housing.

Würzburg, Germany). This application, initially developed for driving simulation, was modified to run a VR simulation from the viewpoint of a pedestrian. The application uses rotational motion data that are tracked and provided by the HMD's built-in motion-tracking unit to render the virtual view in accordance with the user's rotational head movements. For the user's translational movements, the VR system uses an external video-based motion-capture system by *Vicon* (Vicon Motion Systems, Yarnton, U.K.). However, for the demonstration of the latency measurement method, the assessment was limited to rotational movements only.

The simulation is powered by a computer equipped with the *Intel Xeon E5-1620* (Intel, Santa Clara, CA, U.S.) central processing unit (CPU), running at 3.60 GHz. The GPU was a *GeForce GTX 670* (Nvidia, Santa Clara, CA, U.S.) with 2 GB of memory and 1,344 CUDA cores, running at a 915 MHz base clock and boosting up to 980 MHz. It should be noted that with the use of a custom simulation application, the latency assessment does not necessarily represent the HMD's optimized rendering performance as provided by the manufacturer.

A.5.1.2 Analytical Tools

The observing camera used for the latency assessment was an action camera, model *GoPro Hero3+ Black Edition* (GoPro, San Mateo, CA, U.S.). This consumer camera allows for recording at 240 fps with a resolution of 848 px × 480 px. Thus, the achievable temporal resolution (i.e., the time between two frames) is approximately 4 ms.

The recorded video footage was converted into single frames (i.e., image files) using the open-source *VirtualDub* video processing utility (freeware by Avery Lee). When converting video footage into image files, the actual output of exported frames should be verified to be consistent with the video footage's nominal frame rate. This is necessary because some processing software may not be able to handle high frame rates accurately.

The extracted image files can then be evaluated using an image viewer software, such as the open-source *nomacs* software (Technische Universität Wien, Vienna, Austria), which is a lightweight utility capable of switching rapidly between frames. The lightweight feature is of utmost relevance during the analytical procedure: This way, the evaluator can easily observe changes (i.e., movements) between two consecutive frames by switching quickly back and forth between the frames. Common preinstalled image viewers such as *Microsoft's Windows Photo Viewer* tend to buffer when navigating rapidly through images. A strength of the *nomacs* image viewer is also its convenient zoom control, which may be manipulated using the *up* and *down* arrow keys on the keyboard. This feature allows maintenance of the same zoom level during continuous navigation through the video images (*left* and *right* arrow keys). Maintaining the same zoom level is crucial for achieving a superposition of the individual frames, which facilitates the detection of changes.

The software examples suggested here are just a few options. Generally, any application that can play a video file frame by frame may be used, though lightweight programs are recommended to avoid buffering effects.

A.5.1.3 Latency Evaluators

Twenty latency evaluators (15 male, 5 female; $M = 25$ y/o, $SD = 3.5$) took part in the experimental assessment. The latency evaluators were randomly selected and were mostly university students. The majority did not possess any prior experience with VR systems. Because the evaluators were essentially doing method testing by identifying the first frame of physical motion and the first frame of corresponding virtual motion, no personally identifiable information was collected during the experiments. Participants were not compensated, and their names were not associated with any of the measurements they made. After some minimal instructions detailed below, each evaluator repeated the same ten latency assessments five times in total, with a minimum of two days apart between two runs to reduce possible memory effects that could bias the results.

A.5.2 Experimental Procedure

A.5.2.1 Latency Recording

Ten latency sequences were recorded, exported to image files, and reduced to sequences of 200 frames each. The sequences contained the event in which the HMD was given a pulse of rotation around its vertical axis and also included the consequent virtual movement on the HMD screen. The angular rotation rate of the HMD was approximately 150°/s. The rotation direction was chosen so that the right screen edge of the HMD turned toward the camera. This facilitated the observation of virtual movements, which occurred first on the right screen edge due to the loading direction of the virtual image. The virtual scenery showed an urban street scenario with houses and parked cars, as well as some moving vehicles that occasionally passed on the opposite lane (see Figure A-4).

For the present demonstration of the method, the technical parameters (such as rotation direction, rotation rate, observation angle, light conditions, camera settings, and virtual scenario) remained identical for all recordings. Thus, the ten image sequences were hard to differentiate from each other visually. The different measurements were expected to come with an inevitable variance in the determined latency value due to the various unsynchronized latency-inducing processes involved in the VR system (see Section A.2).

A.5.2.2 General Evaluation Assignment

Each latency evaluator was given the same instructions. First, the phenomenon of latency and why it occurs was briefly explained to the participants. The first two image series from a total of ten served solely for familiarization and allowed the participants to get a feeling for the task at hand, and were thus not analyzed afterward. Simply put, the participants were instructed with the following task to repeat for every series of pictures:

- Navigate through the image series and identify the first frame in which the HMD housing starts moving.
- Continue navigating and identify the first frame in which the virtual environment shifts sidewise

A.5.2.3 Instructions for Identifying Relevant Frames

Most of the participating latency evaluators did not possess prior technical knowledge or any experience regarding the task at hand. In fact, many did not know what “latency” is beforehand. They were, consequently, provided with the following helpful hints.

HMD housing movement:

- It will be helpful to first determine the initial occurrence of the movement by fast-forwarding through the image series to identify the approximate location. There, a frame-by-frame analysis is imperative. Progressively comparing two consecutive frames at a time, by switching back and forth rapidly, will help to identify the frame of the first movement.
- With two frames being solely 4 ms apart from each other, there are no major movements to be expected.
- The texture appearance of surfaces may alter across the frames due to changing light. For this reason, it is helpful to concentrate on

high-contrast edges of the HMD housing, making it easy to observe a physical movement.

Corresponding screen contour movement:

- A simple way to note virtual movements is to watch virtual objects close to the right-hand edge of the HMD's screen and to identify when their relative distance to the screen edge changes.
- Although the virtual environment on the HMD's screen is progressively loaded from right to left, the loading image represents a static instant of the virtual environment and does not change in the middle of the loading progress because of the VSync mode. This virtual instant is progressively loaded throughout about 13 ms (refresh rate of the HMD), with the progression being observable on about three consecutive frames (each approximately 4 ms apart). Consequently, changes in the virtual environment may solely occur once a new virtual image is loaded on the HMD's screen, which is approximately once every three recorded frames.
- Due to the low-persistence mode, the progressively loaded image disappears before the next image appears, leaving the pixels black. This makes it easy to identify when the next virtual instant (i.e., a virtual image) starts to load.

A.6 Demonstration Results

The latency values for each sequence (out of 8), experimental run (out of 5), and latency evaluator (out of 20) were computed by determining the number of frames between the identified first real movement and the identified corresponding virtual one. This counted frame number was multiplied with the reciprocal of the camera's frame rate to determine the latency values. There were two sources of variability that affected the latency measurement: Some variability arose due to the technical components, either within the

VR system itself or within the tools used for the measurement, and some variability was due to the variable performance of the human evaluators.

A.6.1 Technological Uncertainty of Presented Method

A.6.1.1 Underlying Latency Fluctuation

Repeated measurements of latency using the proposed technique can vary over time because the various components contributing to the overall system are not necessarily synchronized (see Section A.2). It should be noted that this inherent variance is distinct from the variance attributable to the measurement technique itself and will occur regardless of the chosen latency measurement method. Accordingly, an accurate latency assessment must be based on repeated measurements. The latency values across the eight different assessed sequences showed a noteworthy fluctuation with a standard deviation of 6.3 ms and maximum deviations of ± 12 ms from the marginal mean that was at 84 ms (see Figure A-5). The standard errors due to human variance in judgments for each of the sequences varied between 0.6 ms and 1.4 ms.

A.6.1.2 Sampling Rate of the Measuring Sensor

The sampling rate of the measuring sensor determines the underlying uncertainty associated with the technological accuracy of the proposed latency measurement technique. For the equipment suggested here, the sampling rate is determined by the camera's video frame rate of 240 fps. Thus, successive frames are 4 ms apart. This means that—due to the sample-and-hold nature—a certain event that is observed on a specific frame could have taken place, theoretically, up to 4 ms earlier, at which point the previous frame had been taken. Since this applies to the observation of the HMD housing movement and the observation of the consequent virtual movement, it doubles the variability range of the measurement but also

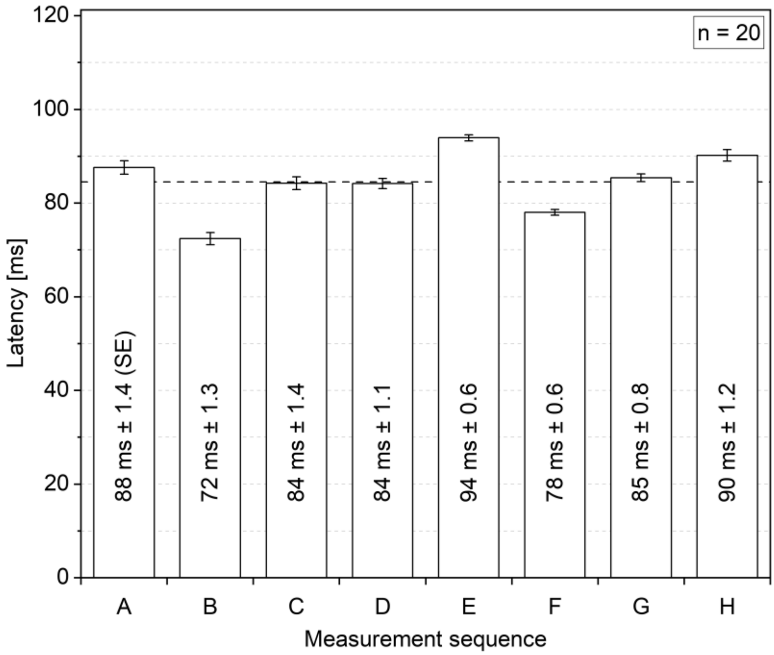


Figure A-5. Determined latency values for eight different measurement sequences for the same VR system under identical conditions: Each sequence was assessed by 20 evaluators with the error bars indicating the standard error due to variance in human judgments. The bold dashed line at 84 ms represents the marginal mean across the eight sequences and by that the latency value to be attributed to the system.

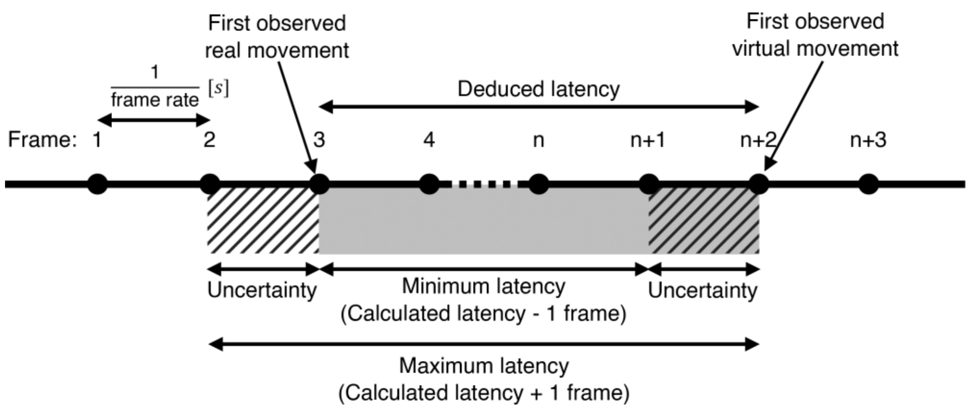


Figure A-6. Sample-and-hold-induced uncertainty of the latency measurement technique.

cancels out the deviations on average. This is because the first part adds a purely positive deviation, whereas the second part adds a purely negative deviation of the same magnitude (see Figure A-6). As a result, the determined end-to-end latency value will involve a range of uncertainty corresponding to the reciprocal value of the frame rate, thus within ± 4 ms in the given example.

A.6.1.3 Inherent Delay of the Measuring Sensor

As discussed in Section A.3.2, the inherent delay of the sensor measuring the latency may affect the determined value. The concept presented here uses nominally one single sensor (i.e., the observing camera) to capture both the real and the virtual movement. However, it should be clarified that technically the image sensor in a camera consists of many million “sub-sensors”: one *photosite* for each pixel. The camera sensor operates either with a *global shutter* (i.e., all photosites are exposed simultaneously) or with a *rolling shutter* (i.e., photosites are exposed row by row). Consequently, the different pixels of a single picture do not represent precisely the same instant when using a rolling shutter. Most current consumer cameras are equipped with complementary-metal-oxide-semiconductor (CMOS) sensors that most commonly—but not exclusively—have a rolling-shutter circuitry (Adler, 2016; Kozacek et al., 2018; Paul, 2016). Depending on the exposure time and captured motion, rolling-shutter exposure may lead to predictable distortion. The rolling shutter usually travels in a top-down direction, meaning that the largest temporal difference is between the pixels captured at the upper image edge and the lower image edge.

If a frame is evaluated at two different areas (e.g., one area showing the movement of the HMD housing and another area the movement of the virtual environment), the two areas might show two slightly different instants. The rolling shutter duration for one frame usually corresponds to the reciprocal value of the footage frame rate. The relative pixel delay between the two areas can be calculated using the following formula:

$$relative\ pixel\ delay = \frac{pixel\ row_{VE} - pixel\ row_{HMD}}{vertical\ pixel\ resolution} \times \frac{1}{frame\ rate}.$$

Formulated in words, the pixel row that shows the HMD housing movement and the pixel row that shows the corresponding virtual movement are to be determined, with the row differential being divided by the sensor's number of rows and multiplied by the reciprocal value of recorded frame rate.

A camera recording at 240 fps will have a rolling-shutter duration of approximately 4 ms top to bottom. Practically, this means that a movement of the HMD's housing can be observed, for example, at about a quarter below the upper image edge, while the area in the frame center captures the HMD's screen showing the virtual movement (see Figure A-7). These areas are temporally apart by one quarter of the shutter time, thus about 1 ms for video recordings at 240 fps.

Consequently, the inherent relative delay of the camera's sensor may be neglected for latency measurements at high frame rates. For lower frame

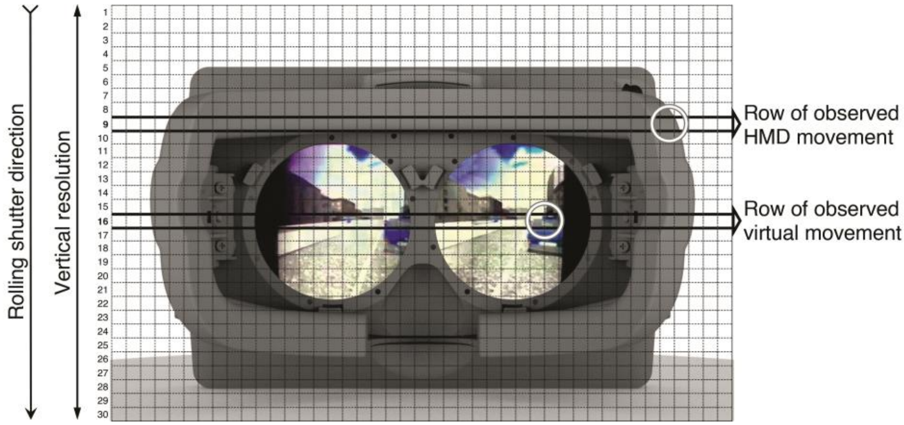


Figure A-7. Pixel rows of a frame are recorded with a temporal shift due to the camera sensor's rolling shutter circuitry. Consequently, different vertical rows of the same snapshot do not show the same instant but are usually a tiny fraction of a second apart.

rates (or higher accuracy), the formula above may be applied and integrated into the calculation of the end-to-end latency. Though evaluating real and virtual movements on a similar vertical level of the image may be the most efficient way to counteract the influence of the rolling shutter on the latency value.

A.6.2 Analytical Uncertainty of Presented Method

When human evaluators judge the recorded footage, certain variances in their decisions related to identifying real and virtual movements may be expected.

A.6.2.1 Variance Between Evaluators

An analysis looking into the variance between the 20 evaluators (with every one of them having evaluated the same eight recordings in identical order) revealed a standard deviation of 4.1 ms across the individually determined mean latency values and thus a standard error of 0.9 ms across the 20 evaluators, with maximum deviations from the marginal mean (84 ms) being at ± 10 ms.

A.6.2.2 Variance Within Evaluators

Furthermore, the occurring variance when the same evaluator repeatedly assesses the same recordings was investigated. In this context, the determined mean latency for each run (containing eight measurements) was compared to the evaluator's individual overall average for all five runs. The mean-variance across these five runs for all 20 evaluators suggests a within-subjects standard deviation that is 1.9 ms on average, leading to an average standard error of 0.9 ms for an evaluator repeating the latency judgments five times. The maximum within-subjects deviations remained within ± 7 ms. Figure A-8 shows the variability between different latency evaluators, while

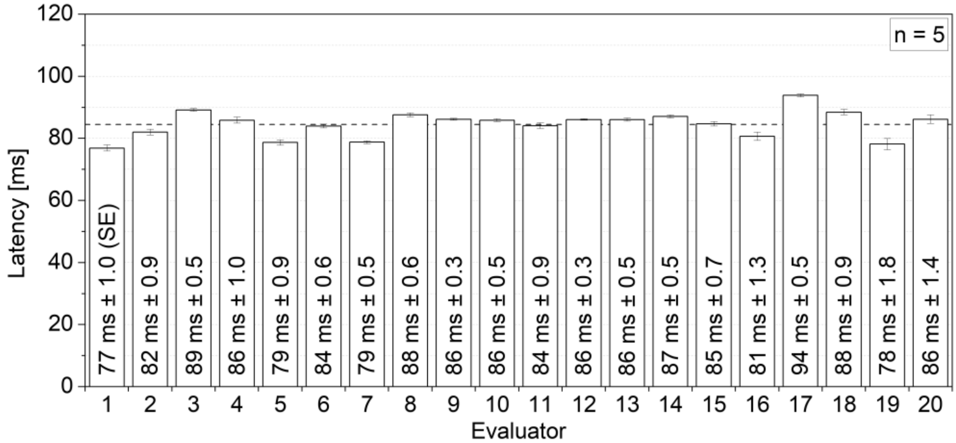


Figure A-8. Average latency values for 20 different evaluators: Each value is based on the mean value of five repetitions of the experiment, in which eight measurements were to be evaluated. The error bars indicate the standard error across the five repetitions of the experiment.

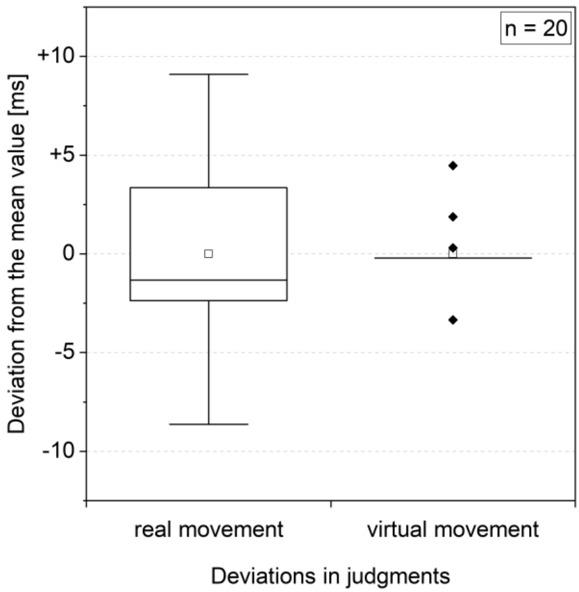


Figure A-9. Deviations from the marginal mean value regarding the points in time at which first movements in the real and virtual environments were observed by 20 evaluators, visualized with box plots.

Box-and-whisker plot structure: Values between the lower and upper quartiles are represented by a box, while whiskers identify estimates within 1.5 times of the interquartile range (IQR) of the lower and upper quartiles. The horizontal line in the box shows the median, and the small square shows the arithmetic mean. Outliers are plotted with diamonds.

the error bars indicate the standard errors across the five repetitions of the experiment.

A.6.2.3 Assessment of Real and Virtual Motion Onsets

Since the evaluation of a single recording consists of identifying the onset of the real movement and its subsequently linked virtual movement, the results of the two different steps were analyzed separately to eventually ascertain differences regarding their impact on the variance of the determined latency values. To this end, the average determined values of each evaluator for the real motion and the virtual one across the five repetitions × eight measurements were compared to the marginal mean values across all evaluators, with the deviations being plotted in Figure A-9.

The distribution shows that an actual variance of the judgments occurred almost exclusively for real motion. In fact, when looking into the individual judgments of the eight different measurements, 97.5% of the virtual motion values across all evaluators were identical.

A.7 Discussion

A.7.1 Costs and Benefits of the Method

A.7.1.1 Simplicity of the Method

The latency measurement technique that is presented here differs from earlier reported methods (see Section A.3). Many previous methods necessitated disassembling the VR system, allowed for measuring only partial system latencies, were suitable solely for one specific movement, or required expensive equipment and time-consuming preparations. Most importantly, none of the methods discussed in Section A.3.1 necessarily allowed for the actual virtual environment to be investigated but substituted the virtual environments with graphical representations that facilitated the latency evaluation. The easy-to-evaluate virtual reference environments in which the latency used to be measured in the past can lead to latency values that are not necessarily representative.

Explained with an analogy: Just as fuel-consumption estimates by car manufacturers, which are based on controlled and optimized laboratory vehicle testing, may not be representative for real-world driving experiences in actual traffic, so the latency testing on isolated virtual-environment systems under optimized and artificial conditions may not represent latency conditions in realistic application-oriented scenarios.

How important it is to determine latency within the actual hardware and software environments of intended application can be seen when comparing our experiment to the one by Raaen and Kjellmo (2015), who used the same HMD hardware in a latency performance testing: Our evaluation used a detailed virtual environment with actual movements being tracked, leading to an average latency value of 84 ms, varying between 72 ms and 94 ms. Raaen and Kjellmo, on the other hand, displayed on the same HMD a plain black "virtual environment" that switched to white as a reaction to any physical movement, capturing this event with a light sensor. They measured an average latency value of 41 ms, varying between 35 ms and 45 ms. When turning off the VSync mode, their latency value even dropped to 4 ms on average, varying between 2 ms and 5 ms. Such a low, potentially unrepresentative latency appears to be solely a sensor delay measurement (see Adelstein et al., 1996): The tracked movement did not require three-dimensional processing, no simulation application required computational expenditure, and the screen switched simply to white without the need of rendering a scene. It is also noteworthy that a small parameter change, such as disabling the VSync mode, can result in a significant change of end-to-end latency.

The method suggested here is reasonably easy to apply, without requiring a deep technical understanding of the specific hardware and software in use. The low-persistence mode that enjoys increasing popularity in VR benefits the evaluator of the footage in allowing to easily discern between different images on the VR display. The method was straightforward and easy to explain to the 20 evaluators, who had no prior knowledge in the field. The time requirement on the evaluators was low, necessitating less than one minute, on average, per evaluation of one latency sequence that included the detection of the real movement and the virtual one (M = 47 s, SD = 18.8, n = 184). The minimum equipment involves a high-frame-rate camera and appropriate software to analyze the recordings. Nowadays, such cameras are readily available since they can be found in many current smartphones. Moreover, all necessary evaluation software is available as freeware.

When using external light sources to illuminate the latency measurement setup, investigators must be aware that these light sources may oscillate along with the frequency of their power source. Although invisible to the human eye in everyday life, this oscillation may cause visible varying light intensity and disturbing artifacts on recordings at a high frame rate (see Steed, 2008; Wu et al., 2013).

A.7.1.2 Number of Evaluated Recordings

Since the system latency varies to some extent (see Section A.6.1.1), repeated recordings are necessary for a full system characterization, regardless of which latency measurement technique is being used. The experiment here showed that the mean latency value began to converge to a fixed value of ± 1 ms after aggregating five recordings (see Figure A-10).

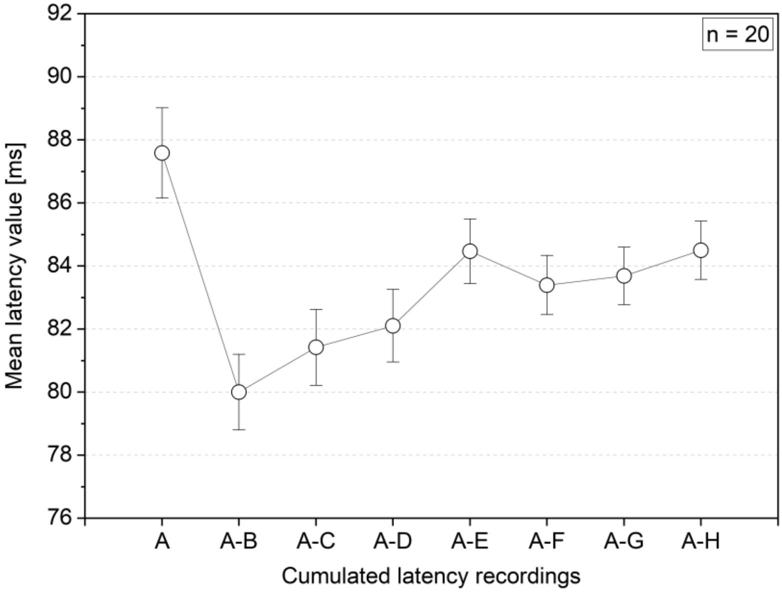


Figure A-10. Mean latency depending on the number of cumulated latency recordings: The error bars indicate the standard error due to human analyzing variability across 20 latency evaluators.

A.7.1.3 Sources of the Measurement Variance

The measurement technique suggested here is reasonably accurate in terms of the technical precision of the measuring tools, with much of the small uncertainty arising from the sample-and-hold nature of video footage (see Section A.6.1.2) and few from the delay between the different pixels across a frame due to the camera's rolling shutter (see Section A.6.1.3). These measurement uncertainties are combined approximately ± 5 ms for recordings at sufficiently high frame rates (≥ 240 fps). Thereby, the uncertainty due to the sample-and-hold mode and the delay between the pixels linearly decrease when the frame rate increases. The experiment showed a standard deviation of 4.1 ms across the assessments of 20 evaluators in terms of the variability of the latency outcome due to different human evaluators (see Section A.6.2.1). When the same evaluator repeated the latency assessments, the values came, on average, with a standard deviation of 1.9 ms across the five repeated assessments (see Section A.6.2.2).

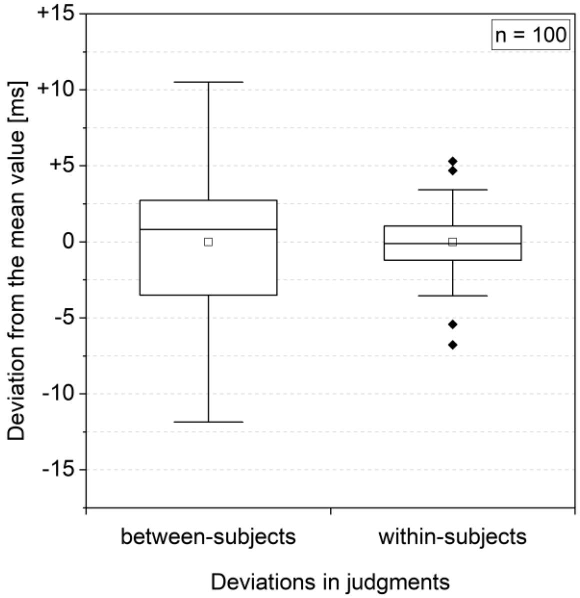


Figure A-11. Deviations from the respective mean latency values between different evaluators and within evaluators when repeating the evaluation: The deviations are based on 100 values deduced across 20 evaluators who repeated the experiment 5 times. For the structure of the box-and-whisker plots see Figure A-9.

Figure A-11 uses box plots to summarize the variation of measured latency values between the different evaluators and between the five repetitions of the experiment (i.e., within evaluators).

A.7.1.4 Evaluation Variability

Figure A-9 shows that the human-induced variance of the determined latency value was due to the difficulty in judging real motion, while the virtual motion was judged identically across all evaluators except for a few sporadic outliers. This may be explained by the nature of the task: Real motion is a continuous movement that is captured at the camera's frame rate. In other words, the evaluator had to observe changes between two frames (i.e., four-millisecond movements) in the real environment. The virtual motion, on the other hand, is a discrete, sample-and-hold movement that does not rely on the camera's frame rate but the refresh rate of the

HMD's screen, displaying a static moment of the virtual environment every 13 ms in the experiment presented here. It is more challenging for evaluators to detect changes that took place over a period of 4 ms than those of linked motions over a period of 13 ms, considering that the rotation rates in the real and virtual environments were equal.

Several strategies may reduce the variance of the determined latency value due to human analytical variability. First of all, it should be noted that the participants in the experiment did not have any prior experience with regard to the task at hand nor much training. The variance between the evaluators tends to decrease with an increased number of evaluated recordings (see Figure A-10). This suggests that some more practice and familiarization with the procedure would eliminate outliers such as observed for a few evaluators. Furthermore, a VR designer can use a group of evaluators to improve measurement reliability. Given the nature of the judgments the evaluators made in our experiment and their apparent ability to benefit from training, the focus could be on one well-trained evaluator, who is familiar with the task and evaluates at least five to ten sequences. Afterward, for verification purposes, an additional two trained evaluators could be used to confirm the latency measurement. Automated frame counting that is detached from human evaluation, such as applied in some other techniques that relied on easy-to-measure environments (see Friston and Steed, 2014; Sielhorst et al., 2007; Wu et al., 2013), are unlikely to benefit the method presented here, for reasons briefly discussed in Section A.4.

A.7.1.5 Summary of the Presented Method

The presented method may be considered even more precise in case the relative latency difference associated with various design options (i.e., the variation of the VR system's latency due to modifications) is of interest (rather than the absolute system latency): The informal observations suggest that repetitive test-retest variability for the same evaluator provide a human-induced standard deviation below 2 ms. However, evaluator performance must first be established to be asymptotic and not susceptible to the effects of further training or fatigue, when repetitive measurements from one single evaluator are used to track system performance changes.

Future investigations may validate the absolute accuracy of the presented method by comparing the evaluators' findings with a ground truth. This could be done, for example, by adding specific artificial delays of known duration to the system and evaluate whether these delay changes are accurately measured during the evaluation process (see Sielhorst et al., 2007). Furthermore, future validation studies may involve a wider range of different VR systems and virtual environments, as well as varying viewpoints.

A variety of additional measures can be taken to improve the accuracy of the presented latency measurement technique and to facilitate the evaluation process. Several possible refinements are discussed in the following section.

A.7.2 Possible Refinements of Described Method

A.7.2.1 Characteristics of Recording Camera

The technological measurement uncertainty of up to ± 5 ms for camera footage at a frame rate of 240 Hz (as used in our demonstration) can be further reduced by using cameras with higher frame rates. For example, current high-end smartphones can go as high as 960 fps at a resolution of 1,920 px × 1,080 px, such as the *Samsung Galaxy S9* or *S10* (Samsung, Seoul, South Korea), outperforming the camera's parameters in the demonstration by factor four for the frame rate and factor five for the image resolution. Such cameras would reduce the hardware-based uncertainty down to approximately ± 1 ms. However, an increase in frame rate may also lead to an increase of the human variance when evaluating the footage: A higher frame rate implies smaller observable movements between two frames, making it harder for human evaluators to detect changes. For this reason, not only the frame rate but also the image resolution and light conditions are relevant parameters to be considered when recording such latency sequences.

A.7.2.2 Background Contrast

A low-contrast background behind the HMD can complicate the visual detection of movement in the real environment. Therefore, a high-contrast backdrop behind the HMD may be added to diminish analytical difficulties regarding the moving contours.

A.7.2.3 Guided Movements

Although it is possible to spin an HMD directly by hand on top of a smooth table surface, as done in our demonstration, this manual procedure does not ensure an identical motion pattern for each recording. Though it is not critical to maintain the identical motion pattern, since similar motion patterns will likely produce the same latency, it may simplify the latency recording and the evaluator's judgments if some constraints are introduced to make repeated movements more similar. Such a constraint could be a simple turntable or a rail system for rotational and translational movements, respectively. A combination of rotational and translational movements may be achieved, for example, by mounting the HMD off-center on a turntable. Tested movements could even be automated by using actuators if higher repeatability of specific motion dynamics is desired, although the benefit of such expenditure may be questionable.

A.7.2.4 Facilitation of Real Motion Detection

The human variance in latency evaluations was caused by the evaluators' limited abilities to identify tiny changes observed between two frames that differ in only a few milliseconds. While there is the possibility to increase the rotation velocity to facilitate detection of motion, this option has its limitations when a controlled rotation is to be maintained, and the camera is required to obtain a sufficiently expansive view of the turning HMD's screen.

A solution may include the installation of a device that visibly signals to the evaluator when the real motion started: This can be easily achieved by simply mounting an LED *on* the turntable with an interrupter switch being mounted *next* to the turntable, interrupting the circuit the moment the turntable starts moving and by that turning the LED off. A typical LED has a turn-off time in the tens of a nanosecond and can thus indicate the start of the rotation without delay. Such an LED indicator will make the observation of tiny motions in the real environment unnecessary and benefit the evaluator, especially at very high frame rates (>> 240 fps).

A.7.2.5 Refined Virtual Motion Detection

Overall, the detection of virtual movements at 75 Hz turned out to pose no difficulties for the evaluators. That said, additional assistance can be provided to the evaluator when the VSync mode is turned off during latency recording. A screen tearing of the virtual image will indicate to the evaluator that a virtual "movement" has taken place (see Figure A-12). Since the virtual image is loaded at a constant rate from right to the left of the HMD's screen, it is possible to calculate the remaining time until the virtual image with the screen tearing has finished loading. This is the time the VSync mode would have theoretically additionally delayed the system under optimized conditions. Note that observing the image loading progress and calculating

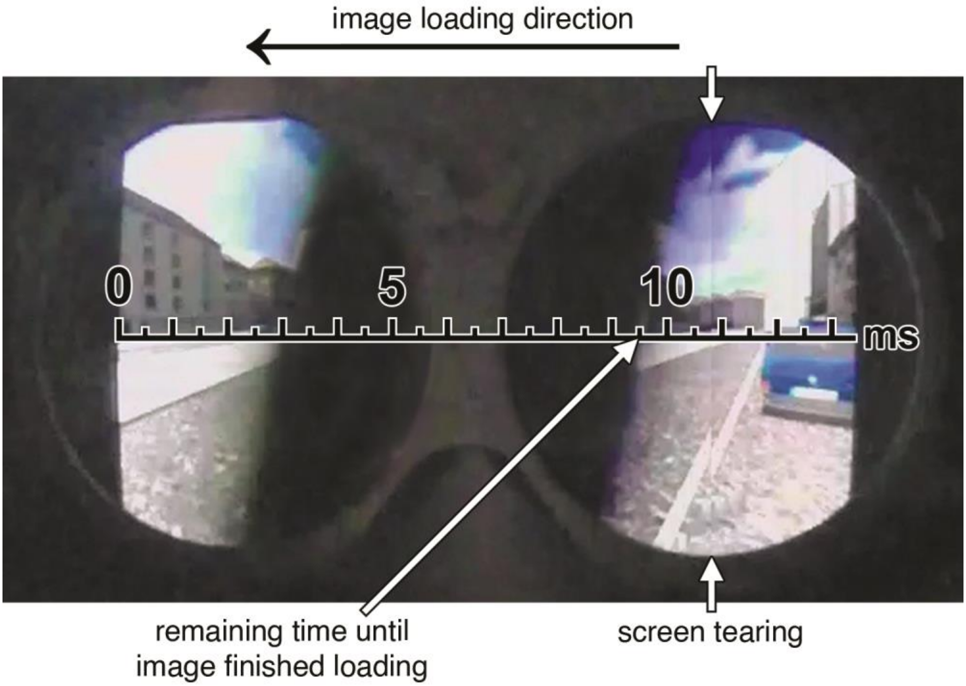


Figure A-12. A screen tearing on the HMD's screen indicates a virtual movement. The remaining loading time of the virtual image at that moment can be inferred since the continuous loading process from right to left is precisely known through the refresh rate. The shown example indicates a remaining loading time of approximately 9.5 ms at 75 Hz. Note that the figure is cropped since the full image must also contain the real-motion-indicating HMD housing (as shown in Figure A-7), unless an LED system, as discussed in Section A.7.2.4, were to be implemented.

the remaining time will also solve the technologically induced uncertainty due to the recording's sample-and-hold nature regarding the virtual motion: The continuous image loading progress on the snapshot may serve as a time scale that eliminates the discrete characteristic of snapshots.

That said, the evaluator must not rely exclusively on observing of screen tearing but use the observation at most as an additional confirmation: Depending on the displayed texture in the virtual environment, it may be quite hard at times to detect the tearing. Scenarios containing straight lines, such as the road marking in Figure A-12, will facilitate the detection of screen tearing. Note that the HMD housing in the screen center may cover a potential tearing, should the tearing happen right underneath the central cover (approximately ten-percent probability with given hardware). Since a poorly configured VSync mode may eventually have quite an impact on the end-to-end latency, it is not advised to turn the mode off for the latency assessment unless it is considered to actually drop the VSync mode for the intended application or the VSync delay is precisely known.

A.7.2.6 Fixed Relative View

Unlike the *Oculus Rift Development Kit 2* in our experiment, many modern HMDs do not allow for lens removal. This becomes particularly problematic when the HMD movement causes a change in the relative angle between the HMD screen and the observing camera. The moving HMD screen can create a false impression of virtual motion. This illusion can even be intensified, when the HMD is recorded with the HMD lenses covering the screen, with different viewing angles through the lenses causing different distortions.

Therefore, a last recommended refinement consists of mounting the camera in a fixed relative position to the HMD on a turntable or rail system (see Figure A-13). This way, the camera moves together with the HMD, maintaining an identical view on the HMD's screen at all times. This will substantially facilitate the evaluator's task, who can then navigate through the recorded frames while maintaining an unaltered view on the HMD's screen. Human evaluators and potential automated evaluation programs will both significantly benefit from such a fixed relative perspective between the HMD screen and the camera sensor. When a fixed relative position between the screen displaying the virtual scenario and the camera cannot be implemented, alternative optical setups with mirrors and lenses that enable a simultaneous capture of the physical motion and the consequent virtual one may be used.

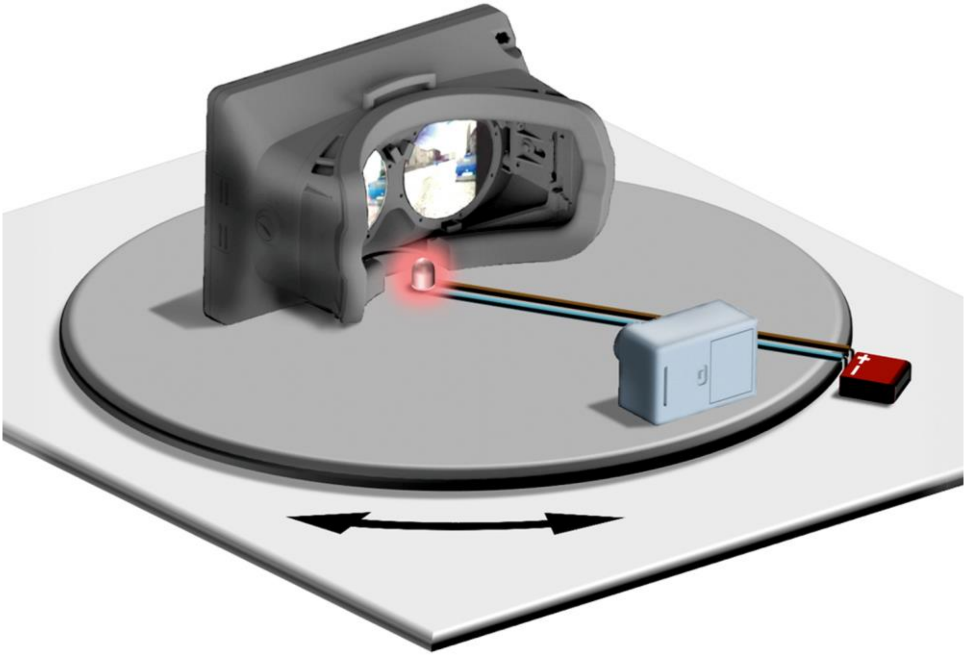


Figure A-13. When the HMD and the observing camera are both mounted on the turntable, the camera has a fixed relative view of the HMD's screen. An LED in the camera's visual field goes off when the turntable starts moving by interrupting the LED's power circuit. This LED's change of state indicates on the footage the moment the real motion started.

A.7.2.7 Summary of the Possible Refinements

A motion control system (e.g., a turntable) can provide a more precise reproducibility of the HMD movement, whereas the luminance and color of the backdrop can be adjusted to aid detection of the HMD's real motion through increased background contrast. Such a backdrop becomes unnecessary if an LED is mounted in the camera's visual field, indicating to the evaluator when motion begins in the real environment by changing the LED's state once the movement starts. This simple means will essentially facilitate the analytical procedure of the evaluator in identifying real motion, reducing the human variance theoretically to zero, and enabling recordings at much higher frame rates, which also reduces the technologically induced uncertainty.

By turning off the VSync of the VR system, the occurrence of screen tearing may help to detect virtual motion, though making it subsequently necessary to arithmetically add the theoretical VSync-induced delay to determine a true end-to-end latency, risking a disparity between the theoretical VSync value and the relevant one. This step seems anyway unnecessary since the detection of virtual movements at typical HMD refresh rates between 75 Hz and 90 Hz has not proven to be challenging. Such measures may rather be of interest if an automated evaluation algorithm were to be implemented, replacing the human evaluator.

Given sufficient space on the turntable or rail system, the camera should be mounted in a fixed relative position to the HMD to ensure an invariable perspective on the HMD's screen. This makes it considerably easier to compare the virtual scenario between two recorded frames, especially if HMD lenses are not removable or an automated evaluation image processing algorithm is used to replace the human evaluators.

Investigators who wish to extrapolate the timeline of captured moments may take advantage of the steady image loading progress on the VR screen. This loading progress is visible due to the low-persistence mode of modern HMD systems.

B ARTICLE II – FULL REPRINT

Pedestrian Simulators for Traffic Research: State of the Art and Future of a Motion Lab

Revised reprint from Feldstein, I. T., Lehsing, C., Dietrich, A., & Bengler, K. (2018). Pedestrian simulators for traffic research: State of the art and future of a motion lab. International Journal of Human Factors Modelling and Simulations, 6(4), 250–265.

Abstract. For decades, classical driving simulators have been a valuable tool for the investigation of human behavior and the validation of advanced driver assistance systems (ADAS). The development of pedestrian simulators, on the other hand, is still in its early stages. However, with increasing complexity, ADAS require a more complex design and evaluation process that is not uniquely limited to driving-simulation assessments but also takes other road users and their perspective into account. Mainly based on the technology surrounding motion capture and virtual reality (VR), pedestrian simulators allow for the investigation of human behavior from the pedestrian perspective in a reproducible, safe, and cost-efficient way. This article will help researchers starting in this research field to gain insight into the state of the art. Potential and possible areas of application of this particular simulator paradigm are briefly discussed, and an overview of some of the technologically most-advanced VR simulators for pedestrian investigations used by various research institutes around the globe is given.

B.1 Motivation

Scenarios on future mobility suggest that the proportion of urban traffic will continuously increase, and technical achievements will enable more driver assistance and active safety functions up to higher levels of automated driving. This evolution raises the question of how urban road users will interact with each other safely and efficiently. Digital human models are used for the simulation of biomechanical consequences, injury patterns, and the assessment of mitigation concepts relating to active safety systems. Crash simulation and standardized evaluation procedures are typical examples of the application of different digital human models in extreme traffic scenarios. However, the goal of active safety and driver assistance is to avoid collisions and emergency braking as well as achieve a highly cooperative behavior among different road users based on more intelligent and anticipative driver assistance systems that comfortably decelerate ahead of collision points (Bengler et al., 2014). This is one goal of the German UR:BAN research initiative (Manstetten et al., 2013). Urban traffic scenarios require the investigation of human interaction and motion patterns in complex scenarios with high replicability. A further question is how highly automated cars will influence and change the street-crossing decisions and the motion patterns of pedestrians in their surroundings.

Additionally, possibilities of interaction between different road users, such as pedestrians and autonomous vehicles, are required to be investigated. These examples show that digital human models and related tools offer robust opportunities to implement new experimental settings for the structured investigation of these effects and research questions. New experimental settings based on well-established precise motion-tracking technologies in combination with digital human models that enable the processing of motion patterns in real-time in a complex network driving-simulation environment shall help to investigate pedestrian-vehicle interactions with a focus on the pedestrian. Research questions, for example, could focus on the pedestrians' acceptance and perception of autonomous vehicles or simply train artificial intelligence with big data of pedestrian movements collected in such virtual environments in a wide range of possible scenarios.

B.2 Previous Work

B.2.1 Evolution of Pedestrian Behavior Investigation

The investigation of pedestrian behavior—in particular concerning vulnerable persons, such as children and elderly persons—has been of interest to road safety researchers for over four decades. However, for a long time, investigations have been limited to field studies or accident evaluations. Experiments in real-life situations were—and still are—either hardly viable due to the potential dangers and risks for participants that impersonate pedestrians, or were not sophisticated enough in terms of presented scenario, raising the question of the ecological validity of the results. For example, Lee et al. (1984) and Demetre et al. (1992, 1993) tried to emulate a road-crossing scenario with the so-called pretend-road method in the early stages of pedestrian investigations. In these road-crossing experiments, children were asked to cross a dummy road section, with the road-crossing decision being based on the car movements that could be observed on a real road, parallel to the artificial one (Figure B-1).

Pedestrian simulators in the form of VR devices are relatively new tools since the necessary technology started becoming accessible and suitable within the last one and a half decades, only. The first setups of these simulators often consisted of simple video settings: A participant was facing some sort of screen—a surround screen in some cases—that displayed a busy road. For gap-acceptance studies, a so-called shout task had to be performed where participants impersonated pedestrians that are willing to cross the virtual street. They had to choose a gap between the displayed passing cars

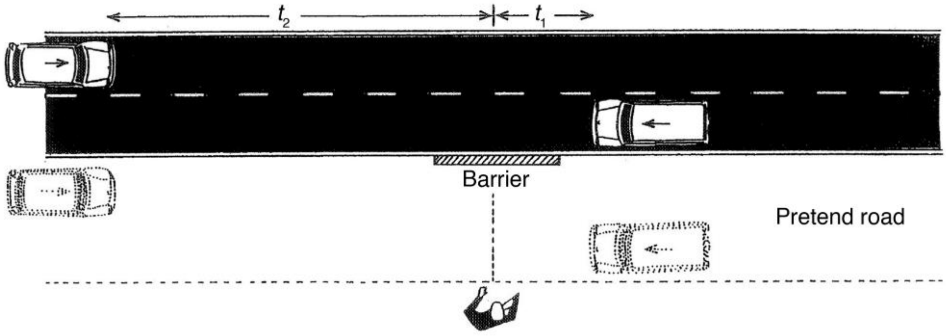


Figure B-1. Pretend-road method (from Demetre et al., 1992).

and expressed their crossing intention either verbally or by pressing a button (Pitcairn & Edlmann, 2000; Oxley et al., 2005). Participants were not capable of interacting with the virtual environment in real time, reducing the possibility of perception-action coupling. Te Velde et al. (2005) were among the first to conduct a road-crossing study with a setup that gave enough space for physical crossing of the artificial street, thus allowing a higher level of interaction with their environment. The mechanical, non-VR setting involved a bicycle that was pulled toward the participant using an electric engine and a cable, thus limiting the scenarios to relatively slow approaches. Te Velde et al. (2005) demonstrated in their simulator study significant differences regarding the participants' crossing behavior when crossing the street physically versus when the crossing intention was indicated verbally.

B.2.2 Pedestrian Simulators Around the Globe

Simpson et al. (2003) were among the first to build a pedestrian simulator that involved a head-mounted display (HMD). Their system used a monoscopic (same image for both eyes) HMD with three rotational and three translational degrees of freedom (DOF). The horizontal field of view (FOV) was limited to 48°, and the virtual environment was a highly simplified rendering.

Figure B-2. The IFSTTAR pedestrian simulator.

In 2003, the French IFSTTAR research institute developed a pedestrian simulator with three screens that enabled single-lane road-crossing investigations (Lobjois & Cavallo, 2007, 2009). In 2010, the device was enhanced and is now using a setting with ten rear-projected screens: two in the front and four on each side (Figure B-2). The CAVE-like (Cave Automatic Virtual Environment) device enables a 180° FOV when standing on the edge of the virtual street and a 300° FOV when standing in the middle. Despite the impressive FOV, the simulator creates a "visual gap" in between the right and the left side due to a missing ground projection. Thus, a virtual passing car will briefly vanish when "switching" from one side of the screens to the other. A *Vicon* motion-capture system (Vicon Motion Systems, Yarnton, U.K.) tracks the user's position, allowing for an accurate visual projection on the screens in accordance with the user's location and height. The seven-meter corridor is also equipped with a 3D sound system. IFSTTAR has a particular interest in investigating behavior of elderly pedestrians in various traffic scenarios (Dommes et al., 2013, 2014, 2015).

The University of Alabama in Birmingham developed a semi-immersive pedestrian simulator to investigate road-crossing behavior, with a focus on children (Schwebel et al., 2008; Byington & Schwebel, 2013). The transportable device can be used for traffic education and training (e.g., in schools). Users standing on a built-up curb in front of three screens indicate their crossing intention by stepping down the curb. This activates a virtual switch from first-person to third-person view, triggering an avatar crossing the street.

In 2011, the Ben Gurion University of the Negev (BGU) built a pedestrian simulator using a 180° spherical screen with a diameter of seven meters. The projection system is capable of 3D rendering, and the integration of systems that allow for physiological measurements is possible. However, in most of their investigations, the participants were merely standing in the center of the laboratory, performing a shout task (Figure B-3). The BGU focused on

Figure B-3. The BGU pedestrian simulator the investigation of hazard perception of child pedestrians (Meir et al., 2013, 2015).

The University of Iowa designed a pedestrian simulator similar to the IFSTTAR model (Jiang et al., 2016; Rahimian et al., 2016). The CAVE-like simulator uses four 3D projectors capable of producing a stereo picture, offering immersive projections to the user's sides and front but also the floor. The user wears active stereoscopic glasses and helmet-mounted markers that are tracked by an *OptiTrack* motion-capture system (NaturalPoint, Corvallis, OR, U.S.), enabling presentation of the correct visual projection for the user. The research team investigates the behavior of children and is specifically interested in the social influence of peers (other pedestrians) on street-crossing behavior.

Morrongiello et al. (2015) at the University of Guelph in Canada built, independently, a pedestrian simulator similar to the one presented in the next section. The user wears a stereoscopic HMD with a 1,280 px × 1,024 px resolution and is tracked by eight motion-capture cameras. The research team's interest concentrates on the street-crossing behavior of children. The 8 m × 5 m room allows for the investigation of street-crossing behavior on a two-lane street.

B.3 A New Pedestrian Simulator

B.3.1 Apparatus

The pedestrian simulator, developed in 2014 at the Chair of Ergonomics of the Technical University of Munich (TUM), is an HMD-based system that consists of the following main components: a control center, a motion-capture system, a custom-built motion suit, and an HMD (Figure B-4).

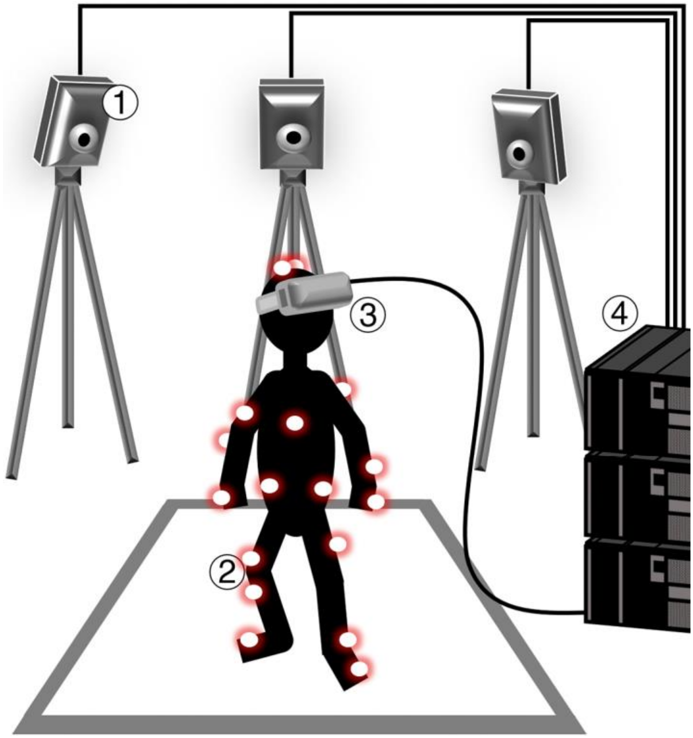


Figure B-4. The TUM pedestrian simulator (stage: 2015): (1) motion-capture camera, (2) motion suit, (3) head-mounted display, (4) control center.

The operating system runs the *Silab* software framework, a traffic simulation software developed by the Würzburg Institute for Traffic Sciences (WIVW, Würzburg, Germany). The simulator software is provided with data about the user's movements that are collected with a *Vicon* motion-capture system. Ten *Vicon T10* cameras are placed around the user, capturing markers placed on the participant's body. The *Vicon Nexus* software analyses the markers' 3D positions accurate to a millimeter in real time, using triangulation algorithms. The original *Vicon* system is operating with reflective markers that can be attached to the participant's body using double-sided tape. Since *Vicon* markers lose their reflective characteristics and reliability over time, it was decided to replace those with LEDs that continuously emit 850-nm infrared light. The LEDs are covered with diffusers and attached to a custom-made full-body spandex suit. This suit is highly elastic and can be worn over clothing by participants with a broad range of body shapes. The markers are connected through hook-and-loop fasteners to the suit, which allows for easy repositioning if necessary. The LEDs are connected through cables sewn into the suit and powered centrally with a battery belt. This motion suit helped significantly to reduce the preparation and mounting time needed to set up the participant from approximately fifteen minutes down to five.

Additionally, the tracking reliability was significantly increased as tests showed that the *Vicon* system could track the LEDs twice, partially thrice in comparison to the original reflective markers. That said, it has also to be noted that the suit, as well as any other clothing, might be shifting on the skin to some extent when participants move. For applications where the tracking of the body limbs needs to be precise to a fraction of a cm, it is still advised to fix markers directly on the participant's skin. Such precision is not required for conventional analyses of human behavior in traffic environments.

The current walking space was limited to 4 m × 2.5 m area. This is mainly due to the limited space of the laboratory and could be—given sufficient room—enlarged by up to 400% using the same equipment. If an even larger experimental area is desired, the motion-capture system might need to be expanded with additional cameras. The required number of cameras depends considerably on the proposed scenario and the expected movements of the participants. Generally, the cameras can capture the LED markers at a distance of up to eight meters, and a minimum of two cameras is necessary to derive the marker's position. However, occlusions through the participant's limbs and body may impede unobstructed tracking and require an augmentation of the camera number. Since the standard lane width of roads in Germany is between 2.75 m and 3.75 m, the currently used area only allows for road-crossing scenarios on one-way streets or streets with traffic islands.

The HMD is an *Oculus Rift Development Kit 2* (Oculus VR, Menlo Park, CA, U.S.). The low-persistence OLED panel enables a stereoscopic view at a 1,920 px × 1,080 px resolution. *Silab* produces slightly differently rendered images for each eye, thus enabling a realistic 3D view (Figure B-5). The HMD has a nominal FOV of 100° diagonally, which may, however, vary depending on the settings (e.g., the lens arrangement). This FOV comes quite close to the human horizontal binocular view of approximately 114° (Howard & Rogers, 1995). Hence, this HMD displays a viewing frustum that covers the central and near-peripheral view as well as most of the mid-peripheral view. The horizontal far-peripheral view, which corresponds to roughly 200° (Harrington, 1971), is not provided by this HMD and nor by other commonly available HMDs on the market, such as the *Sony Project Morpheus* (Sony, Tokyo, Japan) or the *Cinemizer OLED* (Carl Zeiss, Oberkochen, Germany). The panel's visible pixel structure (i.e., grid) is another deficiency that interferes with viewing visually realistic virtual environment content. An asset of the *Oculus* is its weight of only about 440 g. In the past, many HMD manufacturers had difficulties in reducing the weight and achieving a comfortable center of gravity that minimizes neck discomfort.

Figure B-5. Distorted, stereoscopic view through the head-mounted display.

B.3.2 Modeling of the Human Avatar

Numerous studies have shown that the subjective impression of reality within virtual environments is crucial for realistic behavior (Petkova & Ehrsson, 2008; Sanchez-Vives & Slater, 2005; Slater & Usoh, 1994; Witmer & Singer, 1998). This experience, often referred to as *presence* (Witmer & Singer, 1998), can be significantly enhanced by implementing a virtual body in the simulation that imitates the movements of the person situated within the simulator from a first-person view. In the pedestrian simulator presented here, the avatar is created by using *Vicon* motion-capture data to control the human body model in the *Silab* virtual environment.

The tracked human model and created avatar accomplish multiple tasks, such as

- increase of immersion, experience of presence, and spatial and plausibility illusion (Slater, 2009),
- evaluation of body language before, during, and after a simulated traffic encounter,
- measurement of traffic-specific variables, such as time-to-contact (TTC) or deceleration-to-safety time (DST),
- enabling investigation of interaction effects in interconnected simulator setups (see Section B.4).

In traffic investigations, participants need the ability to interact with other road users while not standing out from other simulated pedestrians in the virtual environment since their differentiation could bias driver behavior in linked simulator studies. Therefore, the existing standard pedestrian models in the virtual environment were used and combined with the motion-capture data to create the user avatars.

The skeleton of digital human models in *Silab* consists of 14 joints (Figure B-6) and complies with several boundary conditions, most notably a vertically fixed hip point. The standard *Vicon Plug-in-Gait* model, consisting

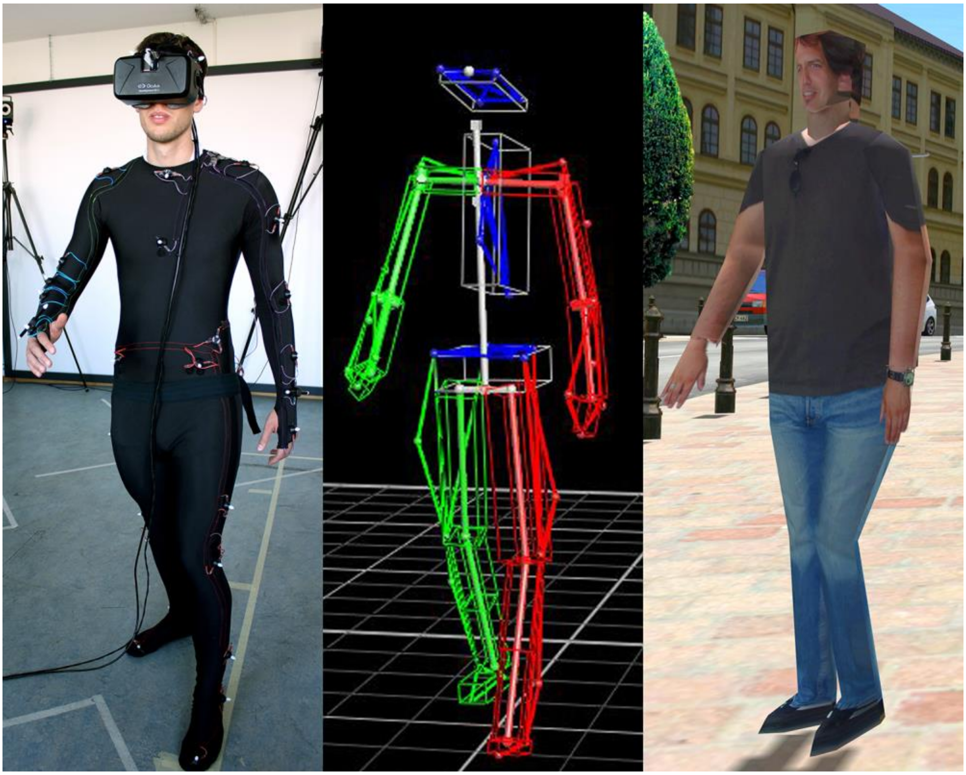


Figure B-6. Creation process of the avatar.

of 17 segments, is merged with the avatar by streaming their position and orientation data to the *Silab* simulation. Within the simulation, the pelvis is fixed in the vertical direction and serves as the central segment of the whole model. The global position and alignment in the *Vicon* coordinate system and local references of all other body parts toward their parental segments deliver the joint positions used for the *Silab* model, moving the avatar following the tracked movements. The avatar could be respawned (i.e., relocated) at predefined places in the simulation in the linked simulator setup. Hence, multiple vehicle-pedestrian encounters are possible without requiring a reset of the simulation.

The quality of the visual representation within the VR plays a vital role in the immersion and presence of a person in the simulator (Wloka, 1995).

The rotational alignment of the view relies on the HMD's inertial measurement unit and its built-in software, which ultimately reduces latency, motion blur, and jitter, while translational head movements are tracked with the *Vicon* system. The *z*-coordinate (vertical coordinate) of the view is fixed, reducing possible effects of simulator sickness induced by vertical wobble when walking.

B.3.3 Preliminary Acquired Results

First studies using this pedestrian simulator revealed that users familiarized themselves quickly with the virtual environment. They showed no signs of difficulties with moving within the virtual environment. Only one participant of seventy-five had to prematurely end the experiment due to nausea. This might be an indication that the latency between the captured movements and the visual representation is low enough to avoid visual-vestibular conflicts such as experienced in stationary driving simulators.

The impression of immersion experienced by the participants was measured using the presence questionnaire (PQ), initially developed by Witmer and Singer (1998) and revised by the UQO Cyberpsychology Lab (Robillard et al., 2002). The questionnaire contained questions concerning the subjectively experienced immersion and was broken down into the five factors "realism," "possibility to act," "quality of interface," "possibility to examine," and "self-evaluation of performance." Overall, the simulator system achieved high PQ scores, exceeding the average scores that were reported by the UQO Cyberpsychology Lab for a series of VR systems that were validated with regard to their effectiveness, suggesting that the pedestrian simulator presented here delivers a satisfyingly high experience of presence (Feldstein et al., 2016).

For the comparison of the users' walking habits when wearing and not wearing an HMD, the walking paces of 30 participants with an average age

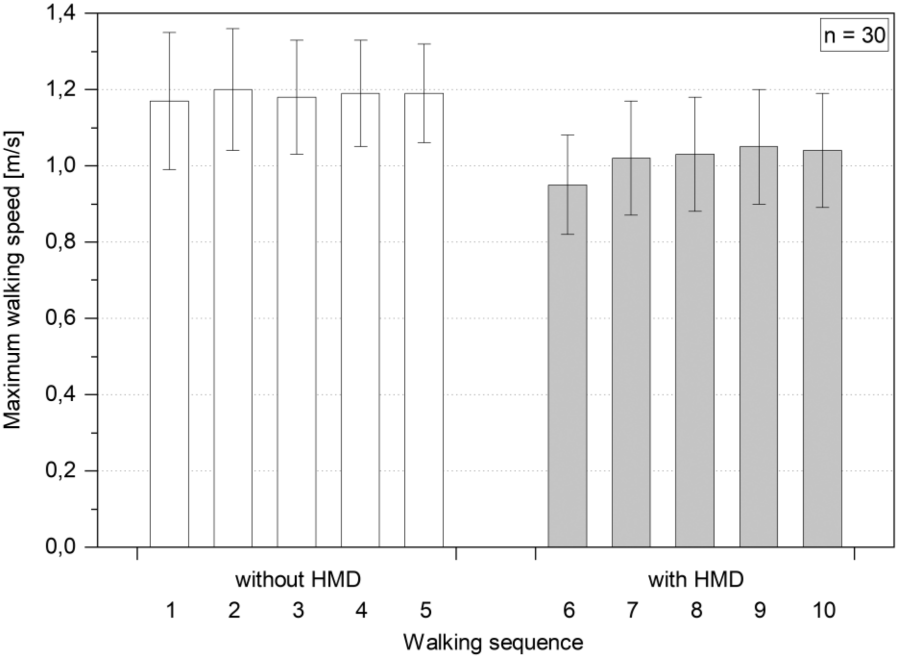


Figure B-7. Maximum walking speeds when walking without and with a head-mounted display, with the error bars indicating the standard deviation.

of 25 y/o (SD = 1.55) were recorded on a four-meter track. These participants had no prior experience with this kind of simulator. First, they were asked to walk the track five times wearing the motion suit, but not the HMD. Subsequently, the participants had to walk the four-meter track again five times, but these times, they were wearing the HMD while being virtually placed in a calm city location (without any interaction with virtual cars or pedestrians). Figure B-7 shows the average maximum walking pace of the participants, revealing a learning and acclimatization process in the virtual environment. However, a repeated-measures ANOVA also suggests that there was still a significant difference observable between the last walk wearing an HMD (M = 1.035 m/s, SD = 0.146 m/s) and the average walk not wearing an HMD (M = 1.183 m/s, SD = 0.144 m/s): $F(1, 29) = 42.97$, $p < .001$, $r = .45$.

B.4 Linked Simulation

The investigation of driver behavior regarding the interaction with other road users—be it motorized or nonmotorized, and assisted or nonassisted—is classically performed in a driving simulator with one participant. This driver usually performs a specific driving task (with or without additional tasks) that addresses the relevant research questions. Besides the well-known discussion about simulator validity related to issues of FOV and resolution (Jamson, 2001), motion cueing (Kemeny, 2001), and especially surrounding motorized traffic (Hancock et al., 2003; Kemeny, 2001; Maag et al., 2012; Rittger et al., 2014; Schindler et al., 2011), urban driver behavior analyses have to consider the interaction between vulnerable road users and drivers (Lehsing et al., 2015).

This interaction process between two or more road users is an essential and bilateral behavior adaption process. In the classical driving simulation approach, drivers are only capable of reacting to the programmed behavior of other road users. This behavior lacks the potential and validity of the interaction mentioned above.

Linking two or more simulators enables the essential process of behavior adaption, and participants can react to each other. Furthermore, the realization of fundamental social interaction in such a synthetic environment needs particular analysis methods that take into account the time-series character of such inter-human processes. One possible approach regarding the assessment of behavior with several road users is cross-correlation analysis. In this case, two signals, for example, the speeds in the situations of interest, are analyzed. These signals are shifted against each other to find their maximum positive or negative correlation and the corresponding lag (time-shift) necessary to gain the maximum correlation coefficient. A recent study (Lehsing et al., 2015) showed that this analysis method is capable of detecting differences in the interaction process between two road users. The approach investigated the necessary lag and cross-correlation coefficient depending on the crossing pedestrian type (human-controlled vs. programmed pedestrian) and the crossing situation type (free lane vs. occlusion vs. zebra crossing). It was shown that in the human-human constellation, the social interaction between these two road users produced a positive lag to gain the maximum correlation coefficient. In this case, the leading (time) series was the driver; in other words, the pedestrian's behavior was dependent upon the driver's behavior. The latter is supported by the assumption that an aware pedestrian considers a safe crossing and behaves defensively. If the car driver does not show yielding behavior, the pedestrian does not cross the street.

Linking two or more simulators enables the essential process of behavior adaption, and participants can react to each other. Furthermore, the realization of fundamental social interaction in such a synthetic environment needs particular analysis methods that take into account the time-series character of such inter-human processes. One possible approach regarding the assessment of behavior with several road users is cross-correlation analysis. In this case, two signals, for example, the speeds in the situations of interest, are analyzed. These signals are shifted against each other to find their maximum positive or negative correlation and the corresponding lag (time-shift) necessary to gain the maximum correlation coefficient. A recent study (Lehsing et al., 2015) showed that this analysis method is capable of detecting differences in the interaction process between two road users. The approach investigated the necessary lag and cross-correlation coefficient depending on the crossing pedestrian type (human-controlled vs. programmed pedestrian) and the crossing situation type (free lane vs. occlusion vs. zebra crossing). It was shown that in the human-human constellation, the social interaction between these two road users produced a positive lag to gain the maximum correlation coefficient. In this case, the leading (time) series was the driver; in other words, the pedestrian's behavior was dependent upon the driver's behavior. The latter

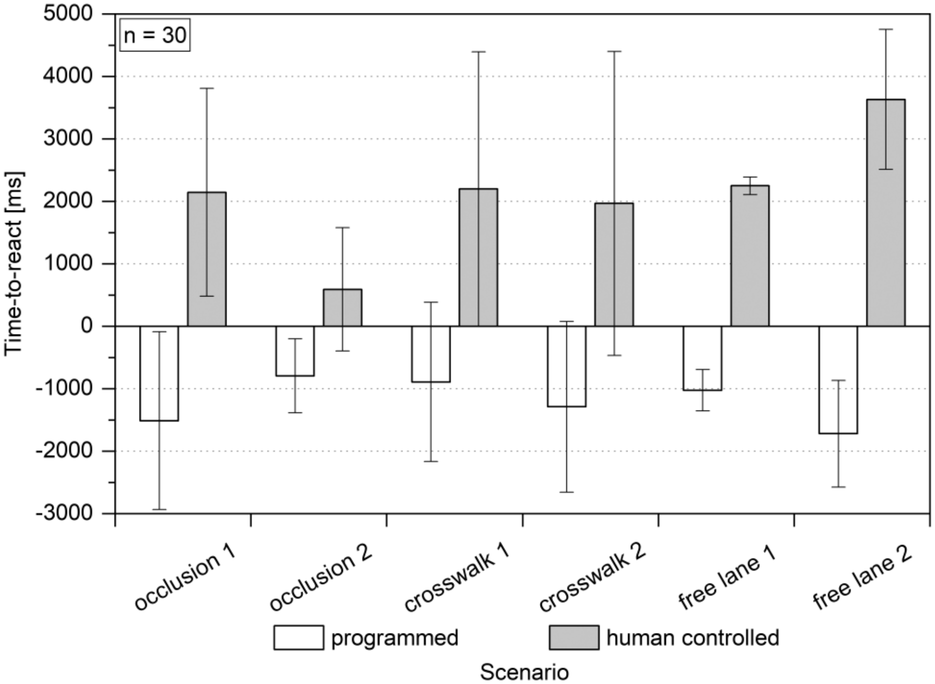


Figure B-8. Time-to-react deduced from the maximum behavior correlation between the vehicle and the pedestrian for different crossing situations, with the error bars indicating the standard deviation.

is supported by the assumption that an aware pedestrian considers a safe crossing and behaves defensively. If the car driver does not show yielding behavior, the pedestrian does not cross the street.

Quite the contrary was observed when the programmed pedestrian crossed the street. The time lag (interpreted as time-to-react) was always negative, which means that the pedestrian was the leading series; in other words, the driver's behavior was dependent upon the pedestrian's behavior (Figure B-8). This is—especially in terms of safety issues—an unrealistic behavior, which was the result of the programming of the pedestrians in the used software framework. One possible use of this kind of unaware pedestrian, who is not paying attention to cars or other traffic-related threats, may be the simulation of pedestrians talking on their phone and other scenarios in which pedestrians are not paying attention to the surrounding traffic. Additionally, a combination of the time series analysis and the classical assessment of safety measures, such as TTC, DST, and post-encroachment time (PET), will result in a more holistic assessment of behavior. A more naturalistic interaction between road users must be afforded.

Beyond the driving specific metrics, the assessment of glance behavior is a further aspect that should be taken into consideration when performing experiments with a multiple-simulator setting (Mourant & Rockwell, 1970). Due to the characteristics of urban traffic and related glance behavior of road users, the analysis of eye-tracking data may support the findings of the driving behavior analysis and, therefore, the validity of the data in general.

In summary, the approach of linking simulators has the potential to increase simulator data validity because the provision of an interaction channel using non-verbal communication facilitates a more realistic behavior. Nevertheless, before use of a multiple-simulator setting is considered, the increased expenditure in terms of time, staff, equipment, and costs should be taken into consideration.

B.5 Discussion

Pedestrian simulators enable research on pedestrian behavior in potentially hazardous traffic situations (e.g., road-crossing scenarios). Becoming an increasingly valuable tool for car manufacturers and original equipment manufacturers (OEMs) in the development process of ADAS (e.g., pedestrian detection and avoidance), the relevance of these simulators will grow given the advent of autonomous and silent cars (e.g., electric vehicles) in the near future. The steady advances in VR technology push the performance of pedestrian simulators, opening new possibilities for reproducible and safe investigations.

More studies are required that compare the behavior of pedestrians in reality with the behavior in the virtual environment to confirm the validity of the presented pedestrian simulator fully. The first studies revealed that the concept of the pedestrian simulator is heading in the right direction. In the future, an HMD with a higher resolution and an invisible pixel grid might increase the visual performance and the quality of the interface, which was demonstrated to have deficits. Additionally, the effect of the missing far-peripheral view may be further investigated.

The simulator is intended for the investigation of pedestrian road-crossing behavior. This case scenario remains one of the most frequent vehicle-pedestrian encounters and, therefore, accounts for numerous potentially hazardous situations. Other case scenarios can also be investigated, but limitations may occur through the restricted freedom of movement within the simulator system. The user's realistic perception regarding the TTC of approaching vehicles needs to be validated in order to ensure the validity of data collected using the simulator system when investigating road-crossing scenarios.

While a CAVE-like system enables a large FOV, the field of regard (FOR) is limited by the size of the screens. An HMD system, on the other hand, has an unlimited FOR but a limited FOV. In the case of the HMD used here, the FOV corresponds to approximately 100° diagonally. The effect of these restrictions on the behavior of the participants should be considered when choosing the system (CAVE vs. HMD).

The investigation of walking habits with participants wearing an HMD revealed that a familiarization process with the system is necessary. Further studies are needed that aim at developing acclimatization strategies that ultimately will adjust HMD walking to real-life walks.

Experiments conducted with the new pedestrian simulator have shown that the self-representation with an avatar helps participants to experience the virtual world in an immersive way and produce realistic traffic encounters. That said, the virtual avatar still offers significant improvement potential. In particular, a more realistic avatar ground contact must be implemented in further developments, also in terms of contact shadows: With the hip being fixed in the vertical level, the feet appear to float on the ground, thus breaking the immersion. Furthermore, a higher number of vertices that form the current avatar and a better texture quality may lead to a higher acceptance of the virtual body and increase the perceptual illusion of body swapping (Petkova & Ehrsson, 2008).

The approach of linking a driving simulator to a pedestrian simulator so that both participants can meet simultaneously in the same virtual environment was introduced. This promising approach facilitates social interaction regarding mutual behavior adaption in the virtual environment of a driving simulation. Experimental evaluations that use this multiple-simulator setup can address a wider range of research questions in traffic-related areas where the interaction between different classes of road users can be assessed as opposed to the reaction to programmed agents used in conventional driving simulators.

The approach of two human beings encountering each other in a safe and reproducible traffic environment shows its potential, especially in urban scenarios where interaction plays a significant role. Critical aspects that may arise when performing experiments with two or more simulators are related to standardization, participants, data analysis, research question, and experimental expenditure. The significant advantage of simulator experiments with a single simulator is the investigator’s control over all aspects that have or do not have to take place in the simulation. When two or more humans meet in the simulation, the variety of behavior is widened and less controllable.

The study design and data collection have to ensure that data acquired with multiple-simulators systems are synchronized, so that accurate analyses of critical events across all metrics and participants may be achieved. Furthermore, the use of a linked simulator setup should be related to the addressed research question and the potential of a more realistic human interaction. For example, if the mutual behavior adaption of two or more road users is not essential for the assessment of the analyzed ADAS, the investigation may be conducted classically with one simulator and less analytical and organizational expenditure.

C ARTICLE III – FULL REPRINT

Impending Collision Judgment from an Egocentric Perspective in Real and Virtual Environments: A Review

Revised reprint from Feldstein, I. T. (2019). Impending collision judgment from an egocentric perspective in real and virtual environments: A review. Perception, 48(9), 769–795.

Abstract. The human egocentric perception of approaching objects and the related perceptual processes have been of interest to researchers for several decades. This article reviews numerous studies that investigated the phenomenon when an object approaches an observer (or the other way around) with the intention to single out factors that influence the perceptual process. A taxonomy of metrics is followed by a breakdown of different experimental measurement methods. Thereinafter, potential factors affecting the judgment of approaching objects are compiled and debated while divided into human factors (e.g., gender, age, and driving experience), compositional factors (e.g., approaching velocity, spatial distance, and observation time), and technological factors (e.g., field of view, stereoscopy, and display contrast). Experimental findings are collated, juxtaposed, and critically discussed. With virtual reality devices having taken a tremendous developmental leap forward in the past few years, they have been able to gain ground in experimental research. Therefore, particular attention in this article is also given to the perception of approaching objects in virtual environments and put in contrast to the perception in reality.

C.1 Introduction

The automotive industry faces new challenges in the course of the vehicles' automation process. The behavior of all surrounding road users must be understood and implemented in the vehicles' artificial intelligence. This also involves, for example, human perception and judgment of an approaching object, such as an oncoming car. While some researchers explored the neural basis on how an approaching object is processed by humans (Billington et al., 2010; Coull et al., 2008; D. T. Field & Wann, 2005), cognitive factors and heuristics that may have an impact on how humans judge an oncoming object are also of profound interest to investigators. While these factors may be related to human individuality or the composition of the environment, the technological components of a virtual setup used for the assessment may also affect the judgment and corrupt the findings to a certain degree. Such virtual environments may be found, for example, in driving simulators or other virtual reality (VR) systems. VR started growing intensely over the past few years, after the entertainment industry started showing interest in this field, with recent technological advances opening up new possibilities also for the research community. Investigators analyzing human behavior tap into the utilization of VR systems due to their advantages over real-life experiments in terms of participant safety, cost efficiency, and reproducibility of investigated scenarios. That said, VR systems initially require an unfortunately often omitted validation process, which analyzes the degree of similarity between participants' perceptions and behavior in the virtual environment and perceptions and behavior in real life. Only then findings acquired in the virtual environment may be reliably projected onto real-life scenarios, and respective conclusions may be drawn.

The investigation of factors that influence impending collision judgments experienced its heyday back in the 1980s and 1990s after Lee (1976) had introduced his concept of *tau*. Nevertheless, this field is gaining much relevance again, for example, in the course of training artificial intelligence in vehicles. The article presented here summarizes and integrates previous findings. Furthermore, the impact of virtual environments on the outcome of collision judgments is examined and discussed critically.

C.2 Taxonomy of Metrics

The research community employs a variety of terms when investigating the remaining time between two objects approaching one another. Widespread terms are *time-to-contact* (e.g., Tresilian, 1991), *time-to-passage* (e.g., Kaiser & Mowafy, 1993), *time-to-arrival* (e.g., Schiff & Oldak, 1990), *time-to-collision* (e.g., Cavallo & Laurent, 1988), and *time-to-coincidence* (e.g., Groeger & Cavallo, 1991). The subtle distinction between these terms may lay in details:

- whether the observer or the object or both of them are moving,
- whether the objects are on a collision course or passing each other, and
- whether the observer is perceiving the collision from an egocentric perspective or rather uninvolved observing than participating in the collision.

Hancock and Manser (1998) strived to organize the terminology and to create a formal taxonomy: An overview of different terms and their characteristics based on Hancock's and Manser's suggestion is given in Table C-1.

The usage of the terms among researchers can be somewhat confusing and inconsistent. Because of these small differences, which have no impact on the underlying concept, and the lack of a formal taxonomy, the terms are often used interchangeably in literature, even being treated as synonyms (see Bootsma & Oudejans, 1993; Parsonson et al., 1996; Schiff & Oldak, 1990; Tresilian, 1991). In addition to the terms listed in Table C-1, Hancock and Manser (1998) also suggested the term *time-to-go*, which is supposed to

Table C-1. Terminology and characteristics of metrics used to describe the remaining time between two objects on a collision or crossing course.

Terminology	Characteristics
Time-to-arrival	{Moving observer} and {stationary object} on a collision course
Time-to-coincidence	{Moving object} and {stationary or moving object} on a collision course from the perspective of a distant, uninvolved observer
Time-to-collision	{Moving observer} and {moving object} on a collision course
Time-to-contact	{Stationary observer} and {moving object} on a collision course
Time-to-passage	{Stationary observer} and {moving object} on a crossing course

describe a stationary observer facing a moving object on a collision course or passing course. Given the proximity to *time-to-contact* in terms of physical parameters, the use of two different terms appears unnecessary. Consequently, the term *time-to-go* has never been able to establish itself within the framework of urban traffic research, even 20 years after Hancock and Manser introduced their taxonomy. In fact, *time-to-go* has presumably not been employed again in this research field since Carel's publication in 1961. Nowadays, the term can be found in aircraft research describing an unrelated metric.

When investigating road-crossing scenarios from the perspective of a pedestrian who is facing an approaching vehicle, *time-to-contact* (TTC) as a metric meets the required characteristics. From the view point of the driver approaching the pedestrian, the corresponding metric is termed *time-to-arrival* (TTA). TTC and TTA can be determined from three-dimensional low-order information or deduced solely from two-dimensional optical variables, such as looming. The *low-order information* method builds on the ratio of the oncoming object's distance to the approaching velocity and is computed as follows:

$$TTC = \frac{distance_{object}}{velocity_{object}} \qquad TTA = \frac{distance_{object}}{velocity_{observer}}.$$

The *looming* method, on the other hand, consists of Lee's (1976) concept using the variable τ, also noted as *tau*, relying on the visually perceived angular size of the approaching object (e.g., vehicle height) that is taken in proportion to the instantaneous rate of its size expansion—perceived while approaching—and is computed as follows:

$$\tau(t) = \frac{\text{(angular separation of any two image points of the object)}}{\text{(rate of angular separation of the image points)}}.$$

While numerous studies have shown that human participants are capable of estimating low-order parameters, such as distance and velocity, humans will mainly rely on the perceived change of optic array to judge an approaching object (Lugtigheid & Welchman, 2011; McLeod & Ross, 1983; Yan et al., 2011). Interestingly, this contrasts machine systems such as autonomous vehicles, which usually deduce temporal distances based on three-dimensional, sensor-tracked parameters such as described in the *low-order information method*.

C.3 Measurement Methodologies

Different study concepts have been introduced for the investigation of participants' judgments of approaching objects, of which an overview is given in Figure C-1. Generally, they can be divided into *estimation tasks* and *discrimination tasks*.

Estimation tasks contain the subsets *coincidence anticipation* and *interceptive action* (Tresilian, 1995). The concept of *coincidence anticipation* requires participants to determine the collision moment with an approaching object after the object has disappeared from sight by extrapolating the afore perceived motion and estimate the moment when the collision would have taken place, for example, by pressing a button at

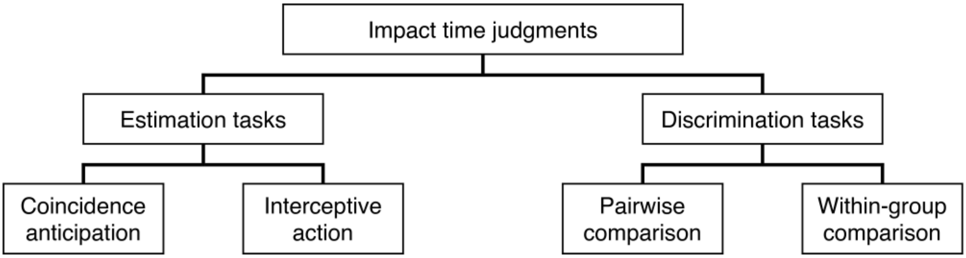


Figure C-1. Overview of classical tasks evaluating participants' impact time judgments

the respective moment (e.g., Cavallo & Laurent, 1988). The disappearance of the object from the observer's sight can be due to occlusion or a virtually induced vanishing. The *interceptive action* concept requires participants to react to a specific incident by performing a specific action. The task can, for example, consist of catching an approaching object (e.g., van der Kamp et al., 1997) or avoiding the object in the last possible moment (e.g., Li & Laurent, 1995). There is little point in asking the participant to express the estimate in numerical values (e.g., the temporal or spatial distance of an approaching vehicle) due to the innate human limitations in making such judgments (see Guzy et al., 1991; Hills, 1980; Loftus, 1979).

Discrimination tasks use standardized psychophysical techniques aimed at measuring psychometric functions to determine, for example, just-noticeable differences (see Green & Swets, 1966). These tasks usually require a *pairwise comparison* leading to the determination of a discrimination threshold. Participants may face, for example, two approaching objects, either simultaneously or successively, and judge the approaches *pairwise* with each other (e.g., DeLucia, 1991). D. Regan and Hamstra (1993) used an uncommon discrimination task variant, following the psychophysical procedure introduced by McKee (1981), in which the participant was asked to estimate whether a given stimulus was lower or higher than the mean of the entire set of stimuli (i.e., *within group*).

The *within-group comparison* remains a rare procedure, while *coincidence anticipations*, *interceptive actions*, and *pairwise comparisons* are the commonly encountered experimental methodologies in this field. When choosing an experimental design, it should be noted that *estimation tasks* and *discrimination tasks* look at different aspects: While an *estimation task* will investigate the perception of the approaching object relative to the observer, the *discrimination task* will determine the participant's ability to detect changes or differences between several stimuli. Consequently, when investigating specific variables in an experiment (e.g., the performance in a real environment vs. a virtual one), the outcome may ultimately depend on the chosen task type because the human estimation ability is cognitively dissociated from the human discrimination ability (see Seward et al., 2007).

C.4 Supportive Depth Cues

When an observer is confronted with an approaching object, the perception of the object's spatial distance and its rate of change are essential heuristics. Therefore, it is of little surprise that depth cues and their availability to the observer affect the judgment of impending collisions. Over time, many different depth cues have been determined, investigated, and discussed regarding their availability and effectiveness over spatial distance. Although depth cues have just a supporting function when judging an approaching object, this section has been dedicated to depth cues, in which the bases and different types are summarized. This shall sensitize researchers who investigate impending collision judgments—especially in virtual environments—as to which cues may affect the depth perception within a specific range and should, therefore, be provided to support the validity of the findings.

Cues that allow inference of three-dimensional depth information from a two-dimensional still image are termed *pictorial depth cues*, whereas other cues are categorized as *nonpictorial depth cues* and are related to motion,

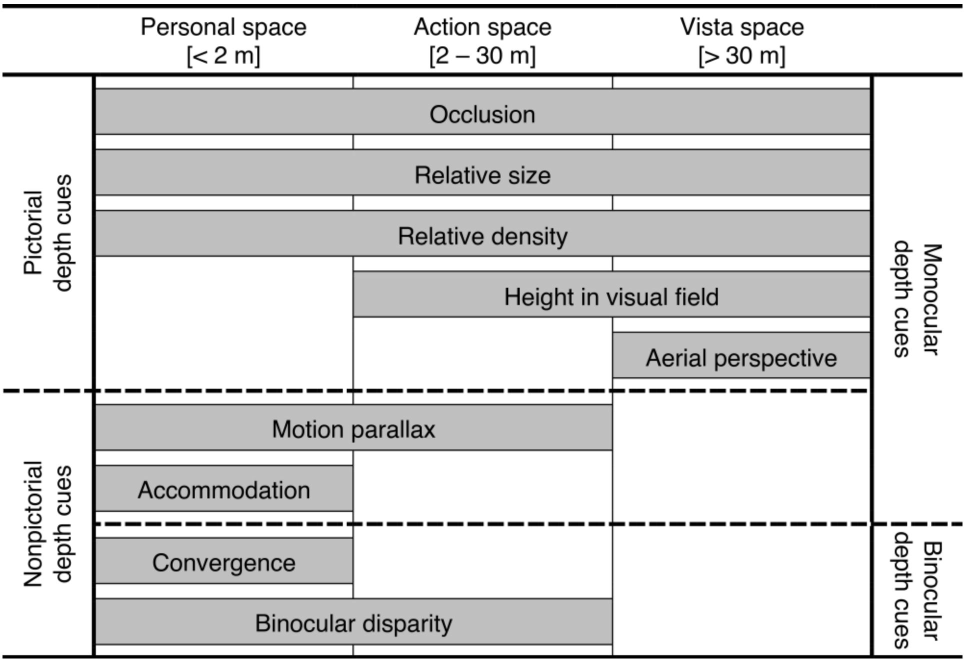


Figure C-2. Overview of different depth cues and their approximate effective range.

the oculomotor system, or stereoscopy. A further distinction can be made by dividing into depth cues that rely either on monocular information or require binocularity. An overview of different depth cues and their effective range, their classification into pictorial and nonpictorial depth cues, and whether they rely on monocular or binocular information is given in Figure C-2, with data based on Cutting and Vishton (1995), Nagata (1989), and Renner et al. (2013).

Pictorial depth cues:

- Occlusion: Objects that are closer visually overlap objects farther away, which allows observers to determine the relative, but not absolute distance toward them.
- Relative size: When objects approach the observer, the object's relative size increases on the retinal image. Thereby, the observer's familiarity with the object's size is relevant, but even when the

object is unknown to the observer, smaller objects may appear relatively farther away (Sousa et al., 2011, 2012).

- Relative density: With increasing distance, the relative retinal density of surface texture will increase. However, relative density is merely around the threshold that is considered effective for perceiving depth (Braunstein, 1976; Cutting & Vishton, 1995; Marr, 2010; K. A. Stevens, 1981).
- Height in visual field: Objects that have a smaller angular distance to the horizon from the observer's point of view (i.e., objects that are vertically closer to the horizon) appear to be farther away. Theoretically, on flat ground, an object's absolute distance can be extracted solely from its height in field, with the effective range lying beyond 2 m for an upright observer, although the effectiveness of this cue diminishes curvilinearly over distance (Cutting & Vishton, 1995; Rand et al., 2011; Sedgwick, 1983).
- Aerial perspective: Objects at a far distance have lower contrast and lower color saturation, usually involving a bluish cast owed to light scattering by the atmosphere. Aerial perspective is the only depth cue with increasing effectiveness over increasing distance, though the effective range will largely depend on the environmental and meteorological conditions (Cutting & Vishton, 1995; Nagata, 1989). Considering that objects at a great distance become indistinct and that the effective range of aerial perspective usually starts at no lower than a minimum of a few hundred meters, this depth cue is ineffective when judging approaching objects such as vehicles.

Nonpictorial depth cues:

- Motion parallax: When an observer is moving, the apparent relative motion of stationary objects against the background gives the observer clues about the relative distance of those objects, with nearby stationary objects having a higher relative motion than those that are farther away. Mathematically, motion parallax can deliver absolute depth information (Ferris, 1972), though effectiveness declines rapidly with increasing distance and ultimately becomes ineffective for distances beyond the action space (Cutting & Vishton, 1995). Additionally, effectiveness is further reduced when the respective object's relative motion is in the axis to the observer rather than lateral.

- Accommodation: Kinesthetic sensations when contracting and relaxing ciliary muscles that are responsible for changing focal length deliver cues about the depth of the object on which the observer is focusing. That said, accommodation is solely effective for distances less than 2 m (Fisher & Ciuffreda, 1988; Okoshi, 1976).
- Convergence: Due to binocularity, the two eyeballs are moved inward to create an intersecting focus point. The extraocular muscles that are responsible for this movement deliver kinesthetic sensations, similar to accommodation, and, consequently, cues about the depth of the focus point. Like accommodation, effectiveness is limited to about 2 m (Cutting & Vishton, 1995; Nagata, 1989; von Hofsten, 1976), although some sources claim that effectiveness may be applicable as far as 10 m (Okoshi, 1976).
- Binocular disparity: Due to horizontal separation of the eyes, the eyes perceive two slightly different images of the scene, which are ultimately fused into one image. The slight differences between the two original images, however, allow the subconscious extraction of depth information by triangulation. The opinions about the effective range of binocular disparity significantly vary across literature, with investigators reporting different thresholds at which binocular disparity as a depth cue becomes ineffective: Some researchers suggest the cue's effectiveness be limited to the

personal and action space (<30 m), while others claim the effectiveness of up to 135 m (see Cavallo & Laurent, 1988; Cutting & Vishton, 1995; Foley, 1991; Lappin, 2014; Nagata, 1989; Palmisano et al., 2010; Schiff, 1980). The varying interpupillary distances (IPD) across participants certainly play a role (discussed in Section C.5.1), but also stereo weakness and stereo blindness are not uncommon in the general population and affect the effective range (Cutting & Vishton, 1995). Clearly, the binocular disparity is most effective in near distances, with the effectiveness linearly decreasing when distances increase.

Additional depth cues that are also commonly mentioned in literature include *texture gradients* (e.g., Gibson, 1950), *linear perspective* (e.g., Kubovy, 1988), *light and shading* (e.g., Boring, 1942), *kinetic depth* (e.g., Wallach & O'Connell, 1953), *kinetic occlusion* (e.g., Kaplan, 1969), and *gravity* (e.g., J. S. Watson et al., 1992). These cues have been neglected in the list presented here, as they represent either some combination of several of the cues discussed above or do not contribute to the extraction of depth information (Cavanagh & Leclerc, 1989; Cutting & Millard, 1984; Cutting & Vishton, 1995).

C.5 Factors Affecting Impending Collision Judgments

C.5.1 Human Factors

Over recent decades, numerous studies have been conducted on the perception of approaching objects. Participants across those studies all have in common that they have persistently underestimated the temporal distances to approaching objects (i.e., TTC or TTA). A meta-analysis attempts to visualize this underestimation effect in Figure C-3, though without taking into account the numerous different variables across the

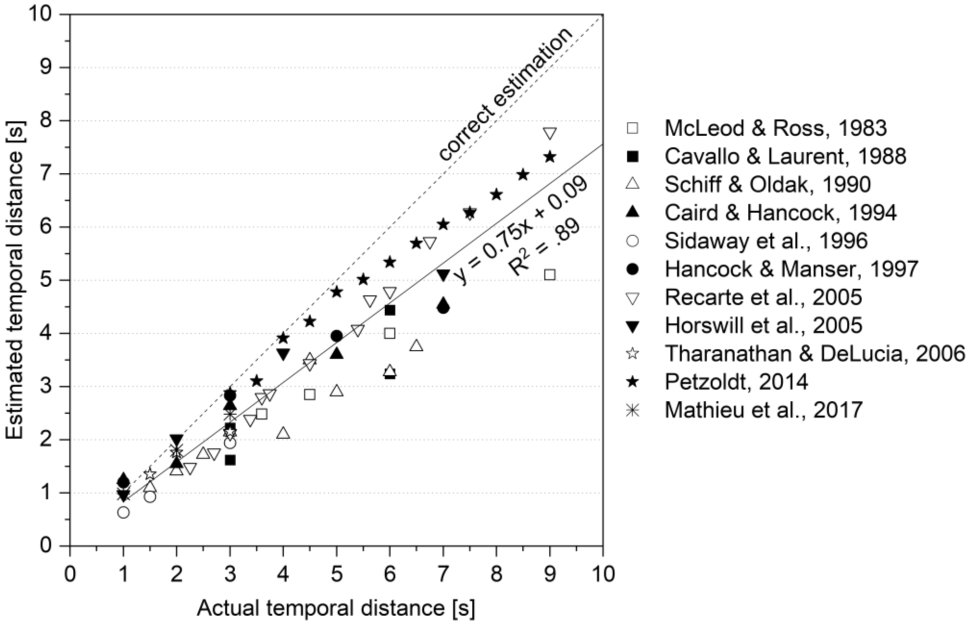


Figure C-3. Meta-analysis of impending collision judgments from an egocentric perspective, across eleven different studies that have used contextual stimuli (i.e., realistic environments containing depth cues), plotted as estimations relative to the actual temporal distances (i.e., time-to-contact and time-to-arrival).

studies, in other words, ignoring the strongly differing experimental designs, the varying number of participants, and the varying number of assessed estimates (see Caird & Hancock, 1994, Cavallo & Laurent, 1988; Hancock & Manser, 1997; Horswill et al., 2005; Mathieu et al., 2017, McLeod & Ross, 1983; Petzoldt, 2014; Recarte et al., 2005; Schiff & Oldak, 1990; Sidaway et al., 1996; Tharanathan & DeLucia, 2006) . Experiments conducted by Geri et al. (2010) and Tharanathan and DeLucia (2006) indicated that temporal distances were judged lower when the observer was moving toward the stationary object (thus the TTA) than when the object was moving toward the stationary observer (thus the TTC). Nevertheless, due to the similarity of events, this meta-analysis takes both scenarios into account. An overview of the experimental characteristics is shown in Table C-2, indicating whether the experiments investigated:

- the TTC or

- the TTA,

and also, whether the experiments

- conducted the study in a real environment,
- displayed a previously recorded video of a real traffic scenario, or
- used a simulated scenario.

Please note that in order to facilitate the reading flow, the research overview in this section and the sections hereinafter use the term TTC, referring eventually also to studies that investigated the TTA.

Despite the fact that all of the analyzed studies contained a lifelike environment, the designs of those experiments varied considerably and are thus difficult to compare. Nevertheless, this meta-analysis allows an approximate idea of the expectable ratio of TTC estimation to the actual

Table C-2. Experimental characteristics of studies examined in the meta-analysis.

		McLeod & Ross (1983)	Cavallo & Laurent (1988)	Schiff & Oldak (1990)	Caird & Hancock (1994)	Sidaway et al. (1996)	Hancock & Manser (1997)	Recarte et al. (2005)	Horswill et al. (2005)	Tharanathan & DeLucia (2006)	Petzoldt (2014)	Mathieu et al. (2017)
Mode	Time-to-arrival	•	•			•		•		•		
	Time-to-contact			•	•		•		•	•	•	•
Environment	Real environment		•					•				
	Recorded video	•		•		•		•	•			
	Simulation				•		•			•	•	•

TTC. A linear regression (R^2 = .89) indicates a ratio of about 75% with the slope having an *SE* = 0.03. Schiff and Oldak (1990) attributed this phenomenon to the biological instinct of self-preservation with built-in or learned tendency to err in the direction of safety in potentially hazardous situations, particularly when the perceptual response system is lacking precision. This is supported by the observation that underestimations occur in studies that have used contextual stimuli (e.g., representations of approaching vehicles and sophisticated backgrounds). Other studies that have used highly simplified stimuli (e.g., an approaching square on a plain white background) varied significantly, indicating partially substantial underestimations (Carel, 1961; Schiff & Detwiler, 1979; Todd, 1981), slight overestimations (Gray & Regan, 1998; Kaiser & Mowafy, 1993), as well as large overestimations of the TTC (DeLucia et al., 2016; Geri et al., 2010).

When looking at gender as a between-subjects factor, significant differences were found, indicating shorter TTC values for female than for male participants (see Caird & Hancock, 1994; Manser & Hancock, 1996; McLeod & Ross, 1983; Schiff & Oldak, 1990). A possible explanation may lie in women's pursuit for greater safety margins (see Evans, 2004; Hills, 1980; Kadali & Vedagiri, 2012; Konečni et al., 1976; Montgomery et al., 2014; Parsonson et al., 1996) and thus an extended underestimation of the TTC, whereas the higher risk tolerance of men by implication leads to higher accuracy of male TTC estimates. In addition, numerous experiments carried out in the past have revealed that men tend to have fewer difficulties with regard to solving spatial tasks when compared with women: A meta-analysis by Voyer et al. (1995) reviewing 190 experiments with regard to spatial visualization, spatial perception, and mental rotation documented partially significant gender differences, with 112 of the experiments in favor of men and 3 experiments in favor of women, while no significant difference was observed in 75 of those experiments. Furthermore, men account for a larger share of traffic, resulting in increased driving experience (see Kirkham & Landauer, 1985; McGuckin & Fucci, 2018; Polus et al., 1988): Increased driving experience leads to more accurate TTC estimates, as a male-only study conducted by Cavallo and Laurent (1988) has shown. Interestingly, Recarte et al. (2005) found no significant difference between the TTC estimates of men and women in their study with a sample of fairly young participants (*M* = 23.5 years, *SD* = 2.5) and similar driving experiences for both genders (*M* = 24,355 km for men, *M* = 27,101 km for women). This could support the hypothesis that driving experience might, to a certain extent, have a larger impact on the estimation than the gender itself, at least when it comes to young participants.

The participants' age also appears to influence the accuracy of TTC estimates: Schiff et al. (1992) observed that elderly participants tend to underestimate the TTC to a greater extent than their younger comparison group. This effect was particularly noticeable when comparing female participants only. The differences between younger male and elderly male participants were less pronounced than for the inter-female comparison, which has a certain consistency with Matthews' (1986) findings of attributing elderly male drivers with apparent overconfidence in traffic. A similar age effect, as well as age × gender interaction, such as reported by Schiff et al. (1992), was observed by Hancock and Manser (1997), and also Andersen and Enriquez (2006) and Dommes et al. (2013) observed a sharp deterioration of adults' TTC judgments with increasing age. On a practical basis, age has been shown to affect road-crossing choices and may be partially related to difficulties of elderly people and children in judging approaching vehicles (see Barton & Schwebel, 2007; Connelly et al., 1998; Demetre et al., 1992, 1993; Dommes & Cavallo, 2011, 2012; Dommes et al., 2012, 2013, 2014, 2015; Hills, 1980; Lee et al., 1984; Lobjois & Cavallo, 2007, 2009; O'Neal et al., 2018; Oxley et al., 2005; Staplin, 1995; Young & Lee, 1987).

Physiological characteristics can also affect motion and depth perception: Banister and Blackburn (1931) measured IPDs of 258 students and investigated the correlation with the students' performance at ball games. The outcome suggests that students with larger eye separation perform better at ball games. This was explained with the superior stereoscopic vision that results from the wider distance between the eyes, increasing the effectiveness of depth cues inferred from binocular disparity and convergence (see Section C.4). The binocular disparity has indeed been demonstrated to be useful for judging motion-in-depth and speed-in-depth in various experiments (González et al., 2010; Khuu et al., 2010).

C.5.2 Compositional Factors

Although some researchers (e.g., Caird & Hancock, 1994; Hancock & Manser, 1997; Sidaway et al., 1996) noted that certain factors lead to a "greater accuracy" of the TTC estimation, it should be specified that due to the consistent underestimation discussed in Section C.5.1, compositional factors that lead to a higher estimation of the temporal distance also increase the accuracy by implication. Consequently, it is more appropriate to classify possible effects as increase/decrease of the TTC estimation because some increasing factors may—with sufficient effect—lead to an exceedance of the actual TTC and by that lower the accuracy again.

Todd (1981), as well as Kaiser and Mowafy (1993), showed that even when lacking spatial information, it is possible to estimate the TTC by solely relying on the perceived looming of objects. That said, Todd (1981) also stated that sensitivity to observe accelerations under those conditions turned out to be extremely poor. There are many other factors and heuristics (e.g., depth cues) that can influence a person's perceptual process. Heuristics such as depth cues can support the perception and estimation process, especially in case of restricted availability of invariants, such as looming, that may result from sensory or cognitive limitations, such as in case of low contrast (DeLucia, 2004). Various studies have sought to investigate the influence of depth cues on the perception of approaching objects. Pictorial depth cues from *occlusion*, *relative size*, *relative density*, and *height in visual field* have been shown to have a significant impact on TTC estimates (DeLucia, 2004; DeLucia et al., 2003; Vincent & Regan, 1997). Depth information from the *aerial perspective*, which is effective at a great distance only, may be neglected though when setting up an experimental traffic environment (see Section C.4).

Special attention has been given to the relative size of an approaching object, with several studies suggesting that this cue can overrule perceived looming, termed the *size-arrival effect*. DeLucia (1991, 1999), DeLucia and Warren (1994), Kappé and Korteling (1995), as well as Stewart et al. (1993) have demonstrated this effect by investigating the estimated impact moment with different-sized same-type objects (e.g., squares displayed on a screen) with virtual self-motion toward the object (i.e., TTA) or virtual approach of the object toward the participant (i.e., TTC). Under identical approaching parameters, participants estimated the respective impact time for the larger objects to be more imminent than for the smaller ones. Van der Kamp et al. (1997) conducted a comparable experiment in a real environment: Four luminous balls of different diameters were approaching the participants in the dark. Interestingly, van der Kamp et al. observed the size-arrival effect solely in monocular viewing conditions and not binocularly. Caird and Hancock (1994), Horswill et al. (2005), as well as Mathieu et al. (2017) looked more application-oriented into the size-arrival effect, comparing the TTC estimates for small motorcycles, large motorcycles, compact cars, full-size cars, and vans. The participant was facing the approaching vehicles displayed on a screen that was either being rendered with simulation software (Caird & Hancock, 1994; Mathieu et al., 2017) or recorded in real traffic environments beforehand (Horswill et al., 2005). In all three studies, a significant effect was observed, with larger vehicles being estimated to be temporally closer than smaller ones.

Vincent and Regan (1997) investigated the *relative density* of surface texture and observed that a mismatch between the expansion of the approaching object's surface texture and the object's looming (i.e., the object's outline expansion) affected TTC estimates significantly, even if the mismatching expansion error was held as low as 10%. In addition to pictorial depth cues, the nonpictorial depth cue motion parallax was also found to be effective (DeLucia et al., 2003). Further nonpictorial depth cues are discussed in Section C.5.3, as cues related to a shifting focal plane and binocularity are more of a technological challenge and not related to environmental composition.

Manser and Hancock (1996) investigated the influence of the trajectory angle of an approaching vehicle and concluded that a larger angle (40°) results in a shorter TTC estimation than straight oncoming trajectories (0°). Hence, they argued that peripheral vision is less efficient in terms of extracting information regarding radial optical flow. This is supported by the findings of D. Regan and Vincent (1995), who observed more difficulties in judging approaching objects in peripheral view than in the foveal visual field.

Caird and Hancock (1994) investigated spatial distance as a variable and observed that a larger distance (61 m) of an approaching vehicle results in disproportionally higher estimates of the TTC than a closer distance (30.5 m). This is consistent with Petzoldt's (2014) findings, according to which a higher velocity (50 km/h) of an approaching vehicle, as opposed to a lower one (30 km/h), leads to a higher TTC estimation as well: For a constant TTC, a higher velocity implies a larger distance, as those parameters are inseparably linked in a linear physical relationship (*velocity* × *TTC* = *distance*). The same effect was also observed with a strongly simplified virtual environment, as the experiment by Kappé and Korteling (1995) demonstrated. Again, an explanation could be that the larger distance may appear to be psychologically less of an imminent threat and thus trigger less need for a subconscious safety margin (see Section C.5.1). Opposite to those findings stand the results of the study that Cavallo and Laurent (1988) conducted in a real environment, revealing no significant differences between estimations made for lower (30 km/h) and higher (90 km/h) velocities, unless visually impoverished conditions were included in the form of a restricted visual field or monocular vision.

The temporal distance which the participant is confronted with also plays a role in the capacity to estimate the TTC of oncoming objects. Schiff and Detwiler (1979), who analyzed a variety of TTC estimates covering 2 s, 4 s, 6 s, 8 s, 10 s, and 16 s, concluded that there is little reason in assessing TTC estimates above 10 s since participants do not seem capable of exploiting perceived information beyond that point. McLeod and Ross (1983), as well as Thomson (1983), deduced an even lower limit of about 8 s, up to which consistent estimates are to be expected. These suggested thresholds are not firm but rather approximate values beyond which estimation quality was observed to deteriorate rapidly. Furthermore, Schiff and Detwiler (1979) state that the observer's critical perceptual-motor adjustments when facing an approaching object occur within the last 4 s prior to contact anyhow.

Studies that examined the effect of observation time on TTC estimates have achieved divided conclusions. While McLeod and Ross (1983), who compared observation times of 2 s, 3 s, 4 s, 5 s, and 6 s, found no significant differences regarding estimation accuracy of the TTC, Manser and Hancock (1996) argued in their study that the duration of vehicle observation has an influence on TTC estimation. However, it should be noted that Manser's and Hancock's observation times—8.92 s, 10.42 s, and 11.92 s—were not dissociated from different spatial distances—80 m, 60 m, and 40 m, respectively—making it more likely to attribute the effect to the different spatial distances than to the small differences in observation time. Nevertheless, Groeger and Cavallo (1991), who analyzed observation times of 2 s and 6 s as an individual variable, found significantly higher and thus more accurate TTC estimates for the longer observation time than for the short one.

Besides all these factors that rather depend on depth cues and parameters of the study design, environmental conditions such as the weather can also have an impact: Snowden et al. (1998), who compared the self-perceived speed of drivers in clear, misty, and foggy conditions, noticed that subjects underestimated their speed significantly when visibility dropped. De Bellis et al. (2018), as well as Gegenfurtner et al. (1999), noted that decreased environmental brightness might lead to an underestimation of driven speed.

C.5.3 Technological Factors

Because most of the studies investigating TTC estimation are carried out with some virtual device or screen, there are technological factors that may influence the perception and should therefore be considered.

The importance of a stereoscopic view for three-dimensional tasks has already been known for centuries, as the publication by Molyneux (1690, pp. 293–294) shows:

> "And as a conclusion to the whole shall only add one experiment that demonstrates we see with both eyes at once; and 'tis, that which is commonly known and practised in all tennis-courts, that the best player in the world hoodwinking one eye shall be beaten by the greatest bungler that ever handled a racket; unless he be used to the trick, and then by custom he gets an habit of using one eye only."

With humans being used to stereoscopic images in everyday life, the reduction to monoscopic depth information will—unsurprisingly—affect their performance in judging their three-dimensional surrounding space.

Various studies sought to compare the perception of approaching objects when provided with monocular information as opposed to binocular one (e.g., Cavallo & Laurent, 1988; Gray & Regan, 1998; van der Kamp et al., 1997). For these experiments, participants observe the three-dimensional scenery partly monocularly and partly binocularly, with one eye being covered for the monocular condition. Alternatively, the effect can also be achieved by just switching between a monoscopic and stereoscopic visual representation of the scenery while participants maintain a binocular view continuously: A monocular view and a monoscopic display both result in a lack of stereoscopic depth cues, such as binocular disparity and convergence. While a stereoscopic head-mounted display (HMD) will provide binocular cues, depth cues resulting from the accommodation will still be lacking because the accommodative demand remains constant in current state-of-the-art HMDs. However, the effectiveness of accommodation is limited to very close distance anyhow (see Section C.4). This should be considered for experiments with near performance ranges, such as surgical simulations (see Satava & Jones, 2003).

Overall, participants performed better when being provided with binocular information, though Cavallo and Laurent (1988) observed an advantage predominantly at closer distances, observing that the effectiveness of binocular disparity was confined to about 75 m in their study. In the event of a small approaching object within the range of a few meters, Gray and Regan (1998) found TTC estimations primarily to rely on binocular cues, concluding that estimation accuracy may significantly improve when both monocular and binocular information is available to the observer. Van der Kamp et al. (1997) also observed the superiority of TTC estimations when binocularity was added. They demonstrated that in the case of binocular vision, lacking information about the relative size of approaching objects (see Section C.5.2) became irrelevant—at least within the close range of 2 m that was investigated in their study.

For now, HMDs provide solely a reduced field of view (FOV), which can affect visual perception. Cavallo and Laurent (1988) compared the impact time estimations of participants who headed toward a stationary object while having a normal visual field and while having a reduced FOV limited to the foveal and parafoveal visual field of about 10°. Estimates with full visual field turned out to be significantly more accurate. However, a post hoc analysis revealed that the difference was only significant for inexperienced drivers. No significant difference was observed for experienced drivers, although it must be noted that the *p*-value ($p = .051$) was barely above the set significance threshold ($p < .05$).

The brightness and contrast of the display may affect the perception of velocities: Takeuchi and De Valois (2000) investigated brightness as a factor when judging velocities of objects displayed on a screen and observed that a lower brightness might lead to a reduced capability of discerning differences between two different velocities. Anstis (2003), Blakemore and Snowden (1999), Stone and Thompson (1992), and Thompson (1982) manipulated the contrast of the visual display in their studies and showed that a reduced contrast typically produces the perception of the object moving at a lower speed.

A large number of studies investigated TTC estimations by making the approaching object virtually vanish at a predefined moment, asking the participants to estimate at what moment the object would have made contact with them by extrapolating the afore-seen motion (see Section C.3). Because the event of a suddenly disappearing vehicle is not very lifelike, Hancock and Manser (1997) compared virtual scenarios in which the approaching vehicle either vanished "magically" into thin air or disappeared through occlusion by driving behind a bush. The results indicated that participants had significantly higher estimation accuracy in the event of occlusion than in the vanishing mode.

Some researchers wondered about the influence of texture integrated into the simulation environment. López-Moliner et al. (2007) examined the effect of surface characteristics of an approaching object on the TTC estimation within a simplistic simulation and concluded that no significant differences could be observed between textured and nontextured surfaces. DeLucia et al. (2003) also argued that generally a richer visual texture of the object or the background surface does not influence TTC estimation performance. Li and Laurent (1995) showed with their experiment conducted in a real environment that an increased texture of an approaching ball did not influence the TTC at which the participant decided to dodge the ball. However, they observed a significant speed increase of the participant's evasive movement when the ball held a texture than when the ball surface was left blank. Studies that observed influences of texture on TTC estimates involved intentional deficiencies in visual representation, such as an expansion of the surface texture that mismatched the perceived looming of an approaching object (Vincent & Regan, 1997) or textured background that compensated for insufficient image contrast (Blakemore & Snowden, 2000).

C.6 Discussion

C.6.1 Influencing Factors

In conclusion, a combination of many factors is incorporated into the judgment of TTCs. An overview of factors discussed throughout Section C.5 is given in Table C-3.

The research field relating to the perception of oncoming objects was stimulated by Lee's concept of optical flow and looming introduced in 1976 and his controversial hypothesis that human TTC estimations rely solely on this monocular cue. Studies that provide participants with solely two-dimensional information may conclude that TTC estimates are made uniquely based on them. However, it is a misconception to believe that other

Table C-3. Overview of factors that may affect the human perception of oncoming objects.

Human factors	Compositional factors		Technological factors
Age	Height in visual field	Depth cues	Abstractness of virtual events
Driving experience	Motion parallax		Brightness
Erring in safety direction	Occlusion		Contrast
Gender	Relative density		Field of view
Physiological characteristics	Relative size		Monoscopy / stereoscopy
Risk tolerance	Approaching angle	Observer-object relation	Fixed / dynamic focal plane
	Approaching velocity		Texture
	Looming		
	Observation time		
	Spatial distance		
	Temporal distance		
	Visibility conditions		

factors and stimuli—once available—do not weigh in. During the 1980s and 1990s, a wave of experiments followed, aiming at identifying factors that may or may not influence TTC judgments. Numerous factors linked to human individuality, study design, experimental environment, and technological aspects were successfully identified and can be partially linked to each other. That said, with recent evolution in the technological field and also in regard to human factors (e.g., women accounting for a larger driver proportion than three decades ago), a new investigation of some of these factors may be of interest.

Substantial differences were observed regarding TTC estimations when the representation of the scenario was realistic and contextual as opposed to abstract experiments that involved simplistic environments. While the results of the realistic experiments varied fairly little with a consistent underestimation of about 75% of the actual TTC (see Section C.5.1), the simplistic studies covered average estimations ranging from 50% (e.g., Schiff

& Detwiler, 1979) up to 200% (e.g., DeLucia et al., 2003) of the actual TTC. Although quality variations of virtual textures showed no influence on TTC estimates (see Section C.5.3), the observation above indicates the importance of an overall realistic environment to achieve realistic results.

When investigating human perception and behavior, human factors will always play a role given individual variability. Those factors may be invariant (e.g., gender), mutational (e.g., age or some physiological characteristics), or trainable (e.g., driving experience or risk tolerance). That said, the factors are not always dissociated from each other: For example, driving experience and risk tolerance are entangled with both gender and age. However, since in the study reported by Recarte et al. (2005) no gender-specific differences could be observed when men and women had the same driving experience, one has to ask if gender can be considered as a factor at all and not merely as an indicator of driving experience and risk tolerance. In the light of the steady decrease in the gap between male and female driving experiences over recent decades (see McGuckin & Fucci, 2018), findings of numerous studies from the 1980s and 1990s may be outdated, and a new evaluation of gender-specific differences is encouraged.

Compositional factors may support (or also corrupt) the participants' perception of their environment and, as a result, affect the task at hand. These factors may be related to the perception of dynamic changes (such as looming), the perception of depth (such as occlusion, relative size, relative density, height in visual field, and motion parallax), or simply to the experimental design (such as approaching angle, approaching velocity, spatial distance, temporal distance, and observation time). When the experiment is conducted outside in a real environment, changing light and weather conditions may also affect the experimental outcome. Compositional factors may interact and support one another and especially compensate for various deficient invariants and heuristics.

Human beings experience limitations when it comes to the extent of the temporal distances up to which reliable estimates can be expected. While Schiff and Detwiler (1979) noted a limitation of 10 s and McLeod and Ross (1983) and Thomson (1983) set the limit at about 8 s (see Section C.5.2), an apparent inconsistency of estimates can already be observed at around 7 s in Caird's and Hancock's (1994) experiment. In general, the longer the TTC is set, the wider the estimation variance is to be expected, and the estimates fall into a certain arbitrariness beyond a certain point.

There are technological factors that impede an experimental scenario within a virtual environment to be perceived the same as it would have been within a real environment. Examples are monoscopic displays instead of stereoscopic ones (that provide binocular depth information), a fixed focal plane instead of a dynamic one, as well as a reduced FOV. Furthermore, some attention should be given to the technical settings of the visual display. Parallels can thereby be drawn between the observed effects of poor visual conditions in a real environment, such as darkness or fog (see Section C.5.2), and those observed due to certain technical limitations of the visual display, such as reduced brightness or contrast (see Section C.5.3), that produce the same perceptual effects regarding TTC estimation.

Many of the different factors are considerably intertwined with each other. For example, the effect of a given FOV on TTC estimates may depend on participants' driving experiences, much as the impact of an approaching object's relative size or velocity on TTC estimates will potentially depend on whether the viewing conditions are monoscopic or stereoscopic. In summary, the complexity and quantity of influencing factors warrant further research to determine in what way different factors interact with each other and what relevance has to be attributed to each of them when collecting human performance data in a virtual environment.

C.6.2 Comparison of Real and Virtual Environments

C.6.2.1 Between-Study Comparison

Virtual environments as a tool represent a mixed bag for perceptual research: Even though they are valuable for investigating scenarios that cannot be easily investigated otherwise, they also possess the potential to distort the collected data. Most of the studies discussed across Section C.5 were conducted using some display system or VR device, without these setups having been validated against real environments. Researchers may observe the same effect over and over again when investigating a specific phenomenon, without taking into account that the effect may be related to the design of the experimental setup. The size-arrival effect, for example, with larger objects leading to shorter TTC estimates (see Section C.5.2), was observed by Caird and Hancock (1994), DeLucia (1991, 1999), DeLucia and Warren (1994), Horswill et al. (2005), Kappé and Korteling (1995), Stewart et al. (1993), as well as Mathieu et al. (2017). Given these numerous experiments with varying study designs and parameters, one might believe this effect to be sufficiently documented. However, it should be noted that all of these experiments have in common of having used some sort of monoscopic display with one single focal plane in close proximity to the participant. Van der Kamp et al. (1997) conducted an experiment in a real environment with a study design similar to some of the ones just mentioned. However, because this experiment was conducted in a real environment, and no screen was involved, binocular information was available to the participants, and the focal plane was shifting with the approaching object. While the size-arrival effect was likewise observed when the participant's view was artificially reduced to monocular vision, a binocular viewing condition abolished this effect, emphasizing the relevance of stereoscopic vision for this kind of task. Study results collected by DeLucia (2005), in which participants were exposed to monoscopic as well as stereoscopic displays, also strongly suggest that binocularity may abolish the size-arrival effect.

Contrary to Caird and Hancock (1994), Kappé and Korteling (1995), as well as Petzoldt (2014), who conducted their studies in a virtual environment using a monoscopic display, Cavallo and Laurent (1988) found no influence of velocity on impact time estimates in a real environment (see Section C.5.2). Although Cavallo and Laurent were investigating the TTA, while the other three experiments primarily studied the TTC (the study by Kappé and Korteling involved both TTC and TTA measures, but an interaction effect was not investigated), the difference can still presumably be traced to the technological characteristics of the setups: Under impoverished visual conditions (i.e., reduced FOV or lacking binocularity), Cavallo and Laurent reproduced similar effects as those that were observed in the three studies using virtual environments.

Experimental environments using displays for visualization of the experimental scenario often provide a reduced FOV that is, in general, significantly smaller than the human far peripheral vision can handle. The experiment conducted by Cavallo and Laurent (1988) in a real environment comparing normal viewing conditions to an artificially reduced FOV (10°) revealed differences in TTA estimations (see Section C.5.3). A post-hoc analysis indicated the differences to be solely significant for a group of inexperienced drivers, and not for an experienced one. This shows that different sources of information may be employed when preconditions differ. Subjects can also rely on nonapparent information sources to fulfill their task as Cavallo's and Laurent's experiment demonstrated: Participants sitting in a vehicle and approaching head-on a stationary object within their central visual field also processed visual information from the periphery. Although it is conclusive in this case that participants may involve the peripheral visual flow to deduce speed (see Warren & Hannon, 1988), there are likely research scenarios in which it may be very complex to identify all influencing factors beforehand. If some of these factors are not correctly incorporated into the experimental setup, it will complicate the transfer of the study's findings to real environments. The current example of Cavallo's and Laurent's experiment illustrates how laborious the validation processes of experimental setups are: Although a setup could have been possibly validated with a group of experienced drivers and no significant difference would have been found between estimates with different fields of view, a small parameter variation, such as the change of the sample's driving experience, could require a new validation of the setup.

C.6.2.2 Within-Study Comparison

Recarte et al. (2005) conducted an experiment in which participants were asked to estimate the impending collision from inside of a vehicle in a real environment, as well as when sitting in front of a screen while watching recorded footage of scenarios from the same point of view and perspective. They noted that estimates made in reality had a lower variance and also a higher correlation with the actual TTA than estimates based on the recorded videos, which was observed for all 16 different experimental conditions arising from combinations of four different speeds (60 km/h, 80km/h, 100km/h, 120 km/h) and four different distances (75 m, 100 m, 125 m, 150 m). While a higher similarity of experimental *distances* led to a higher correlation of the estimates within the respective environment in both environmental conditions, a higher similarity of experimental *speeds* led to a higher correlation of the estimates only in the real environment and not in the virtual one. This suggests that participants experienced more difficulties in differentiating between velocities than in processing different distances when visualized on a screen, as the cross-comparison with the reality setting indicates. Despite the efforts to provide the participants with comparable experimental settings and parameters in the two environments, some differences in estimations were evident, suggesting that information sources in the two environments were employed differently. Moreover, apparent access to additional information sources in reality that were simply not available in the screen setup seems plausible. Despite the highlighted differences between the estimations in both environments, it should nevertheless be mentioned that overall, there was also high comparability of the results in many other characteristics.

Schwebel et al. (2008) conducted one of the very few reported examples of VR validation against a real road traffic environment, with the intention of using the system for research and training purposes: They designed a moderately complex pedestrian simulator, in which the participant stands on an artificial physical curb and faces the scenery of a two-way virtual street. The scenery, which contained cars approaching from both sides, was displayed on three monitors that were arranged in semicircular alignment toward the user. The participant was asked to initiate the street crossing by stepping off the curb whenever feeling safe to do so, stepping onto a pressure plate. This triggered the scenery to morph from the first-person view to a third-person view, and the participant observed an avatar cross the street at a constant personalized speed that was measured beforehand and adjusted for each participant individually. Schwebel et al. conducted an experiment with 74 adults and 102 children that aimed at comparing the behavior in the simulator to real street crossings and ultimately validate the setup as a tool dedicated to training children toward a safe street-crossing behavior. In the experiment, children were asked to perform three different types of tasks: initiating virtual road crossings within the described simulator setup, verbally indicating road crossings facing a real road (*shout task*), and physically indicating road crossings facing a real road by doing the first two steps only (*two-step task*). Adults had an additional fourth task, in which they actually crossed the real road at their discretion.

Schwebel et al. (2008) argued that construct validity was demonstrated through a significant correlation between the behavior in the real and the virtual environment as well as through developmental differences between adults and children observed in both environments. Correlations of parent-reported child temperament and the child's crossing behavior suggested a convergent validity, while the participants' self-reported perception of realism in the virtual environment was an indicator for the simulator's face validity.

Despite the numerous parallels shown between the two environments, it may be hasty to conclude that the simulator is fully validated. Although developmental differences, behavioral patterns, and a realistic impression of the virtual environment are all highly relevant factors when validating a VR setup, they do not provide quantitative information about the perceptual similarity between the real and the virtual environment. For this, Schwebel et al. analyzed two variables: for adults, the *safety gap* (i.e., the TTC after the participant successfully crossed the street), and for children, the *start delay* (i.e., the time elapsed between the last vehicle passing and the initiation of the crossing itself).

The generated *safety gap* may indeed give insight into the pedestrian's perception of approaching vehicles. However, it must be noted that the experimental differences in the study design between the real and the virtual environment add additional variables and uncertainties: In reality, participants were capable of adjusting their walking speed by continuously revaluating the approaching vehicle while crossing and thus speed up or slow down. In the virtual environment, the crossing speed was rigid and visualized by an avatar without the user being capable of influencing the speed during the crossing process.

The *start delay* analyzed for the children, on the other hand, provides very little insight into their effective perception of approaching vehicles. This metric serves predominantly as a proxy measure of the cognitive processing time for judging the next approaching vehicle before initiating the crossing (Thomson et al., 2005). The importance of this metric is beyond question: Plumert et al. (2004) demonstrated that children show a riskier crossing behavior by having a greater *start delay* than adults, despite accepting the same gap in between two passing vehicles. Nevertheless, this metric cannot reveal *how* pedestrians perceive and estimate the approaching vehicle's distance and velocity.

Most importantly, it should be pointed out that the comparison of these two variables—*safety gap* and *start delay*—conducted by Schwebel et al. (2008) investigated the *correlation* between the real and the virtual environment. This means that a significant correlation between the two environments provides minimal information, giving insight solely into the similarity of the *tendency* in both environments. For example, in the event of a significant correlation, a participant with a shorter *safety gap* in the real environment will most likely also have a shorter *safety gap* in the virtual one; the significant correlation will not tell whether the values are identical nor even allow to tell whether there is a fix relative relation between the two data sets. It merely tells that the two data sets have the same tendency. Simulators that underwent such a validation process would allow solely nominal or ordinal measurement scales for research questions (see S. S. Stevens, 1946). For all of these reasons and despite the impressive extent of the study by Schwebel et al., it would be premature to declare an overall validity of the setup solely based on these observations.

David C. Schwebel (University of Alabama at Birmingham, AL, U.S.) kindly provided the raw data collected in his study dating from over a decade ago which allows a more in-depth comparison between the road crossings using the VR system and those observed in reality. The goal of the further analysis provided by the work presented here was to determine whether there are significant differences—despite the observed correlation—in the participants' behaviors when crossing the road in reality and when simulating the crossing in that specific virtual environment. Because only adult participants effectively crossed the road during the experiment and the sample of children skipped this task, the analysis was reduced to the adult sample. Two participants were excluded due to incomplete data, making the reanalyzed sample consist of 72 participants in total.

In the first step of the reanalysis of Schwebel's report, the scenario parameters in the real environment were compared with those in the virtual one: The variables that were examined were the density of the traffic and the participants' road-crossing times. The virtual road-crossing scenario displayed a replicable number of cars per minute. The scenario in reality, on the other hand, was somewhat difficult to control, and the number of vehicles varied somewhat randomly due to changes in traffic conditions. Overall, the virtual crossing scenario contained, on average, 11.95 cars/min (SD = 2.06), while in the real world, participants faced, on average, 13.90 cars/min (SD = 2.32). A paired-samples t-test confirmed a significant difference between the real and virtual car densities: $t(71) = 5.271$, $p < .001$, $r = .53$.

The participants' road-crossing times also revealed significant differences between the two environments: The crossing duration in the virtual environment that was individually determined for the participants before the experiment was, on average, 5.5 s (SD = 0.93). In contrast, the crossing duration in the real environment remained variable during the crossing itself and was, on average, 4.73 s (SD = 0.67). Because the requirements for normal distribution (see A. Field, 2013) were not fulfilled for the crossing times, the samples were analyzed using a Wilcoxon signed-rank test: $T = 108.5$, $p < .001$, $r = .56$. The differences in road-crossing time and traffic density between the real and the virtual environment (shown with box plots in Figure C-4), as well as the fact that road-crossing speeds remained constant in the virtual environment while variable and adaptable to the current road situation in the real environment, complicate a proper comparison between the two environments.

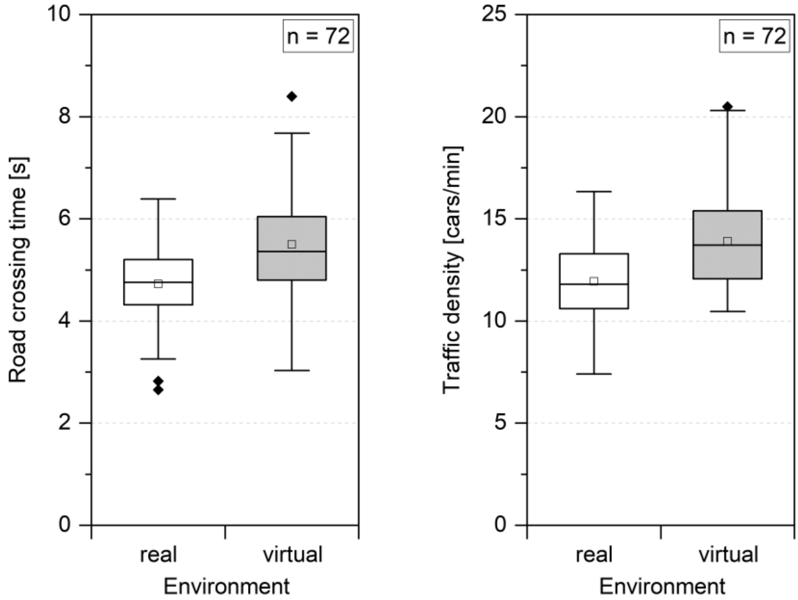


Figure C-4. Box-and-whisker plots show the distribution of participants' road-crossing durations (left) and the traffic density the participants were confronted with during the experiment (right). Box-and-whisker plot structure: Values between the lower and upper quartiles are represented by a box, while whiskers identify estimates within 1.5 times of the interquartile range (IQR) of the lower and upper quartiles. The horizontal line in the box shows the median, and the small square shows the arithmetic mean. Outliers are plotted with diamonds.

The participants' average remaining time to the approaching car after crossing the road (i.e., the safety gap measured as TTC) amounts to 8.95 s (SD = 2.7) for the real road crossing, whereas in the simulated crossings, the safety gap was at 4.97 s (SD = 1.26) only. A paired-samples t-test confirmed the significant difference between the safety gaps in the two environments: $t(71) = 11.967$, $p < .001$, $r = .82$. Considering that the road-crossing time itself differed significantly between the two environments, a further look was given to TTCs at the crossing initiation, neglecting the subsequent crossing durations. Yet, also in this case, the difference between the real environment (M = 13.67 s, SD = 2.66) and the virtual one (M = 10.47 s, SD = 1.78) remained significant: $t(71) = 8.956$, $p < .001$, $r = .73$. The distribution of the safety gaps at crossing initiation and those at crossing completion are shown via box plots in Figure C-5.

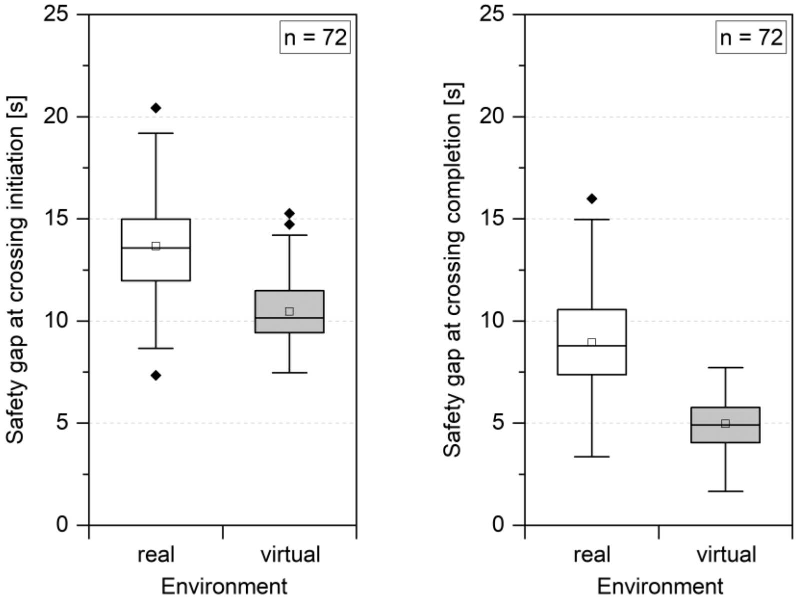


Figure C-5. Box-and-whisker plots show the distribution of the TTC at crossing initiation (left) and after crossing completion (right). See Figure C-4 for the structure of the box-and-whisker plots.

Further attention was given to the start delay that turned out to be, on average, somewhat higher in the virtual setting (M = 0.84 s, SD = 0.36) than in real road crossings (M = 0.58 s, SD = 0.43). A Wilcoxon signed-rank test suggests a significant difference between the two samples: T = 1888.5, p < .001, r = .29. Of particular note are some occurring negative start delays in the real road-crossing scenario, whereas all of the start delays in the virtual setting were positive (see Figure C-6). The negative start-delay values signify that participants in the real scenario at times initiated the crossing (i.e., stepping onto the street) before the last car had entirely passed. In the virtual environment, the participants always waited for the virtual car to have entirely passed.

Finally, the gap size chosen by the participants to cross in between two vehicles was compared: For the real road crossings, participants chose average gap sizes between cars of 14.65 s (SD = 3.13), while in the virtual road crossings, the average gap size amounted to only 9.72 s (SD = 1.26). A paired-

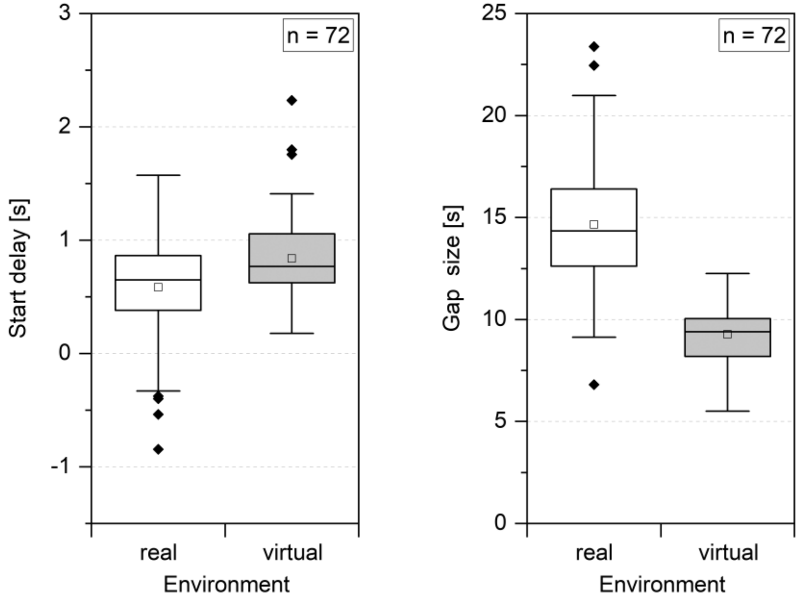


Figure C-6. Box-and-whisker plots visualize the distribution of participants' start delay (left) and the chosen gap size (right). See Figure C-4 for the structure of the box-and-whisker plots.

samples t-test confirmed a significant difference: $t(71) = 14.767$, $p < .001$, $r = .87$. This is also visualized with box plots in Figure C-6.

In conclusion, it must be noted that the VR system, introduced and examined by Schwebel et al. in 2008, showed, after all, some essential and significant differences when compared with real-world behavior. Participants demonstrated a much riskier behavior in the virtual environment by choosing significantly smaller gap sizes and leaving significantly smaller safety gaps when crossing the road. This suggests that participants react to approaching vehicles in virtual environments differently than in real environments. Eventually, this can be due to a differing perception of the vehicles' approach in virtual environments or also due to experiencing a different sense of safety in virtual environments, not perceiving the imminent threat of the vehicle in the same way as in real environments. In addition, the lower—partly negative—start delay at real road crossings suggests that participants act differently—some might say more efficiently—in their familiar (i.e., real) environment.

This VR system example demonstrates that even allegedly validated simulators may bear essential differences to real-world settings and points out that published results of experiments carried out with various VR systems that have not been thoroughly validated beforehand have to be handled with care. This does not imply discarding published results but shall increase the researchers' sensitivity when interpreting those findings. This result shall also not question the use of VR systems for perceptual and behavioral experiments, as these systems come with tremendous advantages and open up whole new research possibilities (see Scarfe & Glennerster, 2015). However, it should emphasize the importance of the validation necessity for VR systems that are intended to be used in behavioral research. Thereby, the validation process should be adopted to the proposed research or education purposes. The degree of similarity regarding perception and behavior between real and virtual environments will indicate which measurement scale may be applicable for research questions (see S. S. Stevens, 1946). So even with significant differences between the two environments, a virtual environment with suitable arrangements may still be a valuable tool for behavioral research or be validly applied for educational purposes, for example, for developing children's cognitive skills in traffic environments (e.g., Demetre et al., 1993; Schwebel et al., 2008).

D ARTICLE IV – FULL REPRINT

Road Crossing Decisions in Real and Virtual Environments: A Comparative Study on Simulator Validity

Revised reprint from Feldstein, I. T., & Dyszak, G. N. (2020). Road crossing decisions in real and virtual environments: A comparative study on simulator validity. Accident Analysis & Prevention, 137, Article 105356.

Abstract. Virtual reality (VR) is a valuable tool for the assessment of human perception and behavior in a risk-free environment. Investigators should, however, ensure that the used virtual environment is validated in accordance with the experiment's intended research question since behavior in virtual environments has been shown to differ from behavior in real environments. This article presents the street-crossing decisions of 30 participants who were facing an approaching vehicle and had to decide at what moment it was no longer safe to cross, applying the step-back method. The participants executed the task in a real environment and also within a highly immersive virtual environment involving a head-mounted display (HMD). The results indicate significant differences between the two settings regarding the participants' behaviors. The time-to-contact (TTC) of approaching vehicles was significantly lower for crossing decisions in the virtual environment than for crossing decisions in the real one. Additionally, it was demonstrated that participants based their crossing decisions in the real environment on the temporal distance of the approaching vehicle (i.e., the TTC). In contrast, the crossing decisions in the virtual environment

seemed to depend on the vehicle's spatial distance, neglecting the vehicle's velocity. Furthermore, a more in-depth analysis suggests that crossing decisions were not affected by factors such as the participant's gender or the order in which they faced the real and the virtual environment.

D.1 Motivation

Investigation of human behavior schemes in traffic increasingly gains relevance, also due to technological progress in the field of automated and assisted driving and the consequent changes in urban traffic environments. Innovative *advanced driver assistance systems* (ADAS) help with the challenges of modern traffic scenarios. Furthermore, intelligent urban traffic management may reduce traffic accidents and increase efficiency. These systems rely on the results of behavioral research in order to work in a safe, reliable, and efficient way.

Because of their vulnerable nature, pedestrians require particular attention, playing an essential role in traffic-related behavioral research. The use of driving simulators has been a common approach for several decades to investigate traffic scenarios without putting the participants at risk. However, due to the composition of most driving simulators, studies focus primarily on the behavior of *drivers*. Nonetheless, it is possible to transfer the driving simulator concept to pedestrian behavior research as well: Virtual reality (VR) setups that can provide an interactive scenario from a pedestrian's perspective will become an indispensable tool for the investigation of pedestrian behavior in a broad range of traffic scenarios, ultimately helping to increase overall pedestrian safety.

D.2 Validation Necessity of Virtual Reality Research Tools

Simpson et al. (2003) developed one of the first VR systems for the investigation of road-crossing scenarios from a pedestrian perspective. They took study participants' reports about high senses of presence as anecdotal evidence for the adequacy of the simulation. That said, they also noted that the collision rate in their virtual street-crossing experiment was higher than the rates in comparable scenarios in literature that were observed in reality.

One of the few reported attempts to validate a VR system for road-crossing scenarios was published by Schwebel et al. (2008), who developed a simulator for road safety education of children. For this, they ran a two-part experiment that compared road crossings in a virtual environment to road crossings on an actual street. They investigated several parameters and argued that construct validity was given since road-crossing decisions in the virtual environment correlated with the decisions that were observed on the real street. Schwebel et al. further substantiated their argument with the developmental differences between participating adults and children that could be observed equally in the real and virtual environments. Furthermore, they claimed face validity due to the participants' self-reported experiences of realism as well as convergent validity due to the correlation of the children's crossing behavior in the simulator with the children's temperament as reported by their parents. With all these parallels between the real and the virtual environment, one could argue that the simulator is fully validated. However, a recent more in-depth analysis of the experiment's raw data revealed that while the road-crossing decisions were indeed correlating in the two environments, the absolute values were significantly different, revealing a remarkably riskier road-crossing behavior in the simulator than on the real street as broadly laid out by Feldstein (2019).

Divergent behavioral patterns are not limited to virtual environments but may also be observed in other display systems, such as demonstrated in an experiment by Recarte et al. (2005). They evaluated the judgments of impending collisions from inside of a real vehicle and compared those to judgments of impending collisions that were recorded and displayed on a screen. While there were great similarities observed between the behavior and decisions in the real environment as compared to the displayed one, the analysis also revealed certain discrepancies: There were greater variances of participants' collision judgments as well as participants' apparent difficulties in discerning different approaching velocities when the scenario was displayed on a screen. Feldstein (2019) juxtaposed reported findings of impending collision judgments that were collected in real environments to those that were generated using a display system and found substantial differences between the different studies' outcomes.

Although these examples show that differences in human perception and behavior between real and virtual environments are an issue, thorough simulator validations are unfortunately often neglected in research. There could be a variety of reasons for the possible discrepancies between real and virtual environments. For example, low immersion, and with this, a low sense of presence may alter participants' emotional responses to threats in the virtual environment (Cummings & Bailenson, 2016; Slater, 2009), and consequently lead to a riskier behavior when compared to scenarios in real life. Additionally, sensory stimuli provided in virtual environments often differ from those available in real environments (e.g., Grechkin et al., 2010; Jones et al., 2016), and thus may lead to divergent reactions. Hence, it is crucial to run a validation process with a new VR system beforehand in order to assure the ecological validity of behavioral effects detected.

D.3 Related Background on the Perception of Approaching Vehicles

Does the perception of approaching vehicles differ between current state-of-the-art VR systems and the real world? This question calls for the consideration of a large number of research criteria and parameters, including the cognitive processing of perceived stimuli, the typical measurement methods used, and the potential influencing factors (here broken down into human, compositional, and technological factors).

D.3.1 Cognitive Processing of Perceived Stimuli

When investigating behavior in urban traffic scenarios, some sensory stimuli and perceptions are more relevant to the participants than others. While the eyes overall account for approximately 80% of all daily sensory input in an everyday situation (Dahm, 2005), the significance of the visual perception system increases in traffic scenarios: More than 90% of processed stimuli in automotive traffic are visual (Hills, 1980). According to Dahm (2005), the human sensory system perceives a data volume that—digitally converted—corresponds to about 8,000,000 bit/s. This stands in stark contrast to the human capacity of processing roughly 7 bit/s in the course of a decision-making process (see Card et al., 1983). Hence, even though vision occupies a majority of the human brain's sensory capacity, only a tiny fraction of visually perceived stimuli is actually processed consciously. Road users, however, are often unaware of their limited sensory processing abilities and may have the illusory impression of perceiving the entire environment (Cavallo & Cohen, 2001). All this emphasizes a relevance of the visual stimuli's quality within a virtual-environment system for traffic research and the necessity to validate visual perception within such an environment. When it comes to investigating pedestrian-vehicle encounters, the pedestrian's accurate visual perception of distances and velocities becomes essential. Investigated traffic scenarios that involve pedestrians are most commonly road-crossing scenarios. In these experiments, participants are typically facing a street with passing vehicles and are asked to cross the street—or indicate such a crossing intention—when feeling safe to do so. Analyzing such a road-crossing task in the course of the simulator's validation process will allow evaluating participants' perceptions and judgments of velocities and distances in the given virtual environment while ensuring application-oriented testing.

D.3.2 Terminology of Metrics

The most common terms to describe the remaining time between an observer and an object on a collision course (i.e., temporal distance) are *time-to-contact*, *time-to-arrival*, and *time-to-collision*. Although these terms are often used interchangeably in literature, a proposal to unravel the terminology suggests to use *time-to-contact* when an object approaches the observer, *time-to-arrival* when the observer approaches an object, and *time-to-collision* when both approach each other simultaneously (Feldstein, 2019). Thus, when pedestrians willing to cross a street face an approaching vehicle, the remaining time between them and the vehicle would be termed *time-to-contact* (TTC).

D.3.3 Measurement Methods

Different experimental concepts investigating the perception of impending collisions have been introduced by previous research. *Discrimination tasks* measure the psychometric function (see Green & Swets, 1966) of participants' sensitivities in distinguishing different parameters from each other (e.g., different approaching velocities). *Estimation tasks*, on the other hand, may require *coincidence anticipation* or *interceptive actions*. In *coincidence anticipation*, participants observe, for example, an approaching object for a limited sequence and subsequently extrapolate the point in time at which the collision would have taken place. When participants have to take *interceptive actions*, they have, for example, to react to an approaching object by catching or avoiding the object (see also Nahin, 2012). The experiment presented here uses an estimation task that requires a combination of coincidence anticipation and interceptive action.

D.3.4 Influencing Factors

Many studies have strived to single out factors that may influence one's perception and judgment of approaching objects. These factors may be related to human individuality, compositional parameters of the experimental design and environment, or technological deficiencies of the VR system. Some of the factors relevant to the study presented in this article are summarized in the following subsections; for a more thorough review, the work by Feldstein (2019) provides a more in-depth insight.

D.3.4.1 Human Factors

Human individual differences have been shown to affect the accuracy of TTC estimates. For this reason, investigators must be careful when comparing the results from different samples. Numerous studies have shown that participants consistently underestimate the TTC as long as the experiments involve contextual environments (i.e., real environments or sophisticated virtual environments). The average estimated TTC amounts to only 75% of the actual TTC, suggesting that the natural instinct of self-preservation with the tendency to err in the direction of safety has an impact on these types of judgments (Feldstein, 2019). This underestimation phenomenon is more pronounced for older adults than younger ones (Hancock and Manser, 1997; Schiff et al., 1992), and is also more pronounced for female participants than male ones (Caird & Hancock, 1994; Manser & Hancock, 1996; McLeod & Ross, 1983; Schiff & Oldak, 1990). On a practical basis, the effect of age on road-crossing choices has been subject to many experiments and is well documented for adults and also for children of all age groups (see Barton &

Schwebel, 2007; Connelly et al., 1998; Demetre et al., 1992, 1993; Dommes & Cavallo, 2011, 2012; Dommes et al., 2012, 2013, 2014, 2015; Hills, 1980; Lee et al., 1984; Lobjois et al., 2013; Lobjois & Cavallo, 2007, 2009; O'Neal et al., 2018; Oxley et al., 2005; Staplin, 1995; Young & Lee, 1987). The reasons for the gender difference is speculative and encourages further investigations. Possible explanations for the gender difference are that women strive for higher safety margins (see Evans, 2004; Hills, 1980; Kadali & Vedagiri, 2012; Konečni et al., 1976; Montgomery et al., 2014; Parsonson et al., 1996), that cognitive abilities related to spatial tasks may differ between men and women (see the meta-analysis by Voyer et al., 1995), and that men possess more driving experience (see Kirkham & Landauer, 1985; McGuckin & Fucci, 2018; Polus et al., 1988). Driving experience affects impact estimations, as has been demonstrated in a male-only study by Cavallo and Laurent (1988). Furthermore, participants' physiological characteristics, such as the interpupillary distance (IPD), may also have an impact on performance related to three-dimensional tasks that require perceptual-motor skills (see Banister & Blackburn, 1931).

D.3.4.2 Compositional Factors

When conducting experiments with participants who judge the TTC of an approaching object, several experimental parameters may affect the perception. These compositional factors are not crucial when comparing real and virtual scenarios—as long as the parameters are identical in both environments—but may matter if cross-comparisons are intended between different studies with varying parameters.

The vehicle velocity affected the TTC judgments in some experiments, with higher velocities evoking disproportionally higher TTC estimates (Kappé & Korteling, 1995; Petzoldt, 2014). Particular attention should be given to the chosen size of the approaching vehicle since a phenomenon, termed the *size-arrival effect*, was reported by numerous studies, observing that larger vehicles lead to shorter TTC estimates than smaller vehicles (e.g., vans vs. compact cars), underlining the importance of vehicle similarity in both environments (see Caird & Hancock, 1994; DeLucia, 1999; DeLucia & Warren, 1994; Horswill et al., 2005; Kappé & Korteling, 1995; Mathieu et al., 2017; Stewart et al., 1993).

A further factor is the participants' observation durations, with Andersen and Enriquez (2006) and Groeger and Cavallo (1991) suggesting that a longer observation time may lead to a higher estimation accuracy, while McLeod and Ross (1983) did not observe an influence of the observation time on the participants' accuracy of judging the approaching vehicle. Although the reported findings were inconsistent, researchers may be advised to give the vehicle a larger approaching phase so that participants have sufficient time to process the situation cognitively. As for the observer's viewing direction, it has been shown that participants performed better when judging vehicles approaching within the foveal visual field than the ones approaching from the periphery (Manser & Hancock, 1996; D. Regan & Vincent, 1995).

It has to be considered that the vehicle's spatial distance at which the TTC has to be judged may affect the TTC judgment, as has been demonstrated by Caird and Hancock (1994), who observed TTC estimates to be disproportionally lower for closer distances. Furthermore, the temporal distance at which the vehicle must be judged should not be chosen too high, since participants experience difficulties in judging vehicles at a temporal distance of more than eight seconds (see Caird & Hancock, 1994; McLeod & Ross, 1983; Schiff & Detwiler, 1979; Thomson, 1983). In fact, Schiff and Detwiler (1979) state that participants undergo critical perceptual-motor adjustments within the last four seconds prior to impact.

D.3.4.3 Technological Factors

While a virtual-environment system may also perform well with only a monoscopic view (Jiang et al., 2017), several studies carried out in the past have demonstrated perceptual deficiencies, for example, regarding the judgment of approaching objects under monocular conditions (Cavallo & Laurent, 1988; Gray & Regan, 1998; van der Kamp et al., 1997). These probably are due to the lack of binocular depth information that is provided in stereoscopic systems.

Since stereo cues and accommodation are not the same in head-mounted displays (HMDs) as in the real world, the stereoscopic rendering of the scenery will affect the quality of the perceived representation of the environment. The fixed focal plane of an HMD removes accommodative depth information. Accommodation cues may be considered to be particularly irrelevant to street-crossing investigations since these cues are generally effective solely within a distance of two meters (Cutting & Vishton, 1995).

State-of-the-art HMDs provide a reduced field of view (FOV). This is of relevance when experimental scenarios are selected that involve, for example, vehicles approaching from the periphery but also tasks that involve continuous self-motion, especially at higher velocities such as in driving simulators: The peripheral optical flow may support the awareness of the observer's own speed (Cavallo & Laurent, 1988). However, this aspect presumably plays a marginal role for moving velocities of pedestrians.

Attention should also be given to the low-level ergonomics of the hardware components worn by the user, such as the HMD, the headphones, and the motion-capture suit, since simple mechanical discomfort may affect the user's task performance. For example, Willemsen et al. (2004) demonstrated that their HMD's weight and inertia contributed partially to a compressed distance perception in virtual environments. An essential element of a VR system is the virtual environment itself, with the investigator facing the question of what degree of realism needs to be implemented. Generally, the richness of visual texture does not affect the participants' judgments of an approaching object's TTC (DeLucia et al., 2003; Li & Laurent, 1995; López-Moliner et al., 2007). However, while Li and Laurent (1995) observed no difference in the timing of avoiding an approaching object, they reported that the participants' evasive movements were significantly faster when some texture was added to the object as compared to a blank object surface. This might indicate that while the estimated TTC was not affected by the texture, the texture still added a relevant component, ultimately increasing the perceived threat by the oncoming object.

Furthermore, the content of the environment needs to be considered since the degree of three-dimensional virtual detail may affect the participants' perceptions and judgments. It has been shown that the approaching object's looming is sufficient to estimate the impact time (Kaiser & Mowafy, 1993; Todd, 1981). Nonetheless, other sources of information—when available—may also affect the judgment: Depth cues, for example, are essential heuristics and support the perception of approaching objects, with *occlusion*, *relative size*, *relative density*, and *height in field* affecting TTC estimates (DeLucia, 1991, 2004; DeLucia et al., 2003; Vincent & Regan, 1997). Hence, the virtual scenario should provide detailed depth cues similar to the real environment when such a comparative experiment is conducted.

Quite a few studies suggest that reduced brightness (de Bellis et al., 2018; Gegenfurtner et al., 1999; Takeuchi & De Valois, 2000) and reduced contrast (Anstis, 2003; Blakemore & Snowden, 1999; Snowden et al., 1998; Stone & Thompson, 1992; Thompson, 1982) may lead to lower velocity estimates—be it the observer's own moving speed or that of an approaching object.

D.4 Method

D.4.1 Virtual-Reality System

The virtual-environment system used in the presented experiment was developed at the Technical University of Munich (TUM) for the investigation of pedestrian behavior in traffic encounters. The simulator hardware consists of the following main components: a stereo HMD, stereo headphones, an optical motion-capture system in combination with a full-body motion-capture suit, and a control center handling the computational processing (Figure D-1). The HMD is an *Oculus Rift Development Kit 2* (Oculus VR, Menlo Park, CA, U.S.), displaying a stereoscopic image, whereas the headphones are from *Bose*, model *QuietComfort 25* (Bose, Framingham, MA, U.S.) with built-in active noise canceling (ANC). The *Vicon* motion-capture system (Vicon Motion Systems, Yarnton, U.K.) was combined with a custom-made, elastic, one-size-fits-all full-body suit that could be worn over regular clothing and was equipped with infrared LED markers in predefined positions, substituted for the less effective reflective markers that are typically used with Vicon systems (Feldstein et al., 2018). The virtual environment was created using the *Silab* traffic simulation software (WIVW, Würzburg, Germany). Further information regarding the technology of this VR system can be found in Feldstein et al. (2016, 2018) as well as in Lehsing and Feldstein (2018).

In the experiment presented here, the viewing conditions were not adjusted to the participants' individual physiology. The identical self-avatar was used for all participants, and the eye height was set to 180 cm. Since the self-avatar remained identical for all participants, positions of the various body limbs and their relative distribution were not optimal for most participants. That said, the quality of the self-avatar played a subordinate role since the focus of this study was on the perception and judgment of approaching vehicles from an egocentric perspective, and no direct contact

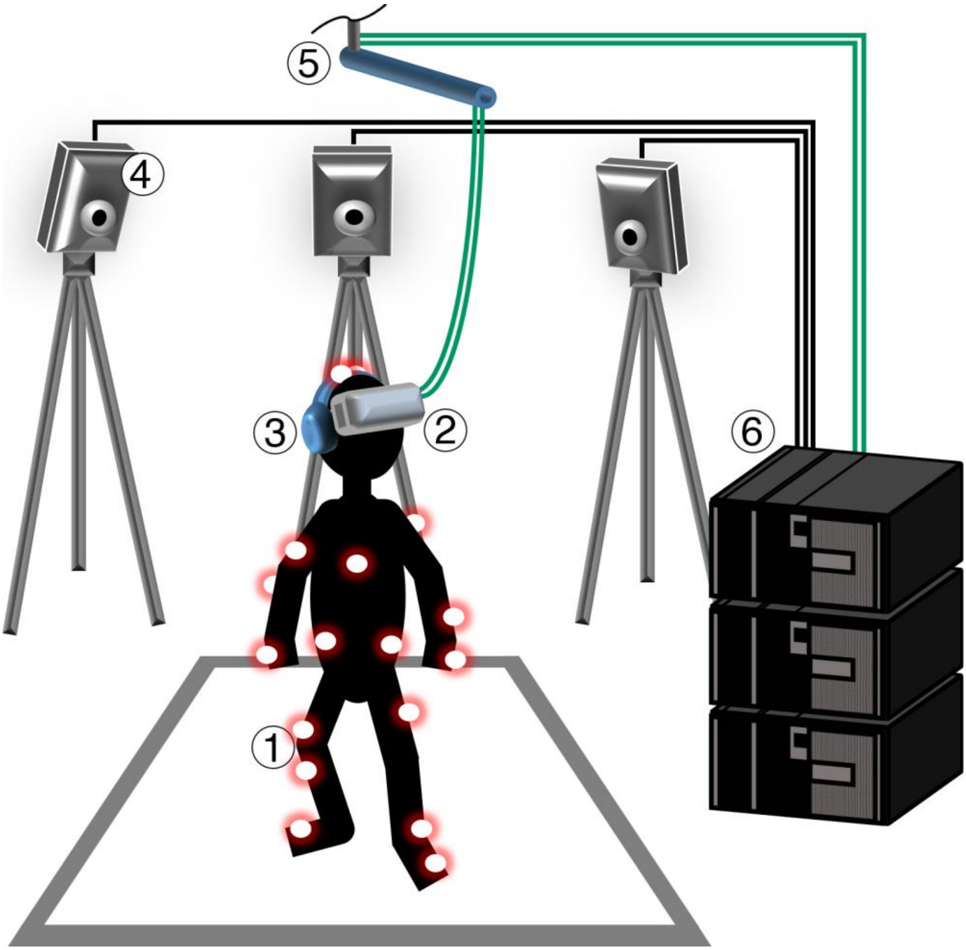


Figure D-1. Simplified illustration of the pedestrian simulator: (1) motion-capture suit, (2) stereo head-mounted display, (3) stereo headphones, (4) motion-capture cameras, (5) dynamic cable-routing system, (6) control center.

or interaction with the self-representation was required. The IPD of the stereoscopic virtual representation was fixed at 63.5 mm, which corresponds to the average human IPD (Best, 1996; Dodgson, 2004).

D.4.2 Experimental Design

When seeking to compare the perception of approaching vehicles in a virtual environment to the one in reality, it makes little sense to ask a participant to numerically estimate physical dimensions such as temporal distances or approaching velocities of oncoming vehicles. Humans generally experience difficulties in attributing a numerical value to such parameters due to innate limitations in making such judgments (Guzy et al., 1991; Hills, 1980; Loftus, 1979). They perform better when asked to match perceptually (e.g., comparing two values with each other) or to react intuitively to a perceived scenario (e.g., verbally or action-based). It should be considered, however, that road-crossing behavior may differ when the crossing intention is verbally expressed (*shout task*) than when the road is actually crossed, such as demonstrated in an experiment by te Velde et al. (2005). This is in accordance with the observation by Goodale and Milner (1992), who dissociated between the vision for perceptual purposes—in this case, the verbal judgment—and the vision for action—in this case, the actual crossing. This may lead to the conclusion that the action component may be necessary for such investigation purposes and that verbal judgments may not fully substitute action-based assessments when it comes to road-crossing investigations.

Furthermore, Hancock and Manser (1997) demonstrated the relevance of realistic events and scenarios within the virtual environment: In their study, participants were asked to cognitively extrapolate the approaching movement of a vehicle after the vehicle disappeared from view. Thereby, judgments for vehicles that disappeared by occlusion (e.g., driving behind a bush) resulted in higher accuracy than for those that vanished "magically" into thin air.

A useful approach to validate the simulator setup may consist of giving participants tasks that are application-oriented and action-based. These tasks need to be implementable in both environments, the real and the virtual one, allowing for subsequent comparison of quantifiable behavioral patterns. A typical traffic situation that involves pedestrians is that of a road-crossing scenario in which the pedestrian has to assess the TTC of an oncoming vehicle within a short time. Simply put, participants standing on the roadside are confronted with approaching vehicles and have to decide whether the time between the vehicle and themselves is sufficient to cross the road or not. While road-crossing tasks in virtual environments cause no complications, they bear a considerable risk in real environments. Either the investigator risks actual collisions and puts the participants' well-being at stake, or it is necessary to solely assess scenarios with the vehicle at an adequate safety distance when the pedestrian crosses the street. In the latter case, the investigator will ultimately not be able to determine the threshold of the minimum TTC at which the participants are still willing to cross the road. The determination of the threshold can be pivotal, though, since possible differences in road-crossing decisions between real and virtual environments may occur somewhere between the TTC of that threshold and the TTC of the mentioned safety distance.

For these reasons, it was decided to let the participants stand at the side of a one-way street with a single lane and let them face a vehicle approaching from the right side at a constant speed. Whenever the participants felt that it was not safe to cross the road anymore, they made a step back onto the sidewalk. Note that a location with a lowered curb is advisable to avoid the participant tripping when stepping back. This signaling technique—hereinafter termed the *step-back method*—requires participants to execute a natural body movement, which is an essential part of the perception process (Goodale & Milner, 1992; te Velde et al., 2005). Facing the approaching vehicle, the step back off the street creates a more intuitive and natural action to express a personal rejection of crossing the road from that point on than methods that merely aim at indicating a crossing intention, such as the *shout task* (e.g., Demetre et al., 1992) or the *two-step task* (e.g., Schwebel et al., 2008). The value of acceptable TTC to cross the street is determined by measuring the TTC at the moment of the step back.

D.4.3 Experimental Environment

The Garching campus of the Technical University of Munich in Germany was chosen as the experiment location and hence remodeled in the simulation software. The circularly paved road next to the building of the Department of Mechanical Engineering (Figure D-2) is demarcated from public traffic by a barrier and used by only a very small number of vehicles for parking and delivery purposes. With the road being a one-way street, there was also no risk of uninvolved vehicles approaching from the opposite direction, thus enabling the participants to concentrate fully on one side. The street has a width of 4.4 m. Participants saw the vehicle approaching from around the corner at a distance of 130 m. Figure D-3 shows the scenario of the two environments.

The study design used two different vehicles in alternating order, enabling greater time efficiency since the vehicle in the experiment required

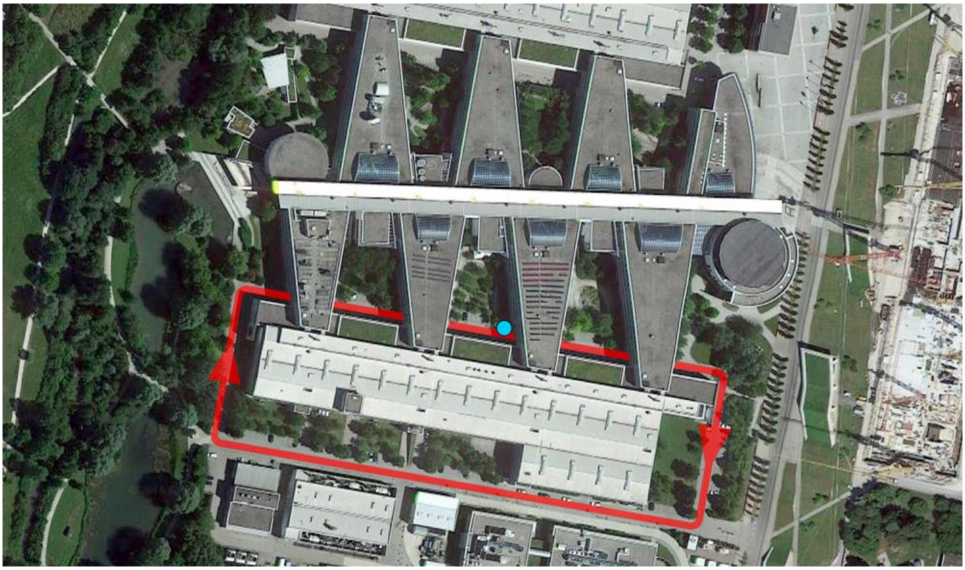


Figure D-2. Experimental area at the TUM campus (adapted from Google Maps, 2017): Arrows on the circular road indicate the circulation direction for vehicles. The spot on the circular road indicates the participant's standing position.

Figure D-3. A participant facing an approaching vehicle and stepping back off the street in reality (left), and the virtual scenario from the point of view of the participant (right).

approximately a full minute to return to the starting point. The experiment in reality used a *BMW 320d* (BMW, Munich, Germany) and a *Mercedes-Benz E350d* (Daimler, Stuttgart, Germany), both with very similar overall dimensions and body type. The vehicles were equipped with digital cruise controls that allowed the driving speed to be set precisely. The simulation used two *BMW 3 series* vehicles with similar colors as the vehicles in the real environment since the respective Mercedes model was not available in the software's digital library. Due to their high overall similarity, the difference of brands should not be of any relevance to the experiment.

Since driving speeds displayed on vehicle dashboards are a few percentage points higher than the cars' actual speed, the cruise controls of the test vehicles were calibrated to the actual speed, which was determined extrinsically with a lidar speed detection device provided by the local police. This laser measurement unit, *Riegl FG21-P* (Riegl Laser Measurement Systems, Horn, Austria), is capable of measuring the distance and velocity of an approaching object: The velocity is deduced from several consecutive distance measurements over a measurement period of 0.4 s, while the distance measurement provides a precision of ± 0.1 m. The device was also used during the experiment to determine the exact position and velocity of the vehicle at the moment that the participant stepped back, crossing a photoelectric sensor line that triggered the speed gun.

D.4.4 Participants

One participant had to be retrospectively excluded since the corresponding data showed several outlier indicators and inconsistencies (see A. Field, 2013), suggesting that the participant had apparent difficulties in understanding and complying with the task at hand. The final sample consisted of 30 participants, who were selected at the Technical University of Munich, therefore representing a relatively homogeneous group with similar sociodemographic characteristics. The 19 male and 11 female participants were, on average, 21.1 y/o (*SD* = 2.1, [17 to 27 y/o]). The average body height was 178.9 cm (*SD* = 10.3, [161 to 198 cm]). Five of the participants who needed visual aids wore contact lenses during the study, and two wore glasses. The participants were not examined regarding their stereoscopic acuity.

All participants were asked to conduct the experiment in the real as well as in the virtual environment. The participants were counterbalanced, with half of them taking the experiment first in reality, whereas the other half started in the virtual environment, creating a between-subjects factor, hereinafter termed *exposure order*.

In the virtual part, the participants were first equipped with the motion-capture suit and the HMD, followed by a short calibration process. Subsequently, they familiarized themselves with the VR system while being loaded into a virtual replica of the actual laboratory in which they explored the virtual environment for a short period of time. When the participants signaled that they felt comfortable in the virtual environment, the actual experiment was loaded: The participants stood on a predefined spot on the edge of the virtual road in a perpendicular position to the street, ready to cross. The participants were told to observe the approaching vehicle coming from the right and step back when personally feeling that it is not safe to cross the street anymore. The vehicles were approaching at a constant speed of 30 km/h, 35 km/h, or 40 km/h, thus typical speeds for urban traffic. Participants were confronted with each velocity five times, thus fifteen encounters, in identical order that was pseudorandomized beforehand. After the vehicle had passed, the participant was asked to step back onto the street and to expect the next vehicle coming from the right.

In the real world, the experiment took place the same as in the virtual one. In rare situations, when uninvolved pedestrians or vehicles interfered with the scenario, the experiment was paused for a short period. The investigator, his assistants, and the drivers were communicating through radio sets, ensuring the safety of the experiment. The moment the participant stepped back (off the street), the velocity and distance of the test vehicle were registered with the lidar speed detection device, thus allowing a deduction of the TTC by dividing the spatial distance between the vehicle and the participant by the vehicle's velocity. In the virtual environment, the TTC was automatically registered the moment the participant stepped back, crossing an invisible virtual plane that was set at the same level as the photoelectric sensor line in reality.

Differences between the two environments were kept to a minimum, thus enabling a valid comparison of the variables of interest. It should be noted that the assessment of approaching velocities at 30 to 40 km/h will ultimately allow conclusions at this speed range only. VR designers should reflect upon which velocities may be of interest for future investigations and incorporate these in the validation study.

D.5 Results

D.5.1 Environment Comparison

The TTC values of approaching vehicles at which the participants decided that they would not cross the street anymore were analyzed using a 2×3

(*environment* × *velocity*) repeated-measures ANOVA with the associated effect size being calculated using Pearson's correlation coefficient *r* (see A. Field, 2013). The analysis revealed a significant main effect of environment, indicating significant differences between real and virtual environments regarding the chosen temporal distances: $F(1, 29) = 63.22$, $p < .001$, $r = .828$. In fact, the marginal mean TTC value in the virtual environment ($M = 2.06$ s, $SD = 0.93$) was 26% lower than the one in the real environment ($M = 2.80$ s, $SD = 1.03$). Furthermore, the interaction of environment × velocity turned out to be significant: $F(2, 58) = 9.96$, $p < .001$. Box plots in Figure D-4 show the distribution of accepted minimum TTCs for crossing the street, for each experimental environment and each velocity of the approaching vehicle. A post-hoc pairwise comparison indicated significantly shorter TTC values in

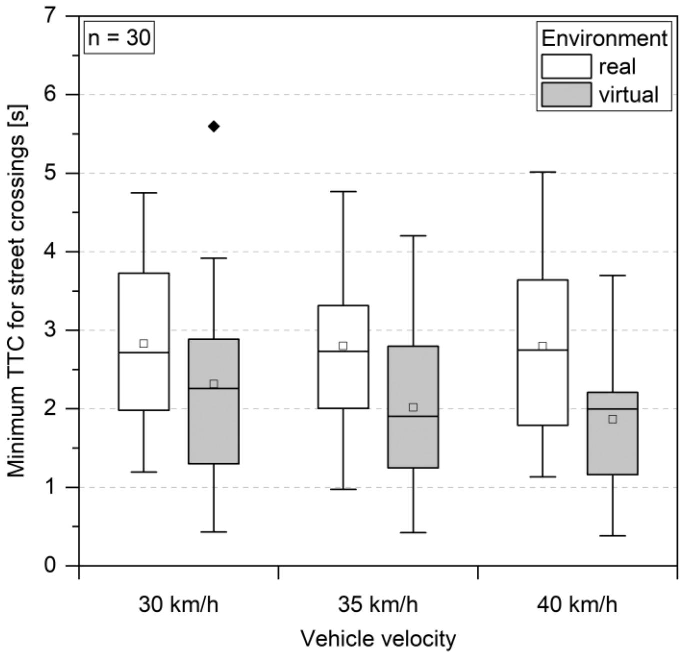


Figure D-4. Box-and-whisker plots show the distribution of the TTC values, which participants judged to be the threshold for safe crossings in the respective scenarios.
Box-and-whisker plot structure: Values between the lower and upper quartiles are represented by a box, while whiskers identify estimates within 1.5 times of the interquartile range (IQR) of the lower and upper quartiles. The horizontal line in the box shows the median, and the small square shows the arithmetic mean. Outliers are plotted with diamonds.

Table D-1. Pairwise comparison of the real and virtual environments regarding the TTC thresholds at which the participants (n = 30) judged the crossing unsafe, broken down into the three approaching velocities.

		Real TTC [s]	Virtual TTC [s]	Pairwise comparison
Approaching velocity	30 km/h	$M = 2.83$ $SD = 0.97$	$M = 2.32$ $SD = 1.09$	$p < .001$, $r = .237$
	35 km/h	$M = 2.80$ $SD = 1.03$	$M = 2.02$ $SD = 0.92$	$p < .001$, $r = .367$
	40 km/h	$M = 2.80$ $SD = 1.10$	$M = 1.87$ $SD = 0.82$	$p < .001$, $r = .427$

the virtual environment for all three vehicle velocities in comparison to the values obtained in the real environment (see Table D-1).

D.5.2 Vehicle Velocity

The ANOVA (see Section D.5.1) also revealed a significant main effect of vehicle velocity: $F(2, 58) = 14.67$, $p < .001$. Pairwise comparisons suggest that TTC values in the real environment remained unaffected by vehicle velocities, while the TTC values in the virtual environment differed significantly from each other, at least partially (see Table D-2).

Table D-2. Pairwise comparison of the different approaching velocities regarding the TTC thresholds at which the participants (n = 30) judged the crossing unsafe, broken down into the two experimental environments.

		Real environment	Virtual environment
Pairwise comparison of the TTC	30 km/h ⇌ 35 km/h	$p > .999$, $r = .013$	$p < .001$, $r = .144$
	35 km/h ⇌ 40 km/h	$p > .999$, $r = .002$	$p = .108$, $r = .086$
	30 km/h ⇌ 40 km/h	$p > .999$, $r = .015$	$p < .001$, $r = .224$

Furthermore, the effect of vehicle velocity on the relative discrepancy between the TTC values in the two environments was investigated with a simple regression analysis. The relative discrepancy was thereby computed as follows:

$$relative\ TTC\ discrepancy\ \Delta_{TTC} = \frac{|TTC_{real} - TTC_{virtual}|}{TTC_{real}}.$$

The simple regression analysis indicated a significant linear relationship between the environment discrepancy of the TTC values and the vehicle velocity: $F(1, 88) = 5.01$, $p = .028$, $R^2 = .054$, with a standardized regression coefficient of $\beta = .232$. The bootstrapped 95% confidence interval of the predictor coefficient, lying between 0.002 and 0.019, reveals a positive linear relation, thus indicating an increasing discrepancy between the TTC values in real and virtual environments for an increasing vehicle velocity. The predictive model suggests an increase of the TTC discrepancy by one percentage point for each additional km/h of the vehicle's approaching velocity:

$$relative\ TTC\ discrepancy\ \Delta_{TTC} = -0.06 + 0.01\ vehicle\ velocity\ \left[\frac{km}{h}\right].$$

D.5.3 Gender

The question of whether the participant's gender affected the outcome of the chosen TTC was investigated in a 2 × 3 × 2 (*environment* × *velocity* × *gender*) mixed-design ANOVA. The results suggest no significant main effect of gender: $F(1, 28) = 3.77$, $p = .062$, $r = .344$. The interaction between environment and gender was also not significant: $F(1, 28) = 0.10$, $p = .753$.

D.5.4 Exposure Order

The between-subjects factor that emerged from the participants starting the experiment either in the real or the virtual environment was investigated using a 2 × 3 × 2 (*environment* × *velocity* × *exposure order*) mixed-design ANOVA. No significant main effect of exposure order was observed: $F(1, 28) = 0.72$, $p = .404$, $r = .158$. The interaction between the environment and the exposure order was also not significant: $F(1, 28) = 1.54$, $p = .225$.

D.5.5 Body Height

The body height and thus the viewing height differed between participants in reality, while the virtual viewing height was not altered, resulting in an identical viewing height for all participants in the virtual environment. The relationship between the discrepancy of the participants' body heights between the real and the virtual environment and the discrepancy of TTC choices between the two environments was investigated with a simple regression analysis. The relative discrepancy of the body heights between the two environments was computed as follows:

$$relative\ height\ discrepancy\ \Delta_{height} = \frac{|height_{real} - height_{virtual}|}{height_{real}}.$$

The relative discrepancy of TTC choices between the two environments was computed as described in Section D.5.2. The simple regression analysis indicated no significant linear relationship between the discrepancy of body height and the discrepancy of TTC choices: $F(1, 28) = 0.43$, $p = .518$, $R^2 = .015$.

D.6 Discussion

D.6.1 Evaluation Method

The presented study introduces a novel type of road-crossing assessment method that contrasts with classical gap-acceptance studies in which pedestrians have to choose a gap between passing vehicles when crossing the road (e.g., Dommes et al., 2015; Kearney et al., 2006; Lobjois et al., 2013; O'Neal et al., 2018). Although participants do not perform an actual road crossing in the method presented here but express their rejection of crossing the street from a specific moment on, it seems plausible that the threshold between accepting and rejecting a road crossing can be measured with both approaches: crossing at the last acceptable moment or denying the crossing at the first unacceptable moment. Future studies that may investigate whether road-crossing decisions for *first-unacceptable* moments differ from those for *last-acceptable* moments are encouraged.

Yannis et al. (2013) observed average chosen gap sizes (n = 243) of 3.29 s (SD = 1.76) for real road crossings on a busy street. These values are quite comparable to the last-moment crossing decisions that were observed in the study presented here with an average TTC of 2.80 s (SD = 1.03) since the observed gap sizes by Yannis et al. (2013) would need to be reduced by the crossing start delay, which is on real streets, on average, 0.58 s (SD = 0.43), as reported by Feldstein (2019). Note that investigators who wish to compare results of gap-acceptance studies to those of studies using the step-back method must consider that gap-acceptance studies usually report the gap between two consecutive vehicles, while the step-back method determines the gap between the pedestrian and the approaching vehicle. For comparability, the determined gap between the two consecutive vehicles in the gap-acceptance study needs to be reduced by the start delay that the pedestrian cognitively requires before initiating the crossing after a vehicle that has just passed (see Figure D-5).

The step-back method has various advantages over classical gap-acceptance experiments: First of all, this method is safe in a real-life experiment since a collision is very unlikely to happen as opposed to when the street is actually crossed. Furthermore, the method is far easier to be implemented in a real environment since a gap-acceptance study would require a car convoy with all vehicles driving at the same precise speed while maintaining a precise time gap. Finally, the step-back method is presumably more time-efficient since participants directly choose their individually

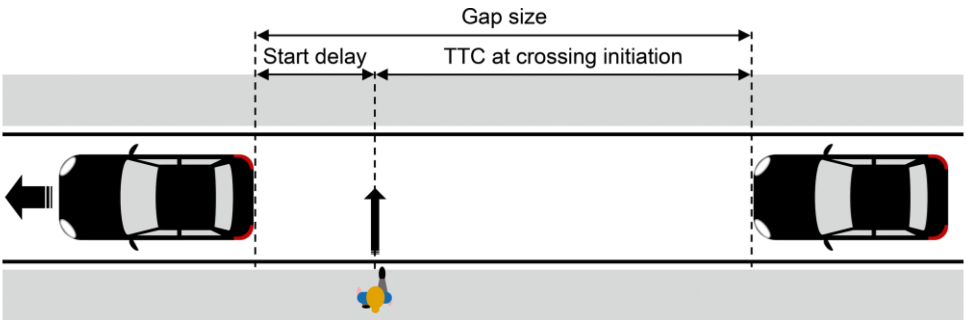


Figure D-5. Relation between gap size, start delay, and TTC at crossing initiation in a gap-acceptance study that investigates pedestrian road crossings.

preferred distance from one continuously approaching vehicle without the need to present a discrete number of vehicles that move with different gaps. Furthermore, researchers who apply the classical gap-acceptance method will have to carefully consider which gap increment they will use in their study. A tiny increment of the gaps will lengthen the experiment and ultimately cause participants to choose a smaller gap than under usual conditions: A lengthy waiting time at the roadside may result in frustration and trigger riskier behaviors, as shown in experiments conducted by Ashworth and Bottom (1977), Bottom and Ashworth (1978), and Ebbesen and Haney (1973). A large increment of the gaps, on the other hand, may corrupt the outcome since participants generally tend to accept a gap that represents a rounded-up value (and usually not a rounded-*down* one) of their individually preferred *minimum* gap size. Therefore, a large increment of the offered gaps will cause a larger discrepancy between the accepted gap size and the theoretically preferred one.

In the study presented here, it should be pointed out that the results need to be reflected upon with certain caution. Like most empirical studies, the experiment came with certain limitations: This includes, for example, that the virtual environment came with technological limitations, such as

- the virtual body height being constant,

- the experimental group being a relatively homogenous group, consisting of adults of fairly young age,
- approaching speeds of vehicles having relatively small variations, and
- the study being conducted on a single-lane one-way street, which produces a different psychological behavior than multi-lane two-way streets.

Some of these parameters and limitations are discussed in the following subsections.

D.6.2 Environment Discrepancies

Despite the massive efforts to assure equal conditions in both environments, the study showed some apparent differences in the road-crossing behavior between the real environment and the given virtual one. That said, these differences must not be overstated: Although 26% may seem large, the average absolute difference was a fraction of a second (M = 0.7 s, SD = 0.5). Furthermore, the experiment analyzed the extreme case: the *minimum* TTC to cross the street. While these shortcomings result in certain limitations for the VR system for experimental traffic investigations, they do not disqualify this system, or VR in general, from being a useful investigation tool. Although it shows that this VR system cannot entirely substitute a scenario in a real environment yet, such virtual environments can still offer a reasonable alternative to the laborious experiments in real environments when certain precautions are taken regarding the experimental design and the interpretation of the findings.

Although the evolution of VR is heading in the right direction, state-of-the-art VR components still have a long path of development ahead to close the gaps between real and virtual environments. For example, current HMDs contain a screen resolution that is still far from the threshold at which pixels and grid structures are not perceivable anymore, the FOV is nowhere near the human visual field of humans, and the screen is still not able to provide the contrast and true black as perceived in reality. To match the resolution capacity of the human fovea, a screen giving a 90° FOV—as commonly implemented in state-of-the-art HMDs—would require a screen resolution of up to 324 megapixels per eye (R. N. Clark, 2017), while current HMDs solely provide a resolution of about one to two megapixels per eye. This necessary screen resolution would have to be increased even further if the actual visual field of humans of up to 160° horizontally per eye (Spector, 1990) were to be implemented. Furthermore, the control center's calculation capacity when rendering the interactive virtual environment in real time has its limitations and calls for compromises regarding the implementation of sophisticated details in three-dimensional virtual space: An increase of the environment complexity will also increase the already perceptible system latency, which will negatively affect the user's perception and task performance. With the continuous progress made in VR technologies that involves a steady augmentation of display resolution, the latency issue will remain a critical challenge in the medium term (Feldstein & Ellis, 2021).

Although it is disputable whether these limitations related to screen resolution, screen contrast, field of view, and environmental details may or may not affect the perception of the approaching vehicle, they do influence the presence feeling in given virtual environment (see the meta-analysis by Cummings & Bailenson, 2016). This may lower the experienced threat of oncoming vehicles in the virtual environment and thus lead to an acceptance of shorter TTCs in comparison to real scenarios. This phenomenon was also observable in the study by Schwebel et al. in 2008 (as reported in the extended analysis by Feldstein, 2019). Other, although probably minor issues for a realistic three-dimensional illusion and depth perception are missing optical aberrations with two-dimensional displays: A realistic curvilinear perspective solely occurs when looking straight, and the fixed focal plane in HMDs prevents common effects of defocus blur (Mather, 1996).

Researchers must be aware that the determined similarity of participants' behaviors in a simulator compared to behavior in reality may indicate what type of measurement scale may be applicable when transferring results from virtual environments to real ones. The different measurement scales—nominal, ordinal, interval, and ratio—allow different basic empirical operations (S. S. Stevens, 1946). Accordingly, future research questions and experimental designs would need to be adapted to the applicable measurement scales in the specific setup. Even if the validation process shows that the simulator does not induce an identical perception and behavior in virtual environments such as in real ones, the setup may nonetheless serve as a useful tool by simply downscaling the research questions. For example, the simulator can be used for comparing two different virtual scenarios regarding the identification of simple equalities or inequalities (nominal scale) or for determining whether the difference is negative or positive (ordinal scale) without concluding the precise numerical discrepancy. In summary, a higher similarity between reality and the simulator's virtual environment will allow the projection of more information from the experimental findings onto reality.

D.6.3 Human Aspects

Several studies carried out in the past have revealed differences between male and female participants regarding the TTC estimation of approaching objects (Caird & Hancock, 1994; Manser & Hancock, 1996; McLeod & Ross, 1983; Schiff & Oldak, 1990), although Feldstein (2019) argued that those differences might be linked to correlating factors, such as driving experience (see Cavallo & Laurent, 1988), or depending on additional factors, such as age (see Hancock & Manser, 1997; Lobjois & Cavallo, 2007; Schiff et al., 1992). Just like Recarte et al. (2005), the study presented here did not observe a significant difference between male and female participants. However, it has to be noted that the compared group sizes of 19 male participants versus 11 female participants are somewhat small, and also that a certain tendency toward a significant difference may be argued given the determined *p*-value ($p = .062$), which is close to the set significance threshold ($p < .05$). A larger sample, or for example, an older one, may potentially result in a significant difference between male and female participants.

The participants had an average body height of 179 cm ($SD = 10.3$), varying between 161 cm and 198 cm. Thus, according to statistical data by Jürgens (1999), the participants had an average eye height of 167 cm. While the participants' individual eye heights remained "as is" in the real environment, the same avatar with an eye height of 180 cm was used for all participants in the virtual environment. A study by Leyrer et al. (2011) showed that a wrong virtual eye height as compared to the participant's real one might lead to decreased accuracy when judging distances. Similar observations were made when the distance perception was investigated in the simulator presented here in another study (Feldstein et al., 2020), suggesting that a higher deviation of the real eye height from the virtual one increases the lack of congruity of distance estimates between real and virtual environments. Surprisingly, the results of the experiment presented here suggest that the discrepancy between the real and the virtual environment when judging approaching vehicles was unaffected by the discrepancy between the real eye height and the virtual one. This can perhaps be explained by the nature of the task since a varying standing height at the roadside (elicited by varying sidewalk heights) is common for pedestrians in everyday situations. Thus, the raised virtual eye height implemented here does not seem to confuse the participants' perceptual processes.

Another anatomical variation among participants is the IPD that allows humans to extract binocular depth cues from the perceived environment. The virtually set IPD in the HMD, however, remained invariable in the presented experiment. Best (1996) demonstrated that a wrongly set IPD did not affect the size estimation of two-dimensional objects, even when the IPD was mismatched to the inner or outer extreme. His observations solely allowed for the conclusion that an extremely mismatched IPD resulted in greater fatigue of the user, but that a fixed IPD set at the general human average may just do the trick. That said, one has to be aware that a mismatch between the user's actual IPD and the one set in the virtual environment may eventually cause eyestrain, vision blur, headaches, and—in extreme cases—double vision (Peli, 1999).

Investigators may consider investing some additional time during the initial configuration and adjust the virtual-environment system individually to the participant's eye height and IPD. Thereby, as a rule of thumb, the participant's eye height corresponds fairly precisely to 93.2% (SD = 0.47) of the body height (Feldstein et al., 2020). There are several user-friendly smartphone applications available for the measurement of the IPD that may be used in this context (e.g., Pundlik et al., 2019).

D.6.4 Vehicle Velocity

The analysis revealed an impact of the vehicle velocity on the discrepancy between the real and the virtual environment regarding the TTC threshold at which the participant considered the crossing not to be safe anymore. This encouraged a deeper look into this phenomenon. Figure D-6 and Figure D-7 show the average gaps between the participants and the approaching vehicle at which the participants set their personal minimum safety distance for crossing the road. These gaps are plotted in terms of temporal distance and spatial distance, respectively.

The graphs suggest that participants were choosing the minimum vehicle distance for road crossings in the real world in a constant-temporal-distance behavior, whereas their decisions in the virtual environment seem to follow a constant-spatial-distance behavior. It is conclusive that the temporal distance of the approaching vehicle remains constant for crossing

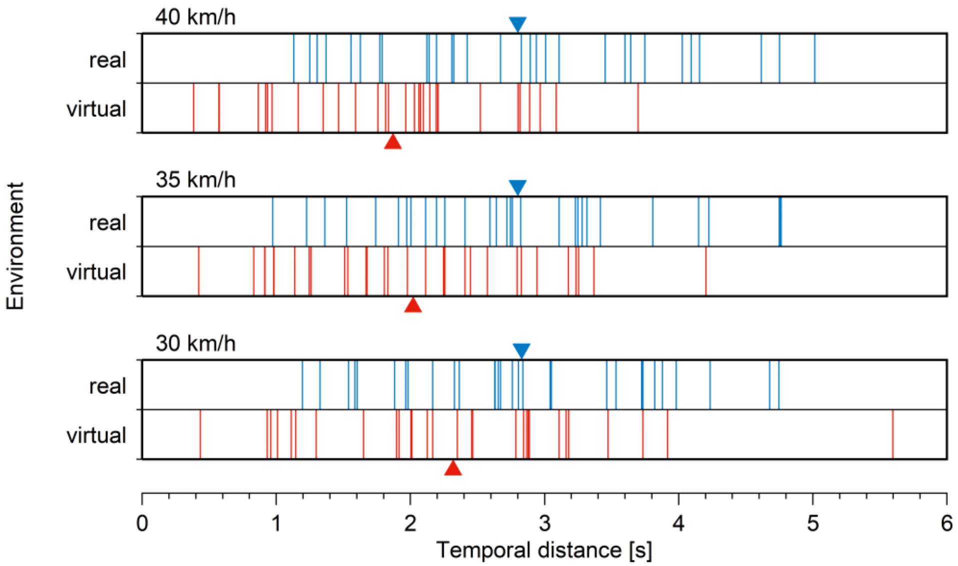


Figure D-6. Minimum distance to the approaching vehicle at which the participant was still willing to cross the road, plotted as temporal distance (i.e., TTC): Each bar represents one of the thirty participants in the respective environment-velocity condition. The triangles represent the mean values of all participants for the respective condition.

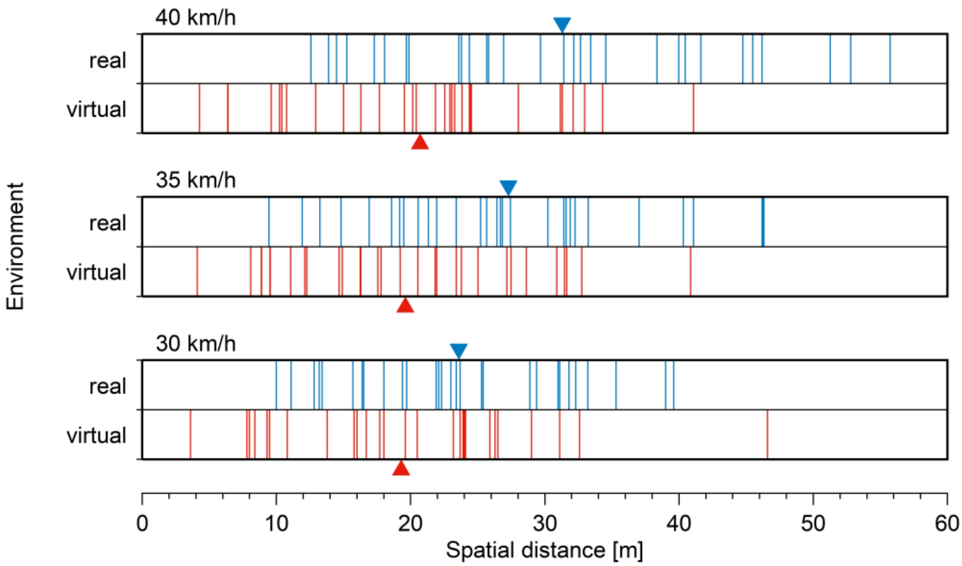


Figure D-7. Minimum distance to the approaching vehicle at which the participant was still willing to cross the road, plotted as spatial distance: Each bar represents one of the thirty participants in the respective environment-velocity condition. The triangles represent the mean values of all participants for the respective condition.

decisions, such as observed in the real environment; after all, crossing requires a certain *time*. In the virtual environment, participants apparently temporal distance of an approaching vehicle, basing their decisions on heuristics such as the vehicle's spatial distance. This inadequacy could be related to the low resolution of the HMD screen that hampered observation of the vehicle and its rate of approach, especially at a far distance. The pairwise comparison of the TTC values for the different approaching velocities (see Section D.5.2) confirmed that there were no significant differences in the real environment, while the TTC values for the different approaching velocities in the virtual environment differed significantly or showed at least such tendencies.

Analogous to the ANOVA in Section D.5.2, a 2 × 3 (*environment* × *vehicle velocity*) repeated-measures ANOVA was conducted, but this time the spatial distance instead of the TTC was investigated. The main effect of environment turned out to be significant: $F(1, 29) = 66.51$, $p < .001$, $r = .834$. The pairwise comparison between the real and virtual environments indicated significant differences for each velocity (see Table D-3). The main effect of vehicle velocity also turned out to be significant: $F(2, 58) = 39.54$, $p < .001$. Pairwise comparison indicated the differences between the spatial distances for the different approaching velocities in the real environment to

Table D-3. Pairwise comparison of the real and virtual environments regarding the spatial-distance thresholds at which the participants (n = 30) judged the crossing unsafe, broken down into the three approaching velocities.

		Real TTC [s]	Virtual TTC [s]	Pairwise comparison
Approaching velocity	30 km/h	$M = 23.6$ $SD = 8.1$	$M = 19.3$ $SD = 9.1$	$p < .001$, $r = .237$
	35 km/h	$M = 27.2$ $SD = 10.0$	$M = 19.6$ $SD = 8.9$	$p < .001$, $r = .367$
	40 km/h	$M = 31.1$ $SD = 12.2$	$M = 20.7$ $SD = 9.1$	$p < .001$, $r = .427$

be significant (see Table D-4). The spatial distances in the virtual environment, on the other hand, showed high similarities: Spatial distances of approaching vehicles with a velocity difference of 5 km/h did not differ significantly from each other. The comparison of vehicles with a velocity difference of 10 km/h suggests a significant difference in the spatial distances, although the *p*-value (p = .046) was barely below the set significance threshold (p < .05). This indicates that participants based their crossing decisions in the virtual environment predominantly—but not exclusively—on the spatial distance. Given a sufficiently high velocity difference, participants will presumably perceive and also incorporate this information into their crossing decisions in the virtual environment.

As discussed in Section D.3.4.2, several experiments carried out in the past observed that TTC estimates were disproportionately affected by the vehicles' approaching velocities (Caird & Hancock, 1994; Kappé & Korteling, 1995; Petzoldt, 2014). However, these experiments were all carried out in virtual environments. The experiment presented here suggests that the difficulties in judging the TTC at varying approaching velocities may be due to technical deficiencies of the experimental setting and may not necessarily apply in a real environment. This assumption may be further supported by Cavallo's and Laurent's (1988) experiment on a real street, who did not

Table D-4. Pairwise comparison of the different approaching velocities regarding the spatial-distance thresholds at which the participants (n = 30) judged the crossing unsafe, broken down into the two experimental environments.

		Real environment	Virtual environment
Pairwise comparison of the spatial distance	30 km/h ⇌ 35 km/h	p < .001, r = .194	p > .999, r = .018
	35 km/h ⇌ 40 km/h	p < .001, r = .166	p = .363, r = .061
	30 km/h ⇌ 40 km/h	p < .001, r = .336	p = .046, r = .077

observe any effect of velocity on TTC estimates. Moore (1953), who observed pedestrians on a real street, also stated that their crossing decisions depended primarily on the time gap to the approaching vehicle and not on the spatial distance.

It should be noted that the results in the experiment presented here were collected with a group of fairly young adults. Oxley et al. (2005) observed in a *virtual* environment for *all* age groups that gap choices for crossing the street were primarily based on vehicles' spatial distances and less on their temporal distances. As for a *real* environment, however, findings by Staplin (1995) suggest that while younger adults incorporate the vehicle's velocity into their crossing decisions, elderly participants seemed unable to judge the vehicle's velocity and hence chose the crossing moment primarily based on spatial distances. Similar observations were also reported by Hills (1980).

It has to be considered that the investigated vehicle velocities were relatively low. A lower velocity implies a longer observation time of the vehicle's approach and allows the vehicle to approach closer to the pedestrian. Some studies suggest that a longer observation time supports participants' judgments of the TTC (Andersen & Enriquez, 2006; Groeger & Cavallo, 1991), which again will afford safer crossing decisions. Since higher velocities will produce crossing decisions at larger distances, an accurate TTC judgment will become progressively more challenging with increasing velocity, as has been shown in several studies (Caird & Hancock, 1994; McLeod & Ross, 1983; Thomson, 1983). The vehicle's looming rate is relatively low at large distances, which challenges the pedestrian's ability to judge the visual expansion rate accurately, especially on screens with insufficient resolution but also in real life. Consequentially, pedestrians will not incorporate the approaching velocity into their crossing decisions anymore, which explains the increasing number of unsafe crossing decisions for high velocities, such as reported, for example, by Dommes and Cavallo

(2011). The observations by Parsonson et al. (1996) on a real street also support the assumption that high velocities may not be judged accurately, ultimately leading to crossing decisions of constant spatial distance.

In conclusion, this means that an older group of participants would have potentially displayed similar crossing decision strategies in both environments, while young participants revealed an essential difference between the two environments regarding the crossing decision strategy. Furthermore, a higher velocity of approaching vehicles could also have triggered a similar crossing strategy in both environments. This demonstrates the difficulty of conducting proper validation studies and emphasizes the attention that must be given to the composition of the participant group (e.g., age) and the parameters of the experimental design (e.g., vehicle velocity).

D.6.5 Exposure Order

Feldstein et al. (2020), who investigated the participants' ability to estimate distances in the VR system presented here, observed that the exposure order of the environments affected the distance estimates. In the study presented here, it did not matter whether the participants started the experiment in the real environment or the virtual one: No significant main effect of the exposure order or significant interaction effect of exposure order and environment was observed. That said, the exposure order may be irrelevant *at this stage* since the differences between the two environments turned out to be significant. When the overall differences between the real and the virtual environment regarding the TTC choices will reach a nonsignificant level, an effect of the environment exposure order may potentially occur, similar to the observation in the study by Feldstein et al. (2020).

D.6.6 Conclusion

When using a VR system for empirical data collection of human behavior, investigators shall consider to which degree the system is validated and thus which experimental designs and research questions can be generally applied when using the system. The experiment presented here showed some significant differences between real and virtual environments in terms of human perception and behavior, underlining the need for such validation studies. Even highly immersive, state-of-the-art VR systems, such as the one introduced here, may bear certain insufficiencies and employ differing cognitive mechanisms: The results of the validation study suggest that participants judged the safety to cross the street in the virtual environment primarily based on the spatial distance of the approaching vehicle, while their decisions in the real environment were based on the vehicle's temporal distance. Not only the road-crossing strategies differed between the two environments, but participants also chose significantly lower temporal and spatial distances to the approaching vehicle in the virtual environment than in the real one. This stimulates further research in this field in order to understand where these differences come from and how the behavioral gap between real and virtual environments can be reduced, also with a view toward necessary enhancements of the technical components.

These findings shall not invalidate findings that have been acquired with virtual environments in general but shall sensitize researchers when interpreting these and emphasize the validation necessity of systems that substitute a real environment. The need for virtual environments for behavioral research, training purposes, or therapies will persist, given their tremendous advantages relating to the users' safety, the scenario's reproducibility, and the efficient cost- and time-reduction.

REFERENCES

Adelstein, B. D., Johnston, E. R., & Ellis, S. R. (1996). Dynamic response of electromagnetic spatial displacement trackers. *Presence: Teleoperators & Virtual Environments*, *5*(3), 302–318.

Adelstein, B. D., Jung, J. Y., & Ellis, S. R. (2001). *Predictive compensator optimization for head tracking lag in virtual environments* [Technical Report NASA/TM-2001209627]. NASA Ames Research Center.

Adelstein, B. D., Lee, T. G., & Ellis, S. R. (2003). Head tracking latency in virtual environments: Psychophysics and a model. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(20), 2083–2087.

Adler, D. (2016). *Rolling shutter versus global shutter* [Online article]. Retrieved November 21st, 2018, from http://bhphotovideo.com/explora/video/tips-and-solutions/rolling-shutter-versus-global-shutter/

Albert, R., Patney, A., Luebke, D., & Kim, J. (2017). Latency requirements for foveated rendering in virtual reality. *ACM Transactions on Applied Perception (TAP)*, *14*(4), Article 25.

Allerton, D. (2009). *Principles of flight simulation*. John Wiley & Sons.

Allison, R. S., Harris, L. R., Jenkin, M., Jasiobedzka, U., & Zacher, J. E. (2001). Tolerance of temporal delay in virtual environments. In H. Takemura & K. Kiyokawa (Eds.), *Proceedings of the IEEE Virtual Reality Conference 2001* (pp. 247–253). IEEE.

Andersen, G. J., & Enriquez, A. (2006). Aging and the detection of observer and moving object collisions. *Psychology and Aging*, *21*(1), 74–85.

Anstis, S. (2003). Moving objects appear to slow down at low contrasts. *Neural Networks*, *16*(5/6), 933–938.

Ashworth, R., & Bottom, C. G. (1977). Some observations of driver gap-acceptance behaviour at a priority intersection. *Traffic Engineering and Control*, *18*(12), 569–571.

Azuma, R., & Bishop, G. (1995). A frequency-domain analysis of head-motion prediction. In S. G. Mair & R. Cook (Eds.), *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)* (pp. 401–408). ACM.

Bailey, R. E., Arthur, J. J., III. & Williams, S. P. (2004). Latency requirements for head-worn display S/EVS applications. In J. G. Verly (Ed.), *Enhanced and synthetic vision III* (pp. 98–109). SPIE.

Banister, H., & Blackburn, J. M. (1931). An Eye Factor Affecting proficiency at ball games. *British Journal of Psychology: General Section*, *21*(4), 382–384.

Barton, B. K., & Schwebel, D. C. (2007). The roles of age, gender, inhibitory control, and parental supervision in children's pedestrian safety. *Journal of Pediatric Psychology*, *32*(5), 517–526.

Batt, S. (2018). *What is "VSync," and should I turn it on or off?* [Blog post]. Retrieved January 22nd, 2019, from http://maketecheasier.com/what-is-vsync/

Bedikian, R. (2013). *Understanding latency: Part 2* [Blog post]. Retrieved January 21st, 2019, from http://blog.leapmotion.com/understanding-latency-part-2/

Begault, D. R. (2000). *3-D sound for virtual reality and multimedia*. NASA Center for AeroSpace Information.

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine*, *6*(4), 6–22.

Bergström, I., Azevedo, S., Papiotis, P., Saldanha, N., & Slater, M. (2017). The plausibility of a string quartet performance in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, *23*(4), 1352–1359.

Bertolini, G., & Straumann, D. (2016). Moving in a moving world: A review on vestibular motion sickness. *Frontiers in Neurology*, *7*, Article 14.

Best, S. (1996). Perceptual and oculomotor implications of interpupillary distance settings on a head-mounted virtual display. In *Proceedings of the IEEE 1996 National Aerospace and Electronics Conference (NAECON '96)* (pp. 429–434). IEEE.

Bhagavathula, R., Williams, B., Owens, J., & Gibbons, R. (2018). The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 2056–2060.

Billington, J., Wilkie, R. M., Field, D. T., & Wann, J. P. (2010). Neural processing of imminent collision in humans. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1711), 1476–1481.

Blakemore, M. R., & Snowden, R. J. (1999). The effect of contrast upon perceived speed: A general phenomenon? *Perception*, *28*(1), 33–48.

Blakemore, M. R., & Snowden, R. J. (2000). Textured backgrounds alter perceived speed. *Vision Research*, *40*(6), 629–638.

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. The MIT Press.

Blissing, B., Bruzelius, F., & Eriksson, O. (2016). Effects of visual latency on vehicle driving behavior. *ACM Transactions on Applied Perception (TAP)*, *14*(1), Article 5.

Boff, K. R., & Lincoln, J. E. (Eds.). (1988). *Engineering data compendium: Human perception and performance (Vol. 1)*. Harry G. Armstrong Aerospace Medical Research Laboratory.

Bootsma, R. J., & Oudejans, R. R. (1993). Visual information about time-to-collision between two objects. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(5), 1041–1052.

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. Apple-Century-Crofts.

Bottom, C. G., & Ashworth, R. (1978). Factors affecting the variability of driver gap-acceptance behaviour. *Ergonomics*, *21*(9), 721–734.

Boxer, B. (2017). *What is your GPU doing when VSync is running in your PC games?* [Blog post]. Retrieved January 23rd, 2019, from http://blog.parsecgaming.com/vsync-technology-impact-on-pc-gaming-and-your-gpu-6307fee70d29/

Braunstein, M. L. (1976). *Depth perception through motion*. Academic Press.

Brimijoin, W. O., Boyd, A. W., & Akeroyd, M. A. (2013). The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE, 8*(12), Article e83068.

Brungart, D. S., Simpson, B. D., & Kordik, A. J. (2005). The detectability of headtracker latency in virtual audio displays. In *Proceedings of the 11th International Conference on Auditory Display (ICAD '05)* (pp. 37–42). Limerick, Ireland.

Byington, K. W., & Schwebel, D. C. (2013). Effects of mobile Internet use on college student pedestrian injury risk. *Accident Analysis & Prevention*, *51*, 78–83.

Caird, J. K., & Hancock, P. A. (1994). The perception of arrival time for different oncoming vehicles at an intersection. *Ecological Psychology*, *6*(2), 83–109.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum.

Carel, W. L. (1961). *Visual factors in the contact analog* [Technical Report 263 820]. General Electric at Cornell University Industry Research Park.

Carmack, J. (2013). *Latency mitigation strategies* [Blog post]. Retrieved November 8th, 2018, from https://danluu.com/latency-mitigation

Cavallo, V., & Cohen, A. (2001). Perception. In P.E. Barjonet (Ed.), *Traffic Psychology Today* (pp. 63–89). Springer.

Cavallo, V., Dang, N.-T., Pala, P., Granié, M.-A., Schneider, S., Maruhn, P., & Bengler, K. (2019). Comparison of HMD and CAVE pedestrian simulators [Poster]. At *2019 Road Safety and Simulation Conference (RSS '19)*. Iowa City, IA, United States.

Cavallo, V., Dommes, A., Dang, N. T., & Vienne, F. (2019). A street-crossing simulator for studying and training pedestrians. *Transportation Research Part F: Traffic Psychology and Behaviour, 61*, 217–228.

Cavallo, V., & Laurent, M. (1988). Visual information and skill level in time-to-collision estimation. *Perception, 17*(5), 623–632.

Cavanagh, P., & Leclerc, Y. G. (1989). Shape from shadows. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 3–27.

Clark, R. N. (2017). Notes on the resolution and other details of the human eye [Blog post]. Retrieved January 15th, 2019, from http://clarkvision.com/articles/eye-resolution.html

Clark, N. (2006). The Airbus saga: Crossed wires and a multi-billion euro delay [Newspaper article]. Retrieved March 22nd, 2020, from https://www.nytimes.com/2006/12/11/business/worldbusiness/11iht-airbus.3860198.html

Connelly, M. L., Conaglen, H. M., Parsonson, B. S., & Isler, R. B. (1998). Child pedestrians' crossing gap thresholds. *Accident Analysis & Prevention*, *30*(4), 443–453.

Coull, J. T., Vidal, F., Goulon, C., Nazarian, B., & Craig, C. (2008). Using time-to-contact information to assess potential collision modulates both visual and temporal prediction networks. *Frontiers in Human Neuroscience*, *2*, Article 10.

Cummings, J. J., & Bailenson, J. N. (2014). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272–309

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, *292*(4), 497–523.

Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*(2), 198–216.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 69–117). Academic Press.

Dahm, M. (2005). *Grundlagen der Mensch-Computer-Interaktion*. Pearson.

Davis, S., Nesbitt, K., & Nalivaiko, E. (2014). A systematic review of cybersickness. In K. Blackmore, K. Nesbitt, & S. P. Smith (Eds.), *Proceedings of the 2014 Conference on Interactive Entertainment (IE '14)* (Article 8). ACM.

de Bellis, E., Schulte-Mecklenbeck, M., Brucks, W., Herrmann, A., & Hertwig, R. (2018). Blind haste: As light decreases, speeding increases. *PLoS ONE*, *13*(1), Article e0188951.

de Kort, Y. A. W., IJsselsteijn, W. A., Kooijman, J., & Schuurmans, Y. (2003). Virtual laboratories: Comparability of real and virtual environments for environmental psychology. *Presence: Teleoperators & Virtual Environments*, *12*(4), 360–373.

DeLucia, P. R. (1991). Pictorial and motion-based information for depth perception. *Journal of Experimental Psychology: Human Perception and Performance, 17(3),* 738–748.

DeLucia, P. R. (1999). Size-arrival effects: The potential roles of conflicts between monocular and binocular time-to-contact information, and of computer aliasing. *Perception & Psychophysics*, *61*(6), 1168–1177.

DeLucia, P. R. (2004). Multiple sources of information influence time-to-contact judgments: Do heuristics accommodate limits in sensory and cognitive processes? In H. Hecht & G. J. P. Savelsbergh (Eds.), *Advances in psychology, vol. 135: Time-to-contact* (pp. 243–286). Elsevier.

DeLucia, P. R. (2005). Does binocular disparity or familiar size information override effects of relative size on judgements of time to contact? *The Quarterly Journal of Experimental Psychology: Section A*, *58*(5), 865–886.

DeLucia, P. R., Kaiser, M. K., Bush, J. M., Meyer, L. E., & Sweet, B. T. (2003). Information integration in judgements of time to contact. *The Quarterly Journal of Experimental Psychology: Section A*, *56*(7), 1165–1189.

DeLucia, P. R., Preddy, D., & Oberfeld, D. (2016). Audiovisual integration of time-to-contact information for approaching objects. *Multisensory Research*, *29*(4/5), 365–395.

DeLucia, P. R., & Warren, R. (1994). Pictorial and motion-based depth information during active control of self-motion: Size-arrival effects on collision avoidance. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 783–798.

Demetre, J. D., Lee, D. N., Grieve, R., Pitcairn, T. K., Ampofo-Boateng, K., & Thomson, J. A. (1993). Young children's learning on road-crossing simulations. *British Journal of Educational Psychology*, *63*(2), 349–359.

Demetre, J. D., Lee, D. N., Pitcairn, T. K., Grieve, R., Thomson, J. A., & Ampofo-Boateng, K. (1992). Errors in young children's decisions about traffic gaps: Experiments with roadside simulations. *British Journal of Psychology*, *83*(2), 189–202.

Di Luca, M. (2010). New method to measure end-to-end delay of virtual reality. *Presence: Teleoperators & Virtual Environments*, *19*(6), 569–584

Dinh, H. Q., Walker, N., Hodges, L. F., Song, C., & Kobayashi, A. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In L. Rosenblum, P. Astheimer, & D. Teichmann (Eds.), *Proceedings of the IEEE Virtual Reality Conference 1999 (VR '99)* (pp. 222–228). IEEE.

DiZio, P., & Lackner, J. R. (1992). Spatial orientation, adaptation, and motion sickness in real and virtual environments. *Presence: Teleoperators & Virtual Environments*, *1*(3), 319–328.

Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. In M. T. Bolas, A. J. Woods, J. O. Merritt, & S. A. Benton (Eds.), *Stereoscopic displays and virtual reality systems XI* (pp. 36–47). SPIE.

Dommes, A., & Cavallo, V. (2011). The role of perceptual, cognitive, and motor abilities in street-crossing decisions of young and older pedestrians. *Ophthalmic and Physiological Optics*, *31*(3), 292–301.

Dommes, A., & Cavallo, V. (2012). Can simulator-based training improve street-crossing safety for elderly pedestrians? *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(2), 206–218.

Dommes, A., Cavallo, V., Dubuisson, J. B., Tournier, I., & Vienne, F. (2014). Crossing a two-way street: comparison of young and old pedestrians. *Journal of Safety Research*, *50*, 27–34.

Dommes, A., Cavallo, V., & Oxley, J. (2013). Functional declines as predictors of risky street-crossing decisions in older pedestrians. *Accident Analysis & Prevention*, *59*, 135–143.

Dommes, A., Cavallo, V., Vienne, F., & Aillerie, I. (2012). Age-related differences in street-crossing safety before and after training of older pedestrians. *Accident Analysis & Prevention*, *44*(1), 42–47.

Dommes, A., Le Lay, T., Vienne, F., Dang, N. T., Beaudoin, A. P., & Do, M. C. (2015). Towards an explanation of age-related difficulties in crossing a two-way street. *Accident Analysis & Prevention*, *85*, 229–238.

Draper, M. H., Viirre, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors*, *43*(1), 129–146.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., & Wenzel, E. M. (1992). On the externalization of auditory images. *Presence: Teleoperators & Virtual Environments*, *1*(2), 251–257.

Ebbesen, E. B., & Haney, M. (1973). Flirting with death: Variables affecting risk taking at intersections. *Journal of Applied Social Psychology*, *3*(4), 303–324.

Ellis, S. R. (1994). What are virtual environments? *IEEE Computer Graphics and Applications*, *14*(1), 17–22.

Ellis, S. R. (1996). Presence of mind: A reaction to Thomas Sheridan's "Further musings on the psychophysics of presence." *Presence: Teleoperators & Virtual Environments*, *5*(2), 247–259.

Ellis, S. R. (2000). Collision in space. *Ergonomics in Design*, *8*(1), 4–9.

Ellis, S. R., Bréant, F., Menges, B. M., Jacoby, R. H., & Adelstein, B. D. (1997). Operator interaction with virtual objects: Effect of system latency. In *Proceedings of the Seventh International Conference on Human-Computer Interaction (HCI '97)* (pp. 973–976). San Francisco, CA, United States.

Ellis, S. R., Mania, K., Adelstein, B. D., & Hill, M. I. (2004). Generalizeability of latency detection in a variety of virtual environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(23), 2632–2636.

Ellis, S. R., Wolfram, A., & Adelstein, B. D. (2002). Three dimensional tracking in augmented environments: User performance trade-offs between system latency and update rate. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(26), 2149–2153.

Ellis, S. R., Young, M. J., Adelstein, B. D., & Ehrlich, S. M. (1999). Discrimination of changes of latency during voluntary hand movement of virtual objects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(22), 1182–1186.

Evans, L. (2004). *Traffic safety*. Science Serving Society.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf & Härtel.

Feldstein, I. T. (2019). Impending collision judgment from an egocentric perspective in real and virtual environments: A review. *Perception*, *48*(9), 769–795.

Feldstein, I. T., Dietrich, A., Milinkovic, S., & Bengler, K. (2016). A pedestrian simulator for urban crossing scenarios. *IFAC-PapersOnLine*, *49*(19), 239–244.

Feldstein, I. T., & Dyszak, G. N. (2020). Road crossing decisions in real and virtual environments: A comparative study on simulator validity. *Accident Analysis & Prevention*, *137*, Article 105356.

Feldstein, I. T., & Ellis, S. R. (2021). A simple video-based technique for measuring latency in virtual reality or teleoperation. *IEEE Transactions on Visualization and Computer Graphics*, *27*(9), 3611–3625.

Feldstein, I. T., Güntner, A., & Bengler, K. (2015). Infrared-based in-vehicle head tracking: A prototype for tracking drivers' head movements in real time. *Procedia Manufacturing*, *3*, 829–836.

Feldstein, I. T., Kölsch, F. M., & Konrad, R. (2020). Egocentric distance perception: A comparative study investigating differences between real and virtual environments. *Perception*, *49*(9), 940–967.

Feldstein, I. T., Lehsing, C., Dietrich, A., & Bengler, K. (2018). Pedestrian simulators for traffic research: State of the art and future of a motion lab. *International Journal of Human Factors Modelling and Simulations*, *6*(4), 250–265.

Feldstein, I. T., & Peli, E. (2020). Pedestrians accept shorter distances to light vehicles than to dark ones when crossing the street. *Perception, 49*(5), 558–566.

Fellgett, P. B. (1974). Ambisonic reproduction of directionality in surround-sound systems. *Nature*, *252*(5484), 534–538.

Ferris, S. H. (1972). Motion parallax and absolute distance. *Journal of Experimental Psychology*, *95*(2), 258–263.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4th edition)*. SAGE Publications.

Field, D. T., & Wann, J. P. (2005). Perceiving time to collision activates the sensorimotor cortex. *Current Biology*, *15*(5), 453–458.

Fink, P. W., Foo, P. S., & Warren, W. H. (2007). Obstacle avoidance during walking in real and virtual environments. *ACM Transactions on Applied Perception (TAP)*, *4*(1), Article 2.

Fisher, S. K., & Ciuffreda, K. J. (1988). Accommodation and apparent distance. *Perception*, *17*(5), 609–621.

Foley, J. M. (1991). Stereoscopic distance perception. In S. R. Ellis, M. K. Kaiser, & A. J. Grunwald (Eds.), *Pictorial communication in virtual and real environments* (pp. 559–566). Taylor & Francis.

Frank, L. H., Casali, J. G., & Wierwille, W. W. (1988). Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator. *Human Factors*, *30*(2), 201–217.

Friston, S., & Steed, A. (2014). Measuring latency in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, *20*(4), 616–625.

Gegenfurtner, K. R., Mayser, H., & Sharpe, L. T. (1999). Seeing movement in the dark. *Nature*, *398*(6727), 475–476.

Geri, G. A., Gray, R., & Grutzmacher, R. (2010). Simulating time-to-contact when both target and observer are in motion. *Displays*, *31*(2), 59–66.

Gerzon, M. A. (1977). Multi-system ambisonic. *Wireless World*, *83*, 43–47.

Gibson, J. J. (1950). *The perception of the visual world*. Houghton Mifflin.

González, E. G., Allison, R. S., Ono, H., & Vinnikov, M. (2010). Cue conflict between disparity change and looming in the perception of motion in depth. *Vision Research*, *50*(2), 136–143.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Google Maps. (2017). Technische Universität München, Fakultät für Maschinenwesen. Retrieved January 7th, 2017, from http://google.de/maps/@48.2650456,11.6680379,387m/data=!3m1!1e3/

Gray, R., & Regan, T. (1998). Accuracy of estimating time to collision using binocular and monocular information. *Vision Research*, *38*(4), 499–512.

Grechkin, T. Y., Nguyen, T. D., Plumert, J. M., Cremer, J. F., Kearney, J. K. (2010). How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions of Applied Perception (TAP)*, 7(4), Article 26.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley & Sons.

Groeger, J. A., & Cavallo, V. (1991). Judgements of time-to-collision and time-to-coincidence. In A. G. Gale, I. D. Brown, C. M. Haslegrave, I. Moorhead, & S. Taylor (Eds.), *Vision in vehicles III* (pp. 27–34). Elsevier.

Gröhn, M., Lokki, T., & Takala, T. (2007). Localizing sound sources in a cave-like virtual environment with loudspeaker array reproduction. *Presence: Teleoperators & Virtual Environments*, *16*(2), 157–171.

Guastavino, C., Larcher, V., Catusseau, G., & Boussard, P. (2007). Spatial audio quality evaluation: Comparing transaural, Ambisonics and stereo. In *Proceedings of the 13th International Conference on Auditory Display (ICAD '07)* (pp. 53–59). Montréal, QC, Canada.

Guzy, L. T., Leibowitz, H. W., & Scialfa, C. T. (1991). A note: Can vehicle speed be estimated accurately? *Journal of Applied Social Psychology*, *21*(2), 172–174.

Hancock, P. A., & de Ridder, S. (2003). Behavioral response in accident-likely situations. *Ergonomics*, *46*(12), 1111–1135.

Hancock, P. A., & Manser, M. P. (1997). Time-to-contact: More than tau alone. *Ecological Psychology*, *9*(4), 265–297.

Hancock, P. A., & Manser, M. P. (1998). Time-to-contact. In A.-M. Feyer & A. Williamson (Eds.), *Occupational injury: Risk, prevention, and intervention* (pp. 44–59). Taylor & Francis.

Harland, D. M. (2005). *The story of space station Mir*. Springer.

Harrington, D. O., & Drake, M. V. (1971). *The visual fields: A textbook and atlas of clinical perimetry*. Mosby.

Hartmann, W. M., & Wittenberg, A. (1996). On the externalization of sound images. *The Journal of the Acoustical Society of America*, *99*(6), 3678–3688.

He, D., Liu, F., Pape, D., Dawe, G., & Sandin, D. (2000). Video-based measurement of system latency. In *Proceedings of the 4th International Immersive Projection Technology Workshop (IPT '00)* (pp. 111–116). Ames, IA, United States.

Hendrix, C., & Barfield, W. (1996). The sense of presence within auditory virtual environments. *Presence: Teleoperators & Virtual Environments*, *5*(3), 290–301.

Hill, M. I., Adelstein, B. D., & Ellis, S. R. (2004). Achieving minimum latency in virtual environment applications. In *Proceedings of the 2004 IMAGE Society Annual Conference*. Scottsdale, AZ, United States.

Hills, B. L. (1980). Vision, visibility, and perception in driving. *Perception*, *9*(2), 183–216.

Hitchner, L. E., & McGreevy, M. W. (1993). Methods for user-based reduction of model complexity for virtual planetary exploration. In J. P. Allebach & B. E. Rogowitz (Eds.), *Human vision, visual processing, and digital display IV* (pp. 622–637). SPIE.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3), Article 33.

Horswill, M. S., Helman, S., Ardiles, P., & Wann, J. P. (2005). Motorcycle accident risk could be inflated by a time to arrival illusion. *Optometry and Vision Science*, *82*(8), 740–746.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford University Press.

Insko, B. E. (2003). Measuring presence: Subjective, behavioral and physiological methods. In G. Riva, F. Davide, & W. A. IJsselsteijn (Eds.), *Being there: Concepts, effects and measurement of user presence in synthetic environments* (pp. 109–119). Ios Press.

Jacoby, R. H., Adelstein, B. D., & Ellis, S. R. (1996). Improved temporal response in virtual environments through system hardware and software reorganization. In M. T. Bolas, S. S. Fisher, & J. O. Merritt (Eds.), *Stereoscopic displays and virtual reality systems III* (pp. 271–285). SPIE.

Jamson, H. (2001). Image characteristics and their effect on driving simulator validity. In *Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 190–195). Snowmass Village at Aspen, CO, United States.

Jerald, J. J. (2010). *Scene-motion-and latency-perception thresholds for head-mounted displays* [Doctoral dissertation]. University of North Carolina at Chapel Hill, Department of Computer Science.

Jiang, Y., O'Neal, E. E., Franzen, L., Yon, J. P., Plumert, J. M., & Kearney, J. K. (2017). The influence of stereoscopic image display on pedestrian road crossing in a large-screen virtual environment. In *Proceedings of the 14th ACM Symposium on Applied Perception (SAP '17)* (Article 6). ACM.

Jiang, Y., O'Neal, E. E., Yon, J. P., Franzen, L., Rahimian, P., Plumert, J. M., & Kearney, J. K. (2016). Acting together: Joint pedestrian road crossing in an immersive virtual environment. In T. Höllerer, V. Interrante, A. Lécuyer, & E. Suma (Eds.), *Proceedings of the IEEE Virtual Reality Conference 2016 (VR '16)* (pp. 191–192). IEEE.

Jiang, Y., O'Neal, E. E., Yon, J. P., Franzen, L., Rahimian, P., Plumert, J. M., & Kearney, J. K. (2018). Acting together: Joint pedestrian road crossing in an immersive virtual environment. *ACM Transactions on Applied Perception (TAP)*, *15*(2), Article 8.

Jones, J. A., Krum, D. M., Bolas, M. T. (2016). Vertical field-of-view extension and walking characteristics in head-worn virtual environments. *ACM Transaction of Applied Perception (TAP)*, *14*(2), Article 9.

Juan, M. C., & Pérez, D. (2009). Comparison of the levels of presence and anxiety in an acrophobic environment viewed via HMD or CAVE. *Presence: Teleoperators & Virtual Environments*, *18*(3), 232–248.

Jung, J. Y., Adelstein, B. D., & Ellis, S. R. (2000). Discriminability of prediction artifacts in a time-delayed virtual environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(5), 499–502.

Jürgens, H. W. (1999). Ergonomische Datensammlung. In H. Schmidtke, (Ed.), *Handbuch der Ergonomie: Band 3 (2nd edition)*. Hanser.

Kadali, B. R., & Vedagiri, P. (2012). Pedestrians' gap acceptance behavior at mid block location. *International Journal of Engineering and Technology*, *4*(2), 158–161.

Kaiser, M. K., & Mowafy, L. (1993). Optical specification of time-to-passage: Observers' sensitivity to global tau. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(5), 1028–1040.

Kaplan, G. A. (1969). Kinetic disruption of optical texture: The perception of depth at an edge. *Perception & Psychophysics*, *6*(4), 193–198.

Kappé, B., & Korteling, J. E. (1995). *Perception of time to contact in driving simulators* [Technical Report TNO-TM-1995-A-53]. TNO Human Factors Research Institute.

Kearney, J. K., Grechkin, T., Cremer, J., & Plumert, J. (2006). Traffic generation for studies of gap acceptance. In *Proceedings of the 2006 Driving Simulation Conference Europe (DSC '06)* (pp. 177–186). Paris, France.

Kemeny, A. (2001). Recent developments in visuo-vestibular restitution of self-motion in driving simulation. In *Proceedings of the 2001 Driving Simulation Conference Europe (DSC '01)* (pp. 15–18). Sophia Antipolis, France.

Khuu, S. K., Lee, T. C. P., & Hayes, A. (2010). Object speed derived from the integration of motion in the image plane and motion-in-depth signaled by stereomotion and looming. *Vision Research*, *50*(9), 904–913.

Kim, K., Rosenthal, M. Z., Zielinski, D. J., & Brady, R. (2014). Effects of virtual environment platforms on emotional responses. *Computer Methods and Programs in Biomedicine*, *113*(3), 882–893.

Kim, Y., Deille, O., & Nelson, P. A. (2006). Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners. *Journal of Sound and Vibration*, *297*(1/2), 251–266.

Kirkham, R. W., & Landauer, A. A. (1985). Sex differences in the distribution of traffic law enforcement. *Accident Analysis & Prevention*, *17*(3), 211–215.

Kiyokawa, K. (2006). An introduction to head mounted displays for augmented reality. In M. Haller, M. Billinghurst, & B. Thomas (Eds.), *Emerging technologies of augmented reality: Interfaces and design* (pp. 43–63). Idea Group.

Kobayashi, M., Ueno, K., & Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: A psychological and physiological study. *Presence: Teleoperators & Virtual Environments*, *24*(2), 163–174.

Kolasinski, E. M. (1995). *Simulator sickness in virtual environments* [Technical Report 1027]. U.S. Army Research Institute for the Behavioral and Social Sciences.

Kolasinski, E. M., & Gilson, R. D. (1998). Simulator sickness and related findings in a virtual environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(21), 1511–1515.

Konečni, V., Ebbeson, E. B., & Konečni, D. K. (1976). Decision processes and risk taking in traffic: Driver response to the onset of yellow light. *Journal of Applied Psychology*, *61*(3), 359–367.

Kozacek, B., Grauzel, J., & Frivaldsky, M. (2018). The main capabilities and solutions for different types of the image sensors. In P. Hockicko, J. Dubovan, & M. Marković (Eds.), *Proceedings of the 12th International ELEKTRO Conference (ELEKTRO '18)* (Article 41). IEEE.

Kreylos, O. (2017). *Projection and distortion in wide-FoV HMDs* [Blog post]. Retrieved September 30th, 2019, from http://doc-ok.org/?p=1649/

Kubovy, M. (1988). *The psychology of perspective and renaissance art*. Cambridge University Press.

Kuhlen, T., Assenmacher, I., & Lentz, T. (2007). A true spatial sound system for CAVE-like displays using four loudspeakers. In R. Shumaker (Ed.), *Proceedings of the Second International Conference on Virtual Reality (ICVR '07)* (pp. 270–279). Springer.

Kurabayashi, H., Otani, M., Hashimoto, M., & Kayama, M. (2015). Development of dynamic crosstalk cancellation system for multiple-listener binaural reproduction. *Acoustical Science and Technology*, *36*(6), 537–539.

Lai, R. (2017). *Chinese startup's "8K" VR headset is surprisingly advanced* [Blog post]. Retrieved September 30th, 2019, from https://www.engadget.com/2017-10-12-pimax-8k-vr-headset.html

Lappin, J. S. (2014). What is binocular disparity? *Frontiers in Psychology*, *5*, Article 870.

Larsen, J. J. & Pilgaard, M. (2015) *The effect of spatial audio on immersion, presence, and physiological response in games* [Master's thesis]. Aalborg University, Department of Architecture and Media Technology.

Larsson, P., Västfjäll, D., Olsson, P., & Kleiner, M. (2007). When what you hear is what you see: Presence and auditory-visual integration in virtual environments. In *Proceedings of the 10th Annual International Workshop on Presence* (pp. 11–18). ISPR.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, *5*(4), 437–459.

Lee, D. N., Young, D. S., & McLaughlin, C. M. (1984). A roadside simulation of road crossing for children. *Ergonomics*, *27*(12), 1271–1281.

Lehsing, C. (2019). *Untersuchungen zum Interaktionsverhalten urbaner Verkehrsteilnehmer* [Doctoral dissertation]. Technical University of Munich, Department of Mechanical Engineering.

Lehsing, C., & Feldstein, I. T. (2018). Urban interaction – Getting vulnerable road users into driving simulation. In K. Bengler, J. Drüke, S. Hoffmann, D. Manstetten, & A. Neukum (Eds.), *UR:BAN Human factors in traffic* (pp. 347–362). Springer.

Lehsing, C., Kracke, A., & Bengler, K. (2015). Urban perception-a cross-correlation approach to quantify the social interaction in a multiple simulator setting. In *Proceedings of the IEEE 18th International Conference on Intelligent Transportation Systems (ITSC '15)* (pp. 1014–1021). IEEE.

Lehsing, C., Ruch, F., Kölsch, F. M., Dyszak, G. N., Haag, C., Feldstein, I. T., Savage, S. W., & Bowers, A. R. (2019). Effects of simulated mild vision loss on gaze, driving and interaction behaviors in pedestrian crossing situations. *Accident Analysis & Prevention*, *125*, 138–151.

Lentz, T. (2006). Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments. *Journal of the Audio Engineering Society*, *54*(4), 283–294.

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

Leyrer, M., Linkenauger, S. A., Bülthoff, H. H., Kloos, U., & Mohler, B. (2011). The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (AGVP '11)* (pp. 67–74). ACM.

Li, F. X., & Laurent, M. (1995). Occlusion rate of ball texture as a source of velocity information. *Perceptual and Motor Skills*, *81*(3), 871–880

Liang, J., Shaw, C., & Green, M. (1991). On temporal-spatial realism in the virtual reality environment. In *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology (UIST '91)* (pp. 19–25). ACM.

Lindau, A. (2009). The perception of system latency in dynamic binaural synthesis. In M. Boone (Ed.), *Proceedings of the NAG/DAGA 2009 International Conference on Acoustics* (pp. 1063–1066). Rotterdam, the Netherlands.

Link, N. K., Kruk, R. V., McKay, D., Jennings, S. A., & Craig, G. (2002). Hybrid enhanced and synthetic vision system architecture for rotorcraft operations. In J. G. Verly (Ed.), *Enhanced and Synthetic Vision* (pp. 190–201). SPIE.

Lobjois, R., Benguigui, N., & Cavallo, V. (2013). The effects of age and traffic density on street-crossing behavior. *Accident Analysis & Prevention*, *53*, 166–175.

Lobjois, R., & Cavallo, V. (2007). Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task. *Accident Analysis & Prevention*, *39*(5), 934–943.

Lobjois, R., & Cavallo, V. (2009). The effects of aging on street-crossing behavior: from estimation to actual crossing. *Accident Analysis & Prevention*, *41*(2), 259–267.

Loftus, E. F. (1979). *Eyewitness testimony*. Harvard University Press.

López-Moliner, J., Brenner, E., & Smeets, J. B. (2007). Effects of texture and shape on perceived time to passage: Knowing “what” influences judging “when.” *Perception & Psychophysics*, *69*(6), 887–894.

Luebke, D., & Hallen, B. (2001). Perceptually driven simplification for interactive rendering. In S. J. Gortler & K. Myszkowski (Eds.), *Proceedings of the 12th Eurographics Workshop on Rendering Techniques* (pp. 223–234). ACM.

Lugtigheid, A. J., & Welchman, A. E. (2011). Evaluating methods to measure time-to-contact. *Vision Research*, *51*(20), 2234–2241.

Maag, C., Mühlbacher, D., Mark, C., & Kruger, H. P. (2012). Studying effects of advanced driver assistance systems (ADAS) on individual and group level using multi-driver simulation. *IEEE Intelligent Transportation Systems Magazine*, *4*(3), 45–54.

Mackensen, P. (2004). *Auditive localization. Head movements, an additional cue in localization* [Doctoral dissertation]. Technical University Berlin, Faculty I – Humanities.

Mallaro, S., Rahimian, P., O’Neal, E. E., Plumert, J. M., & Kearney, J. K. (2017). A comparison of head-mounted displays vs. large-screen displays for an interactive pedestrian simulator. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)* (Article 6). ACM.

Mania, K., Adelstein, B. D., Ellis, S. R., & Hill, M. I. (2004). Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization (AGVP '04)* (pp. 39–47). ACM.

Manser, M. P., & Hancock, P. A. (1996). Influence of approach angle on estimates of time-to-contact. *Ecological Psychology*, *8*(1), 71–99.

Manstetten, D., Bengler, K., Busch, F., Färber, B., Lehsing, C., Neukum, A., Petermann-Stock, I., & Schendzielorz, T. (2013). “UR:BAN MV Human Factors in Traffic” – a German Research Project to Increase Safety in Urban Traffic. In *Proceedings of the 20th ITS World Congress*. Tokyo, Japan.

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. The MIT Press.

Masiero, B., & Vorländer, M. (2014). A framework for the calculation of dynamic crosstalk cancellation filters. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(9), 1345–1354. IEEE.

Mather, G. (1996). Image blur as a pictorial depth cue. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *263*(1367), 169–172.

Mathieu, J., Bootsma, R. J., Berthelon, C., & Montagne, G. (2017). Judging arrival times of incoming traffic vehicles is not a prerequisite for safely crossing an intersection: Differential effects of vehicle size and type in passive judgment and active driving tasks. *Acta Psychologica*, *173*, 1–12.

Matthews, M. L. (1986). Aging and the perception of driving risk and ability. *Proceedings of the Human Factors Society Annual Meeting*, *30*(12), 1159–1163.

McGuckin, N., & Fucci, A. (2018). *Summary of Travel Trends: 2017 National Household Travel Survey* [Technical Report FHWA-PL-18-019]. Federal Highway Administration, U.S. Department of Transportation.

McKee, S. P. (1981). A local mechanism for differential velocity detection. *Vision Research*, *21*(4), 491–500.

McLeod, R. W., & Ross, H. E. (1983). Optic-flow and cognitive factors in time-to-collision estimates. *Perception*, *12*(4), 417–423.

Meehan, M., Razzaque, S., Whitton, M. C., & Brooks, F. P., Jr. (2003). Effect of latency on presence in stressful virtual environments. In J. Chen, B. Froehlich, B. Loftin, U. Neumann, & H. Takemura (Eds.), *Proceedings of the IEEE Virtual Reality Conference 2003 (VR '03)* (pp. 141–148). IEEE.

Meir, A., Oron-Gilad, T., & Parmet, Y. (2015). Are child-pedestrians able to identify hazardous traffic situations? Measuring their abilities in a virtual reality environment. *Safety Science*, *80*, 33–40.

Meir, A., Parmet, Y., & Oron-Gilad, T. (2013). Towards understanding child-pedestrians' hazard perception abilities in a mixed reality dynamic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *20*, 90–107.

Merriken, M. S., Johnson, W. V., Cress, J. D., & Riccio, G. E. (1988). Time delay compensation using supplementary cues in aircraft simulator systems. In *Proceedings of the AIAA Flight Simulation Technologies Conference* (pp. 295–303). AIAA.

Mine, M. R. (1993). *Characterization of end-to-end delays in head-mounted display systems* [Technical Report TR93-001]. The University of North Carolina at Chapel Hill.

Mohler, B. J., Bülthoff, H. H., Thompson, W. B., & Creem-Regehr, S. H. (2008). A full-body avatar improves egocentric distance judgments in an immersive virtual environment. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APVG '08)* (p. 194). ACM.

Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., & Bülthoff, H. H. (2010). The effect of viewing a self-avatar on distance judgments in an HMD-based virtual environment. *Presence: Teleoperators & Virtual Environments*, *19*(3), 230–242.

Molyneux, W. (1690). *Dipotrica nova: A treatise of dioptricks, in two parts*. Benjamin Tooke.

Montgomery, J., Kusano, K. D., & Gabler, H. C. (2014). Age and gender differences in time to collision at braking from the 100-car naturalistic driving study. *Traffic Injury Prevention*, *15*(Supp. 1), 15–20.

Moore, R. L. (1953). Pedestrian choice and judgment. *Journal of the Operational Research Society*, *4*(1), 3–10.

Morrison, G. (2014). *Black frame insertion: Busting blur from Oculus to LCD TVs* [Blog post]. Retrieved January 18th, 2019, from https://www.cnet.com/tech/home-entertainment/black-frame-insertion-busting-blur-from-oculus-to-lcd-tvs/

Morrongiello, B. A., Corbett, M. R., Milanovic, M., Pyne, S., & Vierich, R. (2015). Innovations in using virtual reality to study how children cross streets in traffic: Evidence for evasive action skills. *Injury Prevention*, *21*(4), 266–270.

Mourant, R. R., & Rockwell, T. H. (1970). Mapping eye-movement patterns to the visual scene in driving: An exploratory study. *Human Factors*, *12*(1), 81–87.

Nagata, S. (1989). How to reinforce perception of depth in single two-dimensional pictures. In S. R. Ellis, M. K. Kaiser, & A. J. Grunwald (Eds.), *Spatial displays and spatial instruments* (Article 20). NASA.

Nahin, P. J. (2012). *Chases and escapes*. Princeton University Press.

Nordahl, R., & Nilsson, N. C. (2014). The sound of being there: presence and interactive audio in immersive virtual reality. In K. Collins, B. Kapralos, & H. Tessler (Eds.), *The Oxford Handbook of Interactive audio* (pp. 213–233). Oxford University Press.

Okoshi, T. (1976). *Three-dimensional imaging techniques*. Academic Press.

O'Neal, E. E., Jiang, Y., Franzen, L. J., Rahimian, P., Yon, J. P., Kearney, J. K., & Plumert, J. M. (2018). Changes in perception–action tuning over long time scales: How children and adults perceive and act on dynamic affordances when crossing roads. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 18–26.

Oxley, J. A., Ihsen, E., Fildes, B. N., Charlton, J. L., & Day, R. H. (2005). Crossing roads safely: An experimental study of age differences in gap selection by pedestrians. *Accident Analysis & Prevention*, *37*(5), 962–971.

Palmisano, S., Gillam, B., Govan, D. G., Allison, R. S., & Harris, J. M. (2010). Stereoscopic perception of real depths at large distances. *Journal of Vision*, *10*(6), Article 19.

Parsonson, B. S., Isler, R. B., & Hansson, G. J. (1996). *Driver behaviour at rural T-intersections* [Research Report 56]. Transit New Zealand.

Paul, J. (2016). *Rolling shutter vs global shutter: What's the difference?* [Blog post]. Retrieved November 21st, 2018, from http://premiumbeat.com/blog/know-the-basics-of-global-shutter-vs-rolling-shutter/

Pausch, R., Crea, T., & Conway, M. (1992). A literature survey for virtual environments: Military flight simulator visual systems and simulator sickness. *Presence: Teleoperators & Virtual Environments*, *1*(3), 344–363.

Peli, E. (1999). Optometric and perceptual issues with head-mounted displays. In P. Mouroulis (Ed.), *Visual Instrumentation: Optical Design and Engineering Principles* (pp. 205–276). McGraw-Hill.

Pelzer, S., Masiero, B., & Vorländer, M. (2014). 3D reproduction of room auralizations by combining intensity panning, crosstalk cancellation and Ambisonics. In S. Weinzierl, M. Vorländer, F. Zotter, H.-J. Maempel, & A. Lindau (Eds.), *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics 2014* (pp. 182–188). Berlin, Germany.

Peñín, L. F. (2000). Teleoperation with time delay: A survey and its use in space robotics. In P. Putz & M. Winnendael (Eds.), *Proceedings of the 6th ESA Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA '00)* (Article 3.5b-4). Noordwijk, the Netherlands.

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: perceptual illusion of body swapping. *PLoS ONE, 3*(12), Article e3832.

Petzoldt, T. (2014). On the relationship between pedestrian gap acceptance and time to arrival estimates. *Accident Analysis & Prevention*, *72*, 127–133.

Phillips, L., Ries, B., Kaeding, M., & Interrante, V. (2010). Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments. In B. Lok, G. Klinker, & R. Nakatsu (Eds.), *Proceedings of the IEEE Virtual Reality Conference 2010 (VR '10)* (pp. 115–1148). IEEE.

Pitcairn, T. K., & Edlmann, T. (2000). Individual differences in road crossing ability in young children and adults. *British Journal of Psychology*, *91*(3), 391–410.

Plumert, J. M., Kearney, J. K., & Cremer, J. F. (2004). Children's perception of gap affordances: Bicycling across traffic-filled intersections in an immersive virtual environment. *Child Development*, *75*(4), 1243–1253.

Poeschl, S., Wall, K., & Doering, N. (2013). Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. In S. Coquillart, J. S. LaViola Jr., & D. Schmalstieg (Eds.), *Proceedings of the IEEE Virtual Reality Conference 2013 (VR '13)* (pp. 129–130). IEEE.

Polus, A., Hocherman, I. & Efrat, E. (1988). Evaluation of the accident rates of male and female drivers. *Transportation Research Record*, *1172*, 42–46.

Pundlik, S., Tomasi, M., Liu, R., Houston, K., & Luo, G. (2019). Development and Preliminary Evaluation of a Smartphone App for Measuring Eye Alignment. *Translational Vision Science & Technology*, *8*(1), Article 19.

Raaen, K., & Kjellmo, I. (2015). Measuring latency in virtual reality systems. In K. Chorianopoulos, M. Divitini, J. Baalsrud Hauge, L. Jaccheri, & R. Malaka (Eds.), *Proceedings of the 14th International Conference on Entertainment Computing (ICEC '15)* (pp. 457–462). Trondheim, Norway.

Rahimian, P., O'Neal, E. E., Yon, J. P., Franzen, L., Jiang, Y., Plumert, J. M., & Kearney, J. K. (2016). Using a virtual environment to study the impact of sending traffic alerts to texting pedestrians. In T. Höllerer, V. Interrante, A. Lécuyer, & E. Suma (Eds.), *Proceedings of the IEEE Virtual Reality Conference 2016 (VR '16)* (pp. 141–149). IEEE.

Rand, K. M., Tarampi, M. R., Creem-Regehr, S. H., & Thompson, W. B. (2011). The importance of a visual horizon for distance judgments under severely degraded vision. *Perception*, *40*(2), 143–154.

Recarte, M. Á., Conchillo, Á., & Nunes, L. M. (2005). Estimation of arrival time in vehicle and video. *Psicothema*, *17*(1), 112–117.

Reddy, M. (1995). A survey of level of detail support in current virtual reality solutions. *Virtual Reality*, *1*(2), 95–98.

Regan, D., & Hamstra, S. J. (1993). Dissociation of discrimination thresholds for time to contact and for rate of angular expansion. *Vision Research*, *33*(4), 447–462.

Regan, D., & Vincent, A. (1995). Visual processing of looming and time to contact throughout the visual field. *Vision Research*, *35*(13), 1845–1857.

Regan, M. J., Miller, G. S., Rubin, S. M., & Kogelnik, C. (1999). A real-time low-latency hardware light-field renderer. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)* (pp. 287–290). ACM.

Rejhon, M. (2018). *Why do some OLEDs have motion blur?* [Blog post]. Retrieved January 18th, 2019, from http://blurbusters.com/faq/oled-motion-blur/

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The perception of egocentric distances in virtual environments: A review. *ACM Computing Surveys*, *46*(2), Article 23.

Riecke, B. E., Schulte-Pelkum, J., & Bülthoff, H. H. (2005). Perceiving simulated ego-motions in virtual reality – Comparing large screen displays with HMDs. In B. E. Rogowitz, T. N. Pappas, & S. J. Daly (Eds.), *Human vision and electronic imaging X* (pp. 344–355). SPIE.

Ries, B., Interrante, V., Kaeding, M., & Phillips, L. (2009). Analyzing the effect of a virtual avatar's geometric and motion fidelity on ego-centric spatial perception in immersive virtual environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology (VRST '09)* (pp. 59-66). ACM.

Rittger, L., Muehlbacher, D., Maag, C., & Kiesel, A. (2015). Anger and bother experience when driving with a traffic light assistant: A multi-driver simulator study. In D. de Waard, J. Sauer, S. Röttger, A. Kluge, D. Manzey, C. Weikert, A. Toffetti, R. Wiczorek, K. Brookhuis, & H. Hoonhout (Eds.), *Proceedings of the Human Factors and Ergonomics Society: Europe Chapter 2014 Annual Conference* (pp. 41–51). Lisbon, Portugal.

Robillard, G., Bouchard, S., Renaud, P., & Cournoyer, L. G. (2002). Validation canadienne-française de deux mesures importantes en réalité virtuelle: l'Immersive Tendencies Questionnaire et le Presence Questionnaire [Poster]. At *25ème congrès de la Société Québécoise pour la Recherche en Psychologie (SQRP '02)*. Trois-Rivières, QC.

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews: Neuroscience*, *6*(4), 332–339.

Satava, R. M., & Jones, S. (2003) Medical applications in virtual reality. In L. J. Hettinger & M. W. Haas (Eds.), *Virtual and adaptive environments: Applications, implications, and human performance issues* (pp. 325–343). Lawrence Erlbaum Associates.

Scarfe, P., & Glennerster, A. (2015). Using high-fidelity virtual reality to study perception in freely moving observers. *Journal of Vision*, *15*(9), Article 3.

Schiff, W. (1980). *Perception: An applied approach*. Houghton Mifflin.

Schiff, W., & Detwiler, M. L. (1979). Information used in judging impending collision. *Perception*, *8*(6), 647–658.

Schiff, W., & Oldak, R. (1990). Accuracy of judging time to arrival: Effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 303–316.

Schiff, W., Oldak, R., & Shah, V. (1992). Aging persons' estimates of vehicular motion. *Psychology and Aging*, *7*(4), 518–525.

Schindler, J., Harms, C., Noyer, U., Richter, A., Flemisch, F., Köster, F., Bellet, T., Mayenobe, P., & Gruyer, D. (2011). JDVE: A joint driver-vehicle-environment simulation platform for the development and accelerated testing of automotive assistance and automation systems. In P. C. Cacciabue, M. Hjälmdahl, A. Luedtke, & C. Riccioli (Eds.), *Human modelling in assisted transportation: Models, tools and risk methods* (pp. 233–240). Springer.

Schwebel, D. C., Gaines, J., & Severson, J. (2008). Validation of virtual reality as a tool to understand and prevent child pedestrian injury. *Accident Analysis & Prevention*, *40*(4), 1394–1400.

Sedgwick, H. A. (1983). Environment-centered representation of spatial layout: Available visual information from texture and perspective. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 425–458). Academic Press.

Seward, A. E., Ashmead, D. H., & Bodenheimer, B. (2007). Using virtual environments to assess time-to-contact judgments from pedestrian viewpoints. *ACM Transactions on Applied Perception (TAP)*, *4*(3), Article 18.

Sidaway, B., Fairweather, M., Sekiya, H., & Mcnitt-Gray, J. (1996). Time-to-collision estimation in a simulated driving task. *Human Factors*, *38*(1), 101–113.

Sielhorst, T., Sa, W., Khamene, A., Sauer, F., & Navab, N. (2007). Measurement of absolute latency for video see through augmented reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07)* (pp. 215–220). IEEE.

Simpson, G., Johnston, L., & Richardson, M. (2003). An investigation of road crossing in a virtual environment. *Accident Analysis & Prevention*, *35*(5), 787–796.

Singer, M. J., & Witmer, B. G. (1999). On selecting the right yardstick. *Presence: Teleoperators & Virtual Environments*, *8*(5), 566–573.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators & Virtual Environments*, *8*(5), 560–565.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557.

Slater, M., Linakis, V., Usoh, M., & Kooper, R. (1996). Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '96)* (pp. 163–172). ACM.

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators & Virtual Environments*, *9*(1), 37–51.

Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators & Virtual Environments*, *9*(5), 413–434.

Slater, M., & Usoh, M. (1994). Body centred interaction in immersive virtual environments. In N. M. Thalmann & D. Thalmann (Eds.), *Artificial life and virtual reality* (pp. 125–148). John Wiley & Sons.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *6*(6), 603–616.

Snowden, R. J., Stimpson, N., & Ruddle, R. A. (1998). Speed perception fogs up as visibility drops. *Nature*, *392*(6675), 450.

Soares, F., Silva, E., Pereira, F., Silva, C., Sousa, E., & Freitas, E. (2020). The influence of noise emitted by vehicles on pedestrian crossing decision-making: A study in a virtual environment. *Applied Sciences*, *10*(8), Article 2913.

Soomro, A. (2015). *Reduce input lag in PC games: The definitive guide* [Blog post]. Retrieved January 21st, 2019, from http://displaylag.com/reduce-input-lag-in-pc-games-the-definitive-guide/

Sousa, R., Brenner, E., & Smeets, J. B. (2011). Judging an unfamiliar object's distance from its retinal image size. *Journal of Vision*, *11*(9), Article 10.

Sousa, R., Smeets, J. B., & Brenner, E. (2012). Does size matter? *Perception*, *41*(12), 1532–1534.

Spector, R. H. (1990). Visual fields. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical methods: The history, physical, and laboratory examinations* (pp. 565–572). Butterworths.

Staplin, L. (1995). Simulator and field measures of driver age differences in left-tum gap judgments. *Transportation Research Record*, *1485*, 49–55.

Steed, A. (2008). A simple method for estimating the latency of interactive, real-time graphics simulations. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology (VRST '08)* (pp. 123–129). ACM.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93.

Stevens, K. A. (1981). The information content of texture gradients. *Biological Cybernetics*, *42*(2), 95–105.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680.

Stewart, D., Cudworth, C. J., & Lishman, J. R. (1993). Misperception of time-to-collision by drivers in pedestrian accidents. *Perception*, *22*(10), 1227–1244.

Stone, L. S., & Thompson, P. (1992). Human speed perception is contrast dependent. *Vision Research*, *32*(8), 1535–1549.

Takeuchi, T., & De Valois, K. K. (2000). Velocity discrimination in scotopic vision. *Vision Research*, *40*(15), 2011–2024.

Tang, J. (2015*). A measurement study of vertical synchronization configurations in PC video games* [Online article]. Retrieved January 21st, 2019, from https://www.cse.wustl.edu/~jain/cse567-15/ftp/vsync/index.html

te Velde, A. F., van der Kamp, J., Barela, J. A., & Savelsbergh, G. J. (2005). Visual timing and adaptive behavior in a road-crossing simulation study. *Accident Analysis & Prevention*, *37*(3), 399–406.

Tharanathan, A., & DeLucia, P. R. (2006). Time-to-collision judgments of constant and non-constant velocities: Implications for rear-end collisions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(22), 2463–2467.

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, *22*(3), 377–380.

Thomson, J. A. (1983). Is continuous visual monitoring necessary in visually guided locomotion? *Journal of Experimental Psychology: Human Perception and Performance*, *9*(3), 427–443.

Thomson, J. A., Tolmie, A. K., Foot, H. C., Whelan, K. M., Sarvary, P., & Morrison, S. (2005). Influence of virtual reality training on the roadside crossing judgments of child pedestrians. *Journal of Experimental Psychology: Applied*, *11*(3), 175–186.

Todd, J. T. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 795–810.

Tresilian, J. R. (1991). Empirical and theoretical issues in the perception of time to contact. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(3), 865–876.

Tresilian, J. R. (1995). Perceptual and cognitive processes in time-to-contact estimation: Analysis of prediction-motion and relative judgment tasks. *Perception & Psychophysics*, *57*(2), 231–245.

Udiljak, J. (2016). *Display persistence, or "But the box said it has a 1 millisecond response time!"* [Blog post]. Retrieved November 13th, 2018, from http://jamesudiljak.com/blog/jekyll/update/2016/06/21/display-persistence.html

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, *9*(5), 497–503.

Valve Corporation (2017). *Field of view (FOV)* [Blog post]. Retrieved September 30th, 2019, from http://valvesoftware.com/en/index/deep-dive/fov/

van der Kamp, J., Savelsbergh, G., & Smeets, J. (1997). Multiple information sources in interceptive timing. *Human Movement Science*, *16*(6), 787–821.

Vincent, A., & Regan, D. (1997). Judging the time to collision with a simulated textured object: Effect of mismatching rate of expansion of object size and of texture element size. *Perception & Psychophysics*, *59*(1), 32–36.

von Hofsten, C. (1976). The role of convergence in visual space perception. *Vision Research*, *16*(2), 193–198.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*(4), 205–217.

Warren, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. *Nature*, *336*(6195), 162–163.

Watson, B., Walker, N., Woytiuk, P., & Ribarsky, W. (2003). Maintaining usability during 3D placement despite delay. In J. Chen, B. Froehlich, B. Loftin, U. Neumann, & H. Takemura (Eds.), *Proceedings of the IEEE Virtual Reality Conference 2003 (VR '03)* (pp. 133–140). IEEE.

Watson, J. S., Banks, M. S., von Hofsten, C., & Royden, C. S. (1992). Gravity as a monocular cue for perception of absolute distance and/or absolute size. *Perception*, *21*(1), 69–76.

Wawro, A. (2011). *Geek 101: What is VSync?* [Blog post]. Retrieved from January 21st, 2019, from https://www.pcworld.com/article/491958/geek101-vsync.html

Welch, R. B., Blackmon, T. T., Liu, A., Mellers, B. A., & Stark, L. W. (1996). The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence: Teleoperators & Virtual Environments, 5*(3), 263–273.

Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. I: stimulus synthesis. *The Journal of the Acoustical Society of America*, *85*(2), 858–867.

Willemsen, P., Colton, M. B., Creem-Regehr, S. H., & Thompson, W. B. (2004). The effects of head-mounted display mechanics on distance judgments in virtual environments. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization (AGVP '04)* (pp. 35–38). ACM.

Wilson, D. (2009). *Exploring input lag inside and out* [Blog post]. Retrieved January 21st, 2019, from http://anandtech.com/show/2803/

Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). The factor structure of the Presence Questionnaire. *Presence: Teleoperators & Virtual Environments*, *14*(3), 298–312.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators & Virtual Environments*, *7*(3), 225–240.

Wloka, M. M. (1995). Lag in multiprocessor virtual reality. *Presence: Teleoperators & Virtual Environments*, *4*(1), 50–63.

Wu, W., Dong, Y., & Hoover, A. (2013). Measuring digital system latency from sensing to actuation at continuous 1-ms resolution. *Presence: Teleoperators & Virtual Environments*, *22*(1), 20–35.

Yairi, S., Iwaya, Y., & Suzuki, Y. (2007). Estimation of detection threshold of system latency of virtual auditory display. *Applied Acoustics*, *68*(8), 851–863.

Yan, J. J., Lorv, B., Li, H., & Sun, H. J. (2011). Visual processing of the impending collision of a looming object: Time to collision revisited. *Journal of Vision*, *11*(12), Article 7.

Yannis, G., Papadimitriou, E., & Theofilatos, A. (2013). Pedestrian gap acceptance for mid-block street crossing. *Transportation Planning and Technology*, *36*(5), 450–462.

Young, D. S., & Lee, D. N. (1987). Training children in road crossing skills using a roadside simulation. *Accident Analysis & Prevention*, *19*(5), 327–341.

Ilja Feldstein's dissertation deals with latency within virtual environments that is of unique importance for the development of practical VR systems. He provides specific examples of how latency in VR or similar systems may be conveniently measured and managed, in particular within VR simulation for road-safety research, such as pedestrians who face a virtual crosswalk in the presence of automobile traffic. Measuring and managing the impact of latency on user performance within distant interacting computer systems is likely to remain a continual problem and Dr. Feldstein's work is likely to be consulted well into the future. A remarkable contribution of his dissertation is also his proposed simulator validation technique. His step-back method is itself an interesting new measure that can be utilized in pedestrian simulators as well as corresponding physical situations and suggests a general approach to investigating a variety of risk judgments that balance apparent risk in corresponding virtual and real environments.

Stephen R. Ellis, NASA head scientist