Perception-Based Techniques to Enhance User Experience in Virtual Reality

Von der
Carl-Friedrich-Gauß Fakultät
der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines
Doktoringenieurs (Dr.-Ing.)

genehmigte Dissertation
(kumulative Arbeit)

von
Colin Groth
geboren am 24. Oktober 1994
in Wolfsburg

Eingereicht am:	17. April 2024
Disputation am:	26. Juli 2024
1. Referent:	Prof. Dr.-Ing. Marcus A. Magnor
2. Referent:	Prof. Dr.-Ing. Piotr Didyk

2024

To my father.

Abstract

Virtual reality (VR) ushered in a new era of immersive content viewing with vast potential for entertainment, design, medicine, and other fields. However, the willingness of users to practically apply the technology is bound to the quality of the virtual experience. In this dissertation, we describe the development and investigation of novel techniques to reduce negative influences on the user experience in VR applications. Our methods not only include substantial technical improvements but also consider important characteristics of human perception that are exploited to make the applications more effective and subtle. Mostly, we are focused on visual perception, since we deal with visual stimuli, but we also consider the vestibular sense which is a key component for the occurrence of negative symptoms in VR, referred to as cybersickness. In this dissertation, our techniques are designed for three groups of VR applications, characterized by the degree of freedom to apply adjustments.

The first set of techniques addresses the extension of VR systems with stimulation hardware. By adjusting common techniques from the medical field, we artificially induce human body signals to create immersive experiences that reduce common mismatches between perceptual information.

The second group focuses on applications that use common hardware and allow adjustments of the full render pipeline. Here, especially immersive video content is notable, where the frame rates and quality of the presentations are often not in line with the high requirements of VR systems to satisfy a decent user experience. To address the display problems, we present a novel video codec based on wavelet compression and perceptual features of the visual system.

Finally, the third group of applications is the most restrictive and does not allow modifications of the rendering pipeline. Here, our techniques consist of post-processing manipulations in screen space after rendering the image, without knowledge of the 3D scene. To allow techniques in this group to be subtle, we exploit fundamental properties of human peripheral vision and apply spatial masking as well as gaze-contingent motion scaling in our methods.

Kurzfassung

Die virtuelle Realität (VR) hat eine neue Ära der immersiven Betrachtung von Inhalten eingeläutet, die ein enormes Potenzial für viele Bereiche bietet. Dabei ist die Bereitschaft der Nutzer, die Technologie praktisch anzuwenden, an die Qualität der virtuellen Erfahrung gebunden. Diese Dissertation beschreibt die Entwicklung und Untersuchung neuer Techniken, die negative Einflüsse auf das Nutzererlebnis in VR-Anwendungen reduzieren. Unsere Methoden beinhalten nicht nur wesentliche technische Verbesserungen, sondern berücksichtigen auch wichtige Eigenschaften der menschlichen Wahrnehmung, die unsere Anwendungen effektiver und subtiler gestalten. Diese Arbeit konzentriert sich hauptsächlich auf visuelle Wahrnehmung, betrachtet aber auch den vestibulären Sinn, der eine Schlüsselkomponente für die Entstehung von negativen Symptomen in VR darstellt. In dieser Dissertation werden Techniken für drei Kategorien von VR-Anwendungen entwickelt, die sich durch ihre Flexibilität zur Anpassbarkeit unterscheiden.

Die erste Kategorie von Methoden befasst sich mit der Erweiterung von VR-Systemen durch Stimulationshardware. Durch die Anpassung gängiger Verfahren aus dem medizinischen Bereich induzieren wir künstlich menscheneigene Körpersignale, um immersive Erlebnisse zu schaffen, die übliche Diskrepanzen zwischen Wahrnehmungsinformationen verringern.

Die zweite Kategorie von Methoden konzentriert sich auf Anwendungen, die herkömmliche Grafik-Hardware verwenden, bei denen jedoch Anpassungen der gesamten Rendering-Pipeline möglich sind. Hier sind vor allem immersive Videoinhalte von Interesse, bei denen die Bildraten und die Qualität der Darstellungen oft nicht mit den hohen Anforderungen von VR-Systemen übereinstimmen. Um die Darstellungsprobleme zu lösen, stellen wir einen neuen Videocodec vor, der auf Wavelet-Kompression und Wahrnehmungsmerkmalen des visuellen Systems basiert.

Die dritte Kategorie von Anwendungen ist am restriktivsten und erlaubt keine direkte Modifikation der virtuellen Welt. Hier wenden unsere Methoden Post-Processing-Manipulationen auf das gerenderte Bild, ohne Kenntnis der 3D-Szene vorauszusetzen. Damit die Techniken dieser Kategorie subtil wirken können, nutzen wir grundlegende Eigenschaften des menschlichen peripheren Sehens aus.

Acknowledgements

This thesis would not have been possible without the great support of the many wonderful people in my life. I would like to thank all these people for every form of support and the incredible enrichment of my life during this time.

Special thanks go to my supervisor Marcus Magnor who made this journey possible in the first place. Marcus, you have always encouraged and supported me to get the best out of myself and to take my scientific journey to new heights. Being on your team has given me an unforgettable time that I will look back on with great pleasure. You always supported me, had my back, and made it possible to travel to many exciting conferences. For that, I thank you from the bottom of my heart.

Fortunately, I didn't have to spend my time alone, but with great colleagues. Special gratitude goes to JP, who accompanied me from my Master's time and was always supportive. I would also like to thank Steve, who was not only a good person to talk to in the morning but also had great advice on scientific matters. Thanks to Sascha, with whom I was able to spend four years in an office that was not always productive, but never boring. Of course, my PhD was also enriched by all the other teammates and many laughs were shared. Thanks to Moritz (both of them), Flo, Fabian, Leslie, Marc, Martin, Anja, Michelle, Carsten, Susana and Felix. Thanks also to my friends Jannis and Clara. I am very happy that you joined our university and that I was able to get to know you. I also received a lot of help from the academic assistants who supported me over the years, especially Nikkel, Max, and Timon.

A special thanks also goes to Piotr Didyk who generously and openly welcomed me into his lab and gave me incredibly great support during my time in Lugano and beyond. Piotr, you raised my passion and curiosity for human perception to a new level and taught me valuable new skills. I would also like to thank Taimoor, Ata, Jorge, and Luca for the fantastic atmosphere and inspiring conversations during my stay.

Finally, I would like to thank the people I love. You are always there for me and make my life so much better. I know I can always count on you in every situation. I would never have gotten this far without you.

Funding for the works of this dissertation was generously provided by the German Science Foundation — project *Immersive Digital Reality* (DFG MA2555/15-1).

List of Publications

All papers that comprise this cumulative dissertation have been authored by me in cooperation with several collaborators. I am the responsible first author of and main contributor to these publications. Under the supervision and guidance of my advisor Marcus Magnor, I devised the techniques, developed the methods and prototypes, performed the analyses, wrote the manuscript, and presented the work at national and international conferences, both in-person and online. The general contribution of my co-authors is as follows. Marcus Magnor directed the research narrative, providing the big picture and scientific horizon. He also gave practical guidance for thriving in a challenging research environment. Jan-Philipp Tauscher gave support and guidance, especially in data collection, experimental design, statistical analysis, and ethical approvals. Susana Castillo helped me improve my skills in scientific writing, experimental design, and university politics. My office colleague Sascha Fricke participated in countless discussions of implementation problems and project ideas. His framework further helped me to implementing high-performance applications. The help of, and fruitful discussions with Steve Grogorick guided me towards the right direction in hard times. He helped me to achieve efficient implementations and well-designed experiments. Martin Eisemann served as an additional advisor in our lab and provided helpful ideas for the projects. During my research stay at Università della Svizzera italiana in Switzerland, Piotr Didyk hosted me in this group, welcoming me into an open and inspiring environment. He gave me supervision and guidance during my time abroad and after. Nikkel Heesen, Max Hattenbach, and Timon Scholz served as lab assistants for different projects and helped with the implementation of basic methods, as well as the conduction of scientific experiments.

All manuscripts of the six publications A – F in this thesis are reproduced verbatim. Modifications are solely limited to formatting and layout for visual consistency. The structure of the main part of my dissertation follows the generalizability of the methods, as outlined in the introduction.

A

Colin Groth and Jan-Philipp Tauscher and Nikkel Heesen and Max Hattenbach and Susana Castillo and Marcus Magnor:
Omnidirectional Galvanic Vestibular Stimulation in Virtual Reality
Published in *IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG) 2022,* Volume 28, Article 5, Pages 2234–2244. DOI: 10.1109/TVCG.2022.3150506
Presented at *IEEE VR 2022, ACM SIGGRAPH 2022, IEEE VIS 2022.*

The idea for this paper evolved based on early thoughts of my advisor Marcus Magnor and me. A proof-of-concept of GVS for cybersickness mitigation in VR was provided with the bachelor thesis of Max Hattenbach. The implementation of the bachelor thesis was further refined and extended by my lab assistant Nikkel Heesen and me to form the final project. The ethics application and many thoughts for data collection and experimental design were supported by Jan-Philipp Tauscher. Susana Castillo helped with paper writing.

B

Colin Groth and Timon Scholz and Susana Castillo and Jan-Philipp Tauscher and Marcus Magnor:
Instant Hand Redirection in Virtual Reality Through Electrical Muscle Stimulation-Triggered Eye Blinks
Published in *ACM Symposium on Virtual Reality Software and Technology (ACM VRST) 2023,* Article 37, Pages 1–11. DOI: 10.1145/3611659.3615717
Presented at *ACM VRST 2023.*

This project exploited perceptual effects during actively stimulated eye blinks. Leon von der Emde of the Hospital Clinic Bonn gave advice from a medical perspective in the development of the stimulation model and the safety of the approach. Jan-Philipp Tauscher helped to set up the eye tracker and shared thoughts on the experimental design. Timon Scholz assisted with the design of the virtual environment of the VR experiment and with conducting the experiment. Susana Castillo supported the project by proofreading the manuscript. Marcus Magnor helped with the paper writing and gave advice on further directions within the course of the study.

C

Colin Groth and Sascha Fricke and Susana Castillo and Marcus Magnor:
Wavelet-Based Fast Decoding of 360° Videos
Published in *IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG) 2023,* Volume 29, Article 5, Pages 2508–2516. DOI: 10.1109/TVCG.2023.3247080
Presented at *IEEE VR 2023*.

The project idea was developed during constructive discussions with Sascha Fricke. He also provided his Vulkan framework which offered the basic rendering structure for this work. Susana Castillo contributed ideas for experiment design and analysis of the perceptual experiment. Marcus Magnor shared insightful thoughts which allowed for more effective compression and the right focus of the work.

D

Colin Groth and Jan-Philipp Tauscher and Nikkel Heesen and Susana Castillo and Marcus Magnor:
Visual Techniques to Reduce Cybersickness in Virtual Reality
Published in *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) 2021,* Pages 486–487. DOI: 10.1109/VRW52623.2021.00125
Presented at *IEEE VR 2021*.

The gap in research on cybersickness mitigation was pointed out by Marcus Magnor who also helped to develop the concrete idea. Jan-Philipp Tauscher and Susana Castillo gave support and guidance during the whole project, helping with the experiment design, analysis, and writing. My lab assistant Nikkel Heesen supported me in the experiment conduction.

E

Colin Groth and Jan-Philipp Tauscher and Nikkel Heesen and Steve Grogorick and Susana Castillo and Marcus Magnor:
Mitigation of Cybersickness in Immersive 360° Videos
Published in *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) 2021,* Pages 169–177. DOI: 10.1109/VRW52623.2021.00039
Presented at *IEEE VR 2021*.

Since this project came about in parallel to the former paper, the guidance by Jan-Philipp Tauscher and Susana Castillo was also provided for this work. Steve Grogorick helped me in the implementation of a foveated blur filter. Again, Nikkel Heesen supported me in the extensive conduction of the experiment sessions. Marcus Magnor gave helpful advice and clarified the importance of panorama videos in VR. This advice allowed a more innovative publication.

F

Colin Groth and Marcus Magnor and Steve Grogorick and Martin Eisemann and Piotr Didyk:

Cybersickness Reduction via Gaze-Contingent Image Deformation

Accepted to *ACM Transactions on Graphics (ACM TOG) 2024,* Volume 43, Number 4. DOI: 10.1145/3658138

To be presented at *ACM SIGGRAPH 2024*.

This work arose in collaboration with Piotr Didyk during my research visit in 2023 in his lab in Lugano, Switzerland. Piotr Didyk advised me in the project narrative, gave crucial ideas for the implementation, and assisted in paper writing. Steve Grogorick gave guidance for the development of the application. Marcus Magnor gave general advice and support, and helped with the paper writing.

Measurable Contributions

In total, including the previous first-author publications, this dissertation has generated the following contributions and involvements of the author:

3 Peer-reviewed articles in JCR-indexed journals (A, C, F)

1 Peer-reviewed paper in conference proceedings (B)

2 Further peer-reviewed publications (D, E)

1 Honorable mention award (IEEE VR)

2 Invited talks (ACM SIGGRAPH, IEEE VIS)

1 Committee member position (IEEE VR)

15 Conference and journal reviews

6 Co-advised bachelor and master theses

8 Semesters as a teaching assistant for multiple courses

Contents

1 Introduction

> *Virtual reality is not a media experience. When it's done well, it's an actual experience. In the long term, I think we're going to remember the content of what happens in VR because it's going to be as if it happened to us in real life.*
>
> — Jeremy Bailenson (Stanford University)

From a speculative idea to a dynamic, engaging platform, virtual reality (VR) has undergone a remarkable journey of technological innovation and creative vision. VR has the ability to transport users to entirely different worlds that fulfill unmet desires of reality. As VR continues to evolve, it has the potential to shape our interactions with digital content and the world around us. Thereby, VR is not just about more realistic gaming or novel experiences; it is about redefining the way we interact with technology, opening new frontiers in various fields and profoundly impacting our daily lives.

VR has come a long way from its conceptual origins to its current state as an immersive, interactive experience. The roots of VR can be traced back to the 1950s and 1960s, with the pioneering work of Morton Heilig and Ivan Sutherland [Heilig 1962, Sutherland 1968]. Heilig's *Sensorama*, an early form of VR, was a multisensory machine providing a simulated experience of riding a motorcycle through the streets of Brooklyn [Heilig 1962]. Sutherland, often regarded as the father of computer graphics, conceptualized the *Ultimate Display* – an early vision of the perfect VR system [Sutherland et al. 1965]. Sutherland developed the first head-mounted display system which showcased the potential and fascination of the technology [Sutherland 1968]. From today's viewpoint, his creation, known as the *Sword of Damocles*, was a primitive prototype that displayed simple wireframe graphics.

The 1970s and 1980s witnessed significant advancements in computer graphics and 3D modeling, crucial for VR's growth. During that time the first attempts aimed to extend virtual experiences with tactile feedback. Exemplary is the research of Frederick Brooks at that time [Brooks Jr et al. 1990]. Brooks led the development of GROPE, a project that used haptic feedback in VR to study molecular biology. At the same time, VPL Research, founded by Jaron Lanier, published the VPL DataGlove for commercial use of hand interactions in entertainment [Greenemeier 2015]. This meaningful step

evolved to hand interactions being the main medium of communication in augmented and virtual reality today. Also, NASA began experimenting with VR for space simulations. With the Virtual Environment Workstation (VIEW) project, they pushed extensive research in the field ranging from the creation of multisensory and immersive environments to telepresence and teleoperation applications [NASA Ames Research Center 1985].

In the 1990s, public interest and investments in VR technology began to rise. The period was marked by the release of consumer VR products like Nintendo's *Virtual Boy*, a portable console providing stereoscopic 3D graphics [Nintendo Co., Ltd. 1995]. Although these early attempts were often clunky and limited, they sparked people's imagination about the potential of VR.

The 2000s were a period of technological maturation. Graphics processing units (GPUs) became more powerful and affordable, based on which high-quality 3D graphics became more accessible to a broader audience. In this era, multi-user environments became popular. An example is *Second Life*, which offered a glimpse into social interactions within virtual spaces [Linden ResearchF].

The 2010s marked a renaissance in VR, brought on by the introduction of the Oculus Rift [Rubin 2015]. Crowdfunding platforms and increasing interest from tech giants like Facebook, Google, and Sony accelerated the advancements and development of innovative products. Modern VR systems, like the HTC Vive and Oculus Quest, offer high-fidelity, immersive experiences, with applications expanding beyond gaming into education, healthcare, and training.

Nowadays, VR technology is rapidly evolving with applications in a variety of domains and immense potential for future applications. In medicine and healthcare, VR is increasingly used for therapeutic interventions and training to enhance patient care and medical education [Song et al. 2024, Ollonazarovich 2024, Eremita and Chitra 2023, Schwartz et al. 2024]. The application of VR in movie production offers innovative storytelling and filmmaking through immersive experiences that invigorate the observers' engagement [Mateer 2017, Bouville et al. 2016]. The impact of VR in entertainment, particularly through video games and 360° videos, brings a new level of immersion and interactivity, creating more engaging and realistic virtual experiences [Das et al. 2017, Zyda 2005, Hock et al. 2017]. High-risk training is another important field. VR provides safe, realistic environments for training procedures in fields like aircraft simulation [Vora et al. 2002, Cevette et al. 2012] and surgery [Zabaleta Jiménez et al. 2024, Corvino et al. 2024, Weyhe et al. 2018], enabling professionals to practice complex and dangerous tasks without the associated risks. In engineering and design, VR is used as a tool for visual-

ization and prototyping [Chu et al. 2024, Berni and Borgianni 2020]. Finally, VR also finds increasing popularity in the visualization of scientific data where the technology opens up new perspectives in understanding complex information [Cassidy et al. 2020, El Beheiry et al. 2019, García-Hernández et al. 2016].

1.1 Problem Statement

VR stands at the forefront of technological innovation, offering engaging experiences across many fields like education, gaming, and medicine. The essence of VR's potential lies in its ability to deliver immersive experiences and spatial interaction in ways traditional media cannot match. However, the huge potential of VR is undermined by significant challenges that limit the user experience, notably cybersickness and interaction feedback. Negative physical symptoms that arise from movements inside the VR experience, referred to as cybersickness, significantly disrupt user comfort and immersion due to conflicts between the visual sense and the vestibular system. On the other hand, missing tactile feedback of the interaction with virtual objects can disrupt the user's sense of presence and interaction fidelity in VR. Addressing these issues is crucial; overcoming cybersickness and improving interaction fidelity are key aspects to unlock the full potential of VR, making it a more accessible and impactful technology. In this dissertation, I describe the development of subtle methods that preserve the virtual camera movements and presentation but reduce the negative implications of VR, allowing for more enjoyable and engaging experiences.

1.2 Task and Challenges

The goal of this dissertation is to provide effective and subtle methods that allow to mitigate negative effects in VR arising from mismatches of perceptual modalities. Particularly, we consider two of the most detrimental effects in VR scenarios and focus on vestibular-visual discrepancies, causing cybersickness in VR users, and lack in visual-tactile matching, causing breaks in presence due to unmet expectations. For the techniques, we distinguish between three categories of target applications that are increasingly more isolated and differ in the requirements that can be fulfilled (see Figure 1.1).

In the first category, VR systems offer opportunities for hardware extension with additional sensors. Here, we incorporate professional stimulation hardware to artificially induce reactions in the body by simulating signals of either rotational forces or muscle impulses of the eyelid. These effective physiological stimulations are fine-tuned to be

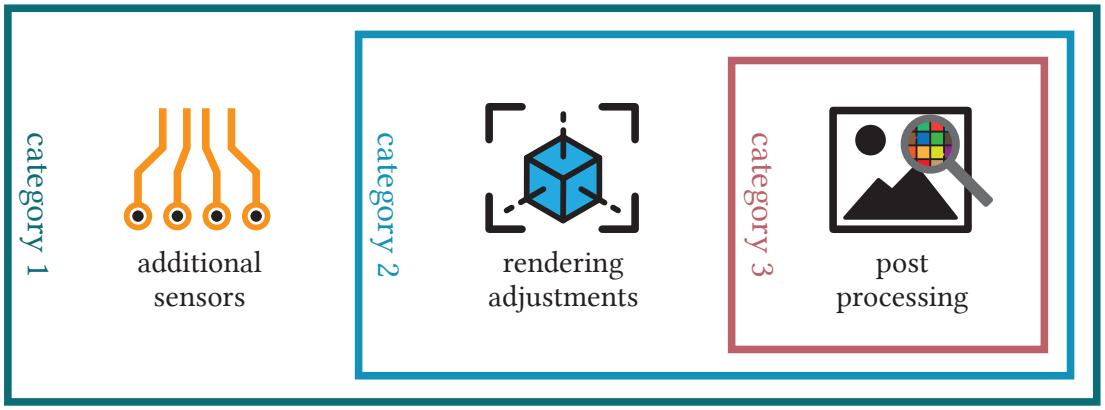

Figure 1.1: The requirements of VR applications can be classified into three categories, which define certain operations that methods, implemented for these applications, can perform. The higher categories only allow a subset of the modification possibilities of the lower categories.

in line with the visual presentation. The core challenge in this category lies in the development of effective stimulation models that are precisely adjusted to stimulate only for the desired outcome with minimal noticeability.

The second category describes applications with common graphics hardware where adjustments of the full render pipeline are possible. The focus of our methods in this category lies on pre-recorded content, where high frame rates and high quality are most challenging for 360° content. With a new way of en- and decoding animation frames based on wavelet transforms and foveation, we are able to increase the playback speed up to four times and bring it in line with rates of rendered 3D geometrical content. Implementing highly effective real-time techniques is the core challenge in this category. Compression efficiency and quality preservation are further requirements that are considered in our implementation.

The third category of applications offers the least flexibility while methods are widely applicable. Without knowledge of the 3D scene, we post-process the rendered images in screen space. Methods in this category are most generalizable due to their separation from the spatial domain. Our methods in this category exploit perceptual properties of peripheral motion perception and visual masking, as well as the effect of change blindness to achieve a significant reduction in cybersickness while preserving their noticeability at the threshold of detectability for VR users. The core challenge in this category is the missing knowledge and control of the 3D scene. Methods are fine-tuned carefully to be subtle yet effective.

Figure 1.2 shows the classification of all papers into the three categories.

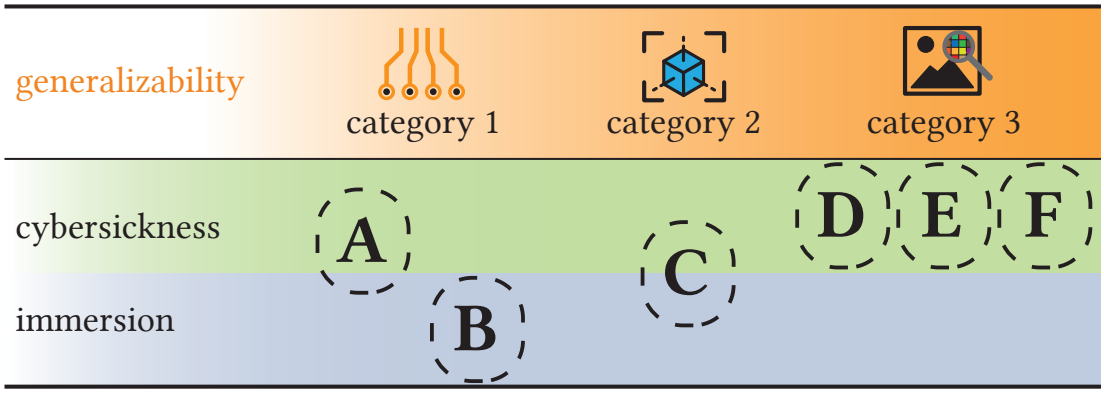

Figure 1.2: Classification of the papers of this dissertation according to the respective main factors of user experience (cybersickness and immersion). The three categories describe the requirements for the target applications to implement the respective methods of the works. Greater modification possibilities are required from applications that are in a lower category while their effect can be more impactful.

1.3 Contributions

The main work that constitutes this dissertation has been published in prestigious scientific conference proceedings or journals. In the following, a full list of all publications of this cumulative dissertation is provided. I am the single first author of all of these publications.

Paper A – Omnidirectional Galvanic Vestibular Stimulation in Virtual Reality

Cybersickness is caused by a mismatch between the visual information and vestibular signals of the balance system. Through precise stimulation of the vestibular system with weak electrical impulses, we induce an impression of self-motion which is synchronized with the visual presentation. While previous works were able to stimulate one fixed axis of rotation, we present a novel stimulation model that allows for precise stimulation of arbitrary rotational movements.

Paper B – Instant Hand Redirection in Virtual Reality Through Electrical Muscle Stimulation-Triggered Eye Blinks

Modern VR systems increasingly provide the ability to interact in a natural way, with realistic representations of the user's hands. Given this high degree of realism, mismatches

between the visual and tactile perception of a scene can quickly break the immersive experience of the user. In this work, we provide a new method to instantaneously redirect the hand movement towards the actual position of a real-world target. Our technique is based on precise stimulation of eye blinks and exploitation of the change blindness effect, and can be used in accordance with previous methods to increase efficiency and error-proofing.

Paper C – Wavelet-Based Fast Decoding of 360° Videos

For 360° videos, the playback speed for high-quality content is often limited due to the complex design of modern codecs. In this work, we propose a codec specifically designed for panoramic videos. Our implementation is based on wavelet compression and allows for viewport-dependent decoding and foveation by design. With our codec, the playback speed of videos is significantly increased to be above the refresh rates of current VR glasses and, therefore, supports more pleasant user experiences for immersive video content.

Paper D – Visual Techniques to Reduce Cybersickness in Virtual Reality

Perceptual properties of the human visual system dictate a lower spatial resolution in the peripheral region of the view. In this work, we explore simple, gaze-contingent image filters and their effect on cybersickness in a user-controlled game environment. While the experimental findings provide practical design suggestions for more pleasant experiences, they also underline the importance of the periphery for cybersickness.

Paper E – Mitigation of Cybersickness in Immersive 360° Videos

This work explores the same visual effects as publication D, but for scenarios where the user has no control over the virtual movements. Using real-world 360° videos as examples, we investigate cybersickness with both, established questionnaires and objective body signals, like EDA, heart rate, and eye movements.

Paper F – Cybersickness Reduction via Gaze-Contingent Image Deformation

In the last paper, we develop a general technique for cybersickness reduction that preserves the scene's visual fidelity. Effective and subtle mitigation of discomforting effects is achieved by content-aware deformation of the image presentation based on calibrated effects of motion perception. As part of our work, we provide a perceptual model that

describes eccentricity-based effective scaling factors of object motion at the threshold of detectability.

1.4 Overview

In this first chapter, the problem domain and approach of the thesis were introduced and an outline of the core contributions and challenges was provided.

Chapter 2 presents extensive background knowledge about the major subjects that concern this dissertation. Particularly, the background comprises the fields virtual reality, visual perception, and vestibular perception. Following, the third chapter provides an updated summary of the relevant related work for the subject.

The following main part presents the six publications that constitute the core scientific contributions of this dissertation. Papers A and B consider the first category of systems and introduce additional stimulation hardware and custom software to evoke body signals by weak electrical impulses. Both publications aim to increase the users' realism of the VR experience. In Paper A, we induce the impression of ego-motion by stimulation of the vestibular system. In Paper B, we stimulate the eyelid muscle to exploit change blindness for more realistic hand-object interactions.

The second category of systems is considered in Paper C, where we introduce a new video codec for 360° videos. With our effective software implementation, we show that wavelet-based encoding techniques are particularly reasonable for surrounding content.

Publications D - F concern application-independent solutions, where the content is manipulated by post-processing the image after the rendering is completed. Here, we investigate the effect of gaze-contingent blurring and occlusions of peripheral content on the arousal of cybersickness, for both virtual worlds (Publication D) and 360° videos (Publication E). In the last publication (Paper F), content-specific geometrical deformations are applied to the frames that reduce cybersickness by subtle reductions of visually induced self-motion.

Chapter 4 discusses valuable findings and results of all publications in a broader picture, followed by promising future research that arises from this dissertation.

2 Background

This chapter provides background knowledge on three key areas, which provides a thorough basis for the dissertation. First, I describe basic properties of the human visual system (HVS) and visual perception which are exploited in our papers to make techniques more efficient and subtle. Subsequently, the vestibular system, which plays a central role in the perception of ego-motion, is discussed in detail. The third section focuses on virtual reality, where I will clarify the most important terms and introduce the effects that are at the core of this dissertation, namely user experience, cybersickness, and interaction effects in VR.

2.1 Visual Perception

> *visual perception is [not] about seeing light; instead, it is about seeing objects, surfaces, and events in the world around us. Light provides the means of achieving this because its spatial and temporal pattern is determined by the layout of the surrounding world.*
> — Vicki Bruce (Newcastle University)

In VR, information is primarily conveyed through visual signals. The illusion of being in a different world is created due to the great importance of visual information for human self-perception. In this thesis, many of the presented techniques alter the visual output of the rendering pipeline to present a slightly different image to the eyes of the observer. However, while we aim to achieve certain effects with this modified version of the output, e.g., a reduction of cybersickness, the visual information of the image that is conveyed to the user should persist. In the techniques presented in this thesis, specific perceptual properties are exploited to develop VR methods with high efficiency. In this section, I will highlight major perceptual effects and provide extensive background information about their occurrence and mode of action.

2.1.1 Physiology of the Human Eye

In this section, a basic introduction to the physiology of the HVS is provided. Understanding the structure and basic functionality of the human visual system will be

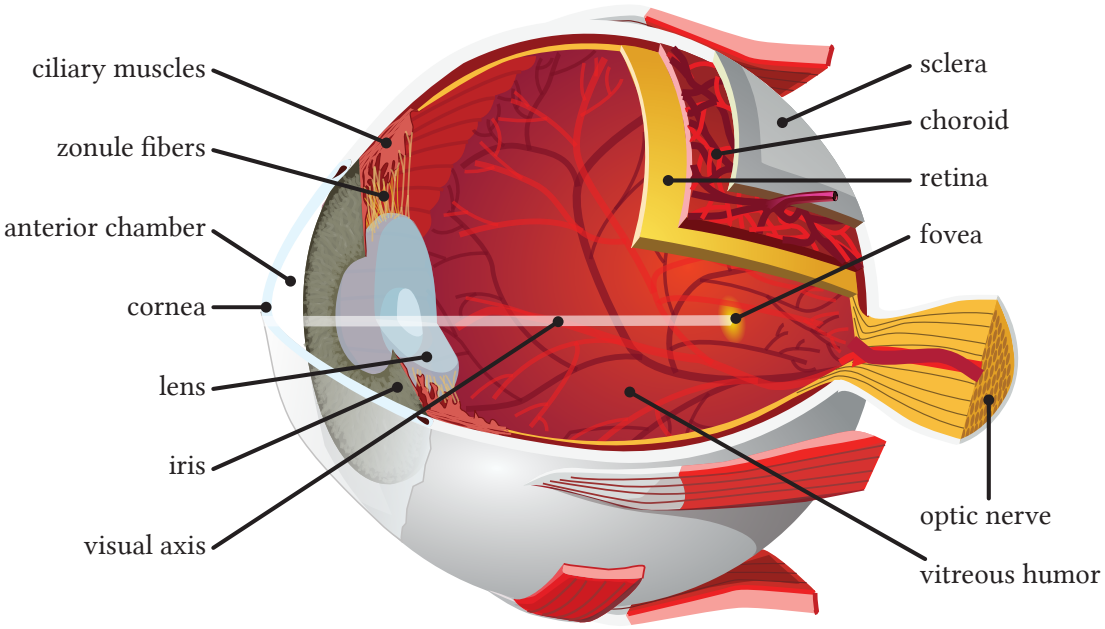

Figure 2.1: Diagram of the human eye (Originally by Chabacano, CC BY-SA 3.0, via Wikimedia Commons)

the foundation for a comprehensive understanding of the subsequent effects of visual perception.

Throughout time, light has been the most potent force to control the evolution of living organisms since the earth was formed [Land and Fernald 1992]. Consequentially, our predecessors formed a visual system to capture crucial information conveyed through light. The first versions of an eye were created already 600 million years ago, as a very simple form of a visual processing apparatus [Lamb et al. 2007]. Since then, the human visual system evolved into a complex sensory system with vast capabilities. In humans, the visual system now is the brain's premier information input and our interpretation of the world highly depends on it [Land and Fernald 1992].

The role of the eye in human perception lies in sensing visual patterns of light which are encoded into neural patterns for high-level processing in the brain [Thompson et al. 2011]. The human eye is a roughly spherical organ with an average diameter of 20mm. The hull of the eyeball is composed of three membranes: the surface consists of the transparent cornea covering the anterior surface and the sclera forming the remainder of the surface of the optic globe, the choroid layer includes blood vessels and high pigmentation to avoid backscatter, and at the innermost retinal layer the photoreceptors are located [Bartlett et al. 1965]. The curvature of the cornea dictates the optical power

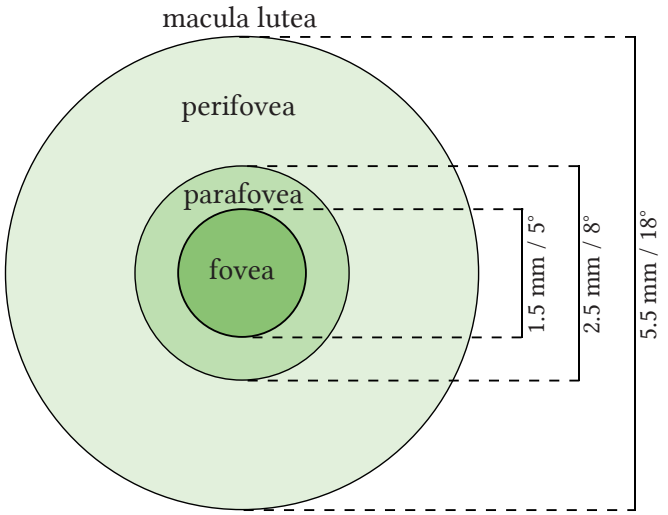

Figure 2.2: Visualization of the retinal regions [Kolb et al. 2020]. The central fovea is surrounded by the parafovea and perifovea area. The macula region describes the combined 18 degree circular area of detailed viewing.

given the high refraction coefficient with the surrounding air [Bartlett et al. 1965]. The eye's lens is positioned posterior to the cornea. Altering the shape of the lens by relaxation and contraction of the ciliary muscles allows for sharp focus of objects at different distances. The iris diaphragm controls the amount of light entering the lens [Bartlett et al. 1965]. The inside of the eye is filled with a viscous fluid, the vitreous humor, which is important for the refraction of the lens with its refractive index closer to the lens than the refraction between lens and air [Bruce et al. 2003].

The visual information captured by the retina is transmitted to the brain through the optic nerve. After initial preprocessing in the chiasma opticum, it reaches the corpus geniculatum laterale for color and contrast processing [Groß 1994]. Subsequently, the primary processing occurs in the cerebral cortex which handles high-level visual tasks like edge orientation estimation, motion detection, and complex shape analysis [Groß 1994]. In parallel, the retinal data is sent to the colliculi superiores. This section of the brain was found to be vital for the assessment of visual motor skills [Groß 1994].

The Retinal Regions

The visual field can be categorized by different circular regions that are defined by eccentricity, as illustrated in Figure 2.2. However, while these regions are defined by a fixed radius for simplified characterization, the decline in photoreceptor density and visual acuity is continuous. The macula lutea describes an approximately 18 degrees wide area

of the retina, placed around the line of sight, where detailed viewing is achieved [Purves et al. 2004]. In the center of the macula lutea lies the fovea centralis, a region about five degrees in diameter that provides the sharpest vision [Purves et al. 2004]. Surrounding the fovea, and comprising the majority of the retinal surface, is the peripheral area. This area, in contrast to the highly populated fovea, has a significantly lower density of photoreceptors, which explains its reduced capacity for detailed vision [Bruce et al. 2003].

2.1.2 Visual Acuity

Visual acuity is a fundamental aspect of human vision, central to our ability to process the detailed spatial patterns of the environment that are received by the human visual system. The light rays that stimulate the photoreceptor cells carry detailed information about the world, shaped by the nature and positioning of the surfaces from which they are reflected. The work of Gibson in 1966 highlights this concept by introducing the ambient optic array, a term that describes how the light carries information about the environment to the human visual apparatus. The structure of the optic array is determined by the nature and position of all reflective surfaces. With his work, Gibson emphasizes how light is more than a mere illuminator, but a carrier of environmental information.

The visual input that is described by the optic array presents highly resolved spatial patterns to the receptors on the retina, serving as the primary input for visual processing [Gibson 1966]. Visual acuity describes the ability of humans to detect these fine spatial patterns of the optic array in different regions of the visual field [Bruce et al. 2003]. The key measure of visual acuity is the minimal visual angle, the smallest angle of an object that can be resolved. The calculation of the angle is demonstrated in Figure 2.3. Under optimal conditions, such as ideal lighting and contrast, human visual acuity can discern patterns separated by as little as one-sixtieth of a degree, known as one minute of arc [Kolb et al. 2020]. When measuring visual acuity, usually by the method of grating, measuring the unit of visual angle allows to achieve independence from the distance of objects.

Three primary factors constrain visual acuity in humans:

Mapping Efficiency: The first factor concerns the efficiency with which spatial light patterns from the optic array are mapped onto the retina. Imperfections in the eye's focus can lead to blurred images on the retina, which causes deterioration of fine details in the spatial patterns [Bruce et al. 2003].

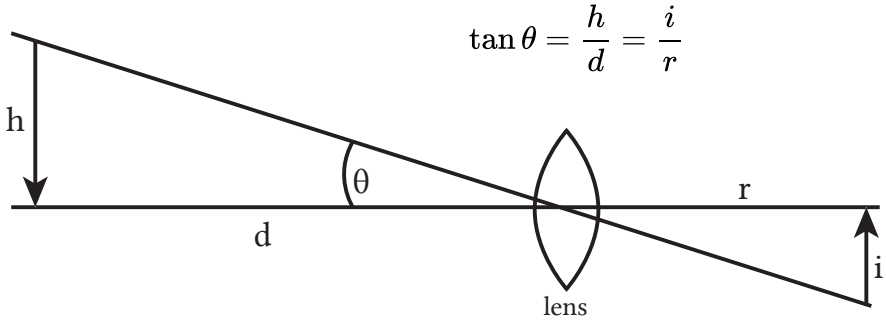

Figure 2.3: Calculation of the visual angle θ for a viewed object of height h at distance d. Here, i and r are the height and distance of the projection surface, respectively. Adapted from Bruce et al. [2003].

Receptor Cell Density: The second factor is the efficiency with which light patterns are transformed into electrical activity by the retina's receptor cells. The efficiency is mainly driven by two factors, the density and pooling of photoreceptors [Bruce et al. 2003]. A higher density allows for more detailed patterns of electrical activity, translating into finer visual resolution. Likewise, the fewer receptors are pooled, the higher the spatial details that can be resolved. This pooling is realized in the nerve cell layers connected to the rods and cones [Bruce et al. 2003]. In the central visual area, the direct connection of foveal receptors to only one or two bipolar cells allows the ganglion axons, which carry visual information over the optic nerve to the visual cortex, to carry information at the finest pattern level of receptor excitement [Bruce et al. 2003].

Neural Processing Efficiency: Finally, the third factor is the efficiency of the neural apparatus to process visual information from the patterns of electrical activity. This complex, high-level process integrates and interprets the electrical signals into coherent visual information. [Bartlett et al. 1965]

Beyond these fundamental factors, visual acuity is subject to various external and internal influences. Age [Young 1987, Ruan et al. 2021], health [Dandona and Dandona 2006], neurological factors [Dandona and Dandona 2006], and environmental conditions [Tognini et al. 2012] can all impact acuity. For instance, age-related macular degeneration significantly reduces the density of receptor cells, directly impacting the second factor of visual acuity [Ruan et al. 2021].

2.1.3 Foveation

In the human retina, the distribution and packing of photoreceptor cells is highly specialized and follows a non-uniform pattern [Leigh and Zee 2015]. Foveation describes this decline in visual acuity starting from the most central part of the retina, the fovea centralis, to the outer part, the periphery, that is more distant from the gaze point [Silverstein 2008]. Foveation is the result of the pooling of information in the periphery as well as the uneven distribution of different photoreceptor types. Thereby, the visual acuity decreases at roughly a logarithmic rate [Kolb et al. 2020].

The retinal distribution of the two primary photoreceptor types, rods, and cones, are described in the work of Bruce and Green [2003]. With an approximate number of 120 million on the surface of the retina, rods are much more prominent than the 7 million cones. Cones, responsible for color vision and high spatial acuity, are densely packed in the fovea, the central region of the retina. The concentration of cones in the fovea enables detection of fine details of the optic array which is necessary for tasks like reading and recognizing faces. Outside the fovea, cone density decreases rapidly, while rod density increases. At around 10 degrees from the center, there are almost no cones anymore. Rods dominate the peripheral retina, the region outside the fovea more distant from the central vision. In the center of the fovea, almost no rods exist. The uneven distribution of rods and cones achieves a significantly higher acuity compared to a uniform distribution. With the specified arrangement of cone and rod photoreceptors, visual perception can resolve spatial details under various lighting conditions.

Understanding foveation is crucial for designing effective computer graphics techniques due to its direct impact on the perception of the output. Exploiting foveation allows the development of foveated renderings, where graphics are rendered in high detail only in the areas where the viewer's acuity is equally high, while details are spared in the peripheral field of vision. The enrichment of techniques with foveation allows for more computational efficiency, while the perceived quality is kept constant [Patney et al. 2016].

2.1.4 Sensitivity

In the former section, the visual acuity of the human eye was described to be at its peak only in the most central part of the retina, with a fast decline in acuity towards the outsides of the retinal image. When considering the design of the eye and the distribution of photoreceptors, the obvious question arises as to why acuity is foveated and not at an equally high level over the entire field of view. One answer to this question

is sensitivity. Naturally, the design of the human eye is a trade-off between sensitivity and visual acuity.

In the human visual system, the central fovea region, with its high concentration of cone receptors, only applies minimal pooling of receptor outputs [Bruce et al. 2003]. With these properties, the central vision at the gaze point allows for high acuity as discussed in Section 2.1.2. In contrast, the larger peripheral area is densely packed with rod receptors and allows for higher sensitivity [Bruce et al. 2003]. For this, the outputs of these peripheral receptors are extensively pooled to optimize the visual system for low-light conditions.

The unique structure of rod receptors makes them particularly suited for low-light vision. Each rod contains a deep stack of pigment-filled layers of a folded membrane in its outer segment, which enhances the probability of photon absorption [Mustafi et al. 2009]. Consequently, rods are activated by light intensities too low to stimulate cone receptors.

In general, light intensity is a crucial factor in the detectability of fine spatial patterns. The rate at which photons hit the receptors varies, even under constant lighting conditions, what is referred to as photon shot noise. Thereby, the variance in photon flux is proportional to the square root of the mean photon count [Bruce et al. 2003]. This means that under bright conditions, the difference in photon flux between two adjacent receptors is high while the variance in photon flux is low. In contrast, with low lighting, the difference between photon counts of adjacent receptors becomes irrelevant since, in comparison, the noise is considerably higher. In order for spatial patterns to become visible in these conditions, the information of the arriving photons have to be averaged over a longer period of time, or by the pooling of multiple receptors.

The movement of objects across the retina further complicates this dynamic. If an object moves too quickly, the necessary threshold time for receptor activation might never be reached, and spatial patterns in the optic array stay undetected. Consequently, the maximum detectable speed of a contrast difference moving over the optic array decreases with decreasing light intensity.

Sensitivity, the ability of the eye's receptor cells to detect light intensity variations, is inversely related to acuity [Bruce et al. 2003]. Enhancements in sensitivity, achievable through larger receptor cell cross-sectional areas or pooling of receptor outputs, result in diminished acuity. This trade-off is a fundamental aspect of the human visual system, balancing sensitivity and resolution according to environmental and physiological needs.

In computer graphics, the sensitivity of visual details in the spatiotemporal domain based on content contrast is described by the contrast sensitivity function (CSF). Per-

ceptual models usually consider a diverse set of parameters like luminance, eccentricity, and content size [Mantiuk et al. 2022, Robson 1966, Tursun and Didyk 2022]. Considering the sensitivity of human perception to visual patterns allows one to discard the rendering of fine details that are below the threshold of visibility, while attention is being steered towards contrast-rich content with high importance. Visual techniques that consider sensitivity can drive attention, enhance virtual experiences and increase computational efficiency.

2.1.5 Motion Perception

Visual motion is a fundamental aspect of perception, driven by movements within the environment that alter the spatial pattern of the optical array. These alterations form spatiotemporal patterns that are rich in information and convey the direction, speed, and type of object movements [Bruce et al. 2003]. Motion perception is a ubiquitous element of visual experiences due to the constant shifts in the retinal image caused by the movements of both eyes, through translation and rotation, and object movement in the environment [Thompson et al. 2011].

Visual motion is a primary cue in human perception because of the information density of the temporally evolving patterns in the optical array [Thompson et al. 2011]. This prominence is attributed to the comparably low ambiguity in motion information as opposed to other perceptual cues. Consequently, motion is a key property in directing attention.

A high fraction of the cells in the visual cortex are tuned to detect visual motion [Thompson et al. 2011], an evolutionary adaptation to manage the richness of motion information. These cells respond to various sources of visual motion, including the relative movement between distinct visible structures and the positional changes of the eyes.

Similar to the processing of static images, motion perception involves extracting information from the optical array. This information is then utilized to deduce the properties of the visual objects, as well as to understand the observer's relationship with the surrounding environment.

Sensing of Visual Motion

Temporal changes in the optical array, caused by relative motion between the eye and the three-dimensional environment, result in alterations of the two-dimensional image received on the retina. The ability to detect motion in a given pattern depends on several

parameters: speed, direction, pattern size, contrast, and eccentricity [Thompson et al. 2011]. The relationship between these parameters for the detectability of motion can be modeled with mathematical functions like the CSF [Mantiuk et al. 2022, Tursun and Didyk 2022]. In general, the smallest size of alternating patterns that can be detected by the average human observer lies at a 1/50th of a degree [Thompson et al. 2011].

Motion detection is inherently complex due to the many influencing factors. It is further complicated by the influence of contrast effects. Similar to brightness contrasts, the perception of motion is affected by conflicting optical flows between two moving patterns [Thompson et al. 2011]. For example, a small pattern moving against an inversely colored homogeneous background requires a speed of at least 0.2 degrees per second to be detectable [Palmer 1999]; when the same pattern is now displayed on a textured background, the required speed drops to just one-tenth of this value.

Motion perception is facilitated by two sensory processes; the image-retina system, which detects changes in the retinal image due to object movements or ego-motion, and the eye-head system, responsible for processing information about eye position and rotation [Gregory 2015]. These processes do not function in isolation but work in unison to enable accurate perception of relative movement and to ensure a comprehensive understanding of motion in the viewer's field of view (FOV).

Motion Characterization

As suggested by Gibson [1950], the perception of the retinal image over time can be understood when examining the two-dimensional vector field that is defined by the projection of the three-dimensional motion of all visible points in the environment. This concept is known as the motion field, while the related term optical flow refers to apparent motion, differing primarily in cases of optical illusions.

The spatiotemporal image on the retina can be represented as a three-dimensional function of two spatial components and a time component (see Figure 2.4). In visual representations of these functions, lines parallel to the time axis indicate patterns that are static relative to the observer, while diagonal lines denote movement [Adelson and Bergen 1985, Bolles et al. 1987].

The HVS's ability to detect moving patterns through space-time filtering is illustrated by perceptual models that operate in the previously defined spatiotemporal space [Adelson and Bergen 1985, Fleet and Langley 1994, Watson and Ahumada 1985]. Based on these models, space-time filtering characterizes moving content by the distribution of size and orientation of lines in the time dimension. In other words, the eye's motion perception is primarily driven by the magnitude and direction of optical flow.

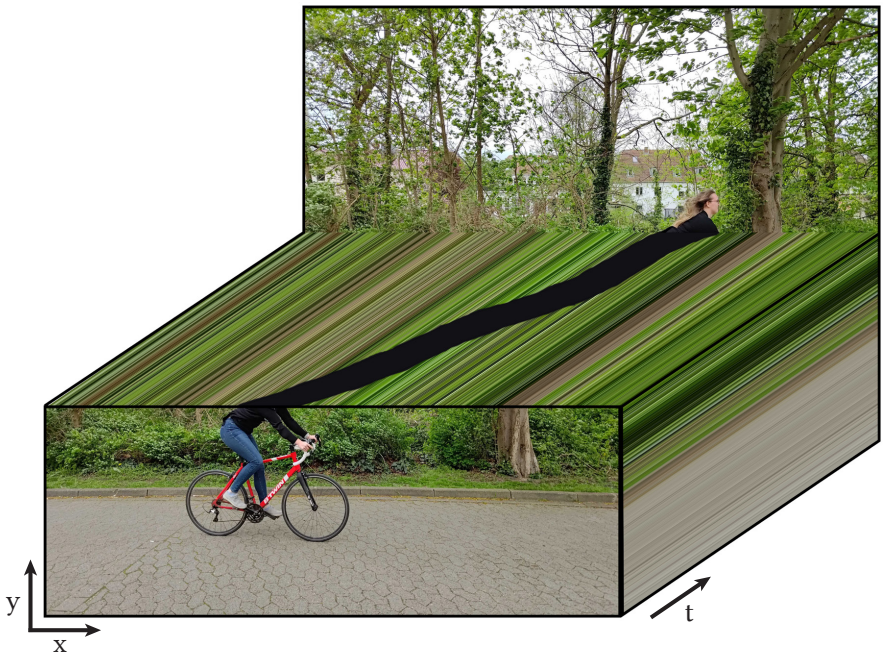

Figure 2.4: Spatiotemporal representation of an image sequence with a static camera. The temporal evolution shows the middle line of the image sequence with the full frame only on the back. The lines that are parallel in time characterize static objects, while the moving cyclist draws a diagonal line to the space-time volume.

Second-order motion refers to motion that is detected from patterns that do not move directly but vary in contrast, brightness, or other display effects [Lu and Sperling 2001]. A popular example to demonstrate this effect is a Gabor patch fixed in position but shifted in phase over time. In real-world scenarios, second-order motion patterns often arise from transparency and occlusion, making visual motion critical for the detection of occlusion boundaries [Hegdé et al. 2004].

The characterization of motion is not solely based on the temporal profile of a pattern on the retinal image, but rather motion is interpreted relative to its surrounding content [Thompson et al. 2011]. For example, a small pattern that is moving against the direction of its background is more easily detected and appears faster compared to a motionless background. This effect of context-related motion can even make static patches appear to be in motion if the surrounding background is moving. Thus, visual perception is considerably more sensitive to relative changes than to absolute values. Relying on relative changes is simpler for a visual system because the detection of absolute magnitudes would require a zero reference and adequate scaling, two properties that are hard to achieve by a natural system. The perception of motion is further in-

fluenced by context [Thompson et al. 2011]. When individual entities are perceived as part of a combined configuration, their movement is interpreted relative to this configuration, while their absolute movement on the retinal image is disregarded. As with all human senses, motion perception is influenced by sensory adaptation [Addams 1834]. The waterfall illusion demonstrates this phenomenon: When staring at a waterfall for a prolonged period of time and shifting the gaze to a static wall afterward creates the illusion of the static wall moving in the opposite direction. The concept of sensory adaptation is explained in more detail in Section 2.1.10.

Local Ambiguity

The aperture problem in visual perception refers to an ambiguity that occurs when observing lines without visible endpoints [Nakayama and Silverman 1988]. The problem arises from the inability of the moving component to induce spatiotemporal changes in the retinal image. An example is the illusion observed with rotating barber poles, where stripes seem to ascend rather than rotate.

Initial processing of visual information is first executed in local regions of the retinal image, which can introduce challenges similar to the aperture problem [Thompson et al. 2011]. However, during further processing, the visual system combines local information to reach a consensus for an estimate on the global two-dimensional motion. This ability relies on the premise that motion within most regions of the FOV only varies minimally [Thompson et al. 2011]. Based on the global consensus, the HVS infers local motion and typically can resolves the ambiguities.

Exceptions to the assumption of minimal variation in movement strength occur at occlusion boundaries, where different sides of the boundary may display significantly varying flow values. Similarly, transparency can present a comparable challenge. The mechanism by which the visual system operates in these situations remains partially unclear. It is hypothesized that the system might selectively combine flow values sharing similar properties like contrast or spatial frequency [Adelson and Movshon 1982].

The minimum temporal changes detectable by the visual system are constrained by multiple factors including contrast and spatial frequency. The general sensitivity threshold lies at a temporal frequency of about 60Hz [Mantiuk et al. 2022]. Changes occurring at faster temporal rates result in a phenomenon known as flicker fusion [Thompson et al. 2011]. In this state, the visual system will not perceive temporal changes anymore, but a static image if no spatial changes occur simultaneously.

Background

Apparent Motion

Apparent motion is a phenomenon where a sequence of discrete animation frames creates the illusion of continuous motion, despite its inherently discontinuous nature [Thompson et al. 2011]. This effect is critical to the success of modern computer graphics. The perception of real movement is significantly enhanced when the direction of spatial displacement of patterns across frames is continuous [Sperling 1976]. The visual system employs distinct processes for short and long-range motion perception. These processes are differentiated by the distance between cues necessary to detect motion. In addition to the distance factor, more rapid temporal frequencies are required for short-range apparent motion [Braddick 1974].

Converting continuous motion into discrete animation frames introduces high frequencies in both time and spatial domains. However, if this added high-frequency information falls outside the limits of perceivability, the two animations will appear the same. Watson et al. [1986] define this interval where differences in time or space are still perceivable as the *window of visibility*. Two strategies allow to move the additional high frequencies in space-time outside of the window of visibility: either an increase of the frame rate to surpass the perception threshold, inducing flicker fusion, or applying a low-pass filter over time to reduce spatial frequencies in the direction of motion, creating motion blur. When two visual motion stimuli pass through the window of visibility with the same space-time frequency spectrum, they are perceived as identical [Watson et al. 1986].

Vection

Vection describes the illusion of self-motion induced by visual stimuli in situations where the viewer is physically stationary [Andersen 1986, Dichgans and Brandt 1978]. A classic example is the experience of sitting in a stationary train and watching another train outside the window begin to move. This situation often creates the sensation that the observer's own train is in motion, while the other train appears static. In this illusion, relative motion is correctly perceived, but the HVS fails to identify absolute motion about the world accurately.

The phenomenon of vection can manifest itself in both linear and rotational movements, which defines the categories of linear vection and circular vection [Thompson et al. 2011]. An example to induce circular vection is an experiment where observers are placed inside a rotating curtain with vertical strips. From the visual input, the observers get tricked into the impression of rotating themselves. Since vection is perceived

as real motion, it can trigger motion sickness just as effectively as actual motion.

Vection predominantly relies on visible motion in the peripheral vision [Brandt et al. 1973]. However, the strength of the vection effect is also influenced by spatial frequency, with high spatial frequencies in the foveal vision enhancing vection and, in contrast, low spatial frequencies promoting the effect in the periphery. This difference in effect intensity is thought to arise from different processing mechanisms for foveal and peripheral information [Brandt et al. 1973].

Vection holds significant importance in computer graphics, particularly in enhancing VR. Through vection, realistic and immersive virtual experiences are fostered by improving the spatial awareness of the user. Understanding the phenomenon is key for addressing cybersickness, which is evoked by the conflict of the visually-induced motion effect of vection and the physical sensation.

2.1.6 Eye Movements

The former section mostly addresses visual motion that is attributed to external motion in the optic array. However, our eyes are not static and motion is also contributed by movements of the eyes themselves. Eye movements are driven by a complex interplay of neurological and muscular systems. They enable the human visual system to gather detailed information of the optic array even without head movements. The primary types of eye movement are discussed in the following: visual fixations, saccades, smooth pursuit, and the vestibulo-ocular reflex.

Fixations

In visual perception, eye fixation is a fundamental feature to allow for accurate vision. Contrary to the assumption that the eye remains fully static when fixating on a motionless target, it performs subtle, involuntary movements. These movements, primarily microsaccades, play a crucial role in maintaining the visibility of the target [Martinez-Conde et al. 2009]. Microsaccades are small, rapid eye movements that occur during fixation, preventing the gaze from fading off the target over time. In addition to microsaccades, the eye also exhibits smooth drift movements during visual fixations (see Figure 2.5). Eye drift is most likely to occur following microsaccades, or even full saccades [Leigh and Zee 2015].

The retinal image is subject to a dynamic process underlined by these micromovements. Movements of the image stimulate visual tracking mechanisms that act in the manner of a negative feedback system to bring the gaze back onto the target [Leigh

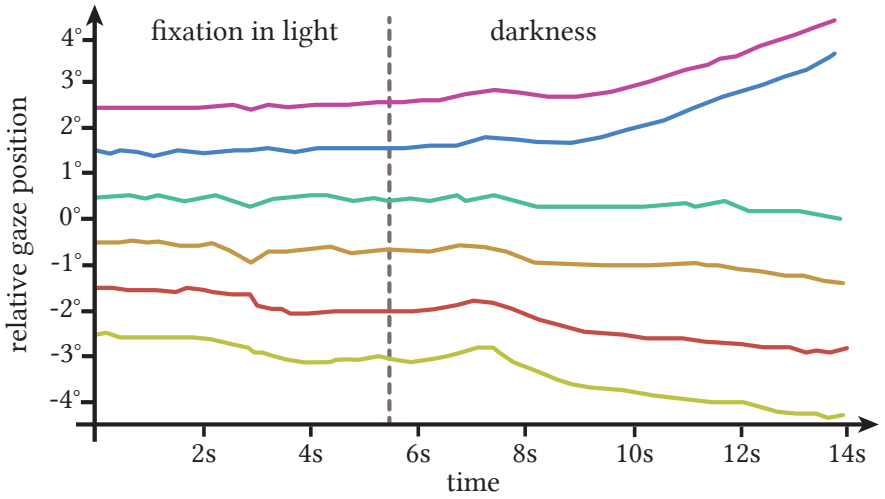

Figure 2.5: Simplified illustration of the gaze stability during fixations based on the results of Leigh and Zee [2015]. Participants were viewing a fixation cross at 1.2 m distance. When the light was turned off the gaze of participants drifted away (darkness section). The direction and strength were found to depend on the starting position of the cross which is indicated by the relative gaze position at the start.

and Zee 2015]. This continuous cycle of drift and readjustment is essential to counteract sensory adaptation (see Section 2.1.10) and the consequential decline in the visibility of static retinal images. Without such movements, "perfect" fixations would lead to temporary blindness due to sensory adaptation.

The stability of visual fixations and correct feedback is heavily dependent on visual marks. When the visual cue of the target is removed, for example by switching off the light, the stability of the fixation decreases and the eye rapidly drifts away from the target [Leigh and Zee 2015].

Saccades

Saccadic eye movements are rapid, brief, and precise shifts in gaze direction executed to reposition the image on the retina [Leigh and Zee 2015]. These movements can be categorized into two types: quick phase movements and voluntary saccades.

Quick phase movements are automatic responses that occur during sustained head rotations. They prevent the eye from reaching the orbital edge by correcting the eye position in the direction of the head rotation. Besides acting as a resetting mechanism, quick phases also provide a preview of the upcoming scene [Schmid and Lardini 1976]. Quick phases act fast and can reach a maximum velocity of 500 degrees per

second [Garbutt et al. 2003].

With the evolution of foveated vision, which concentrates visual acuity in a small retinal area, the ability to redirect gaze to explore the environment independently of head movements became essential. Voluntary saccades represent this type of saccadic movement that facilitates scanning of visual information within the FOV. These saccades can be initiated under various conditions, such as the appearance of new objects, motion, flicker, and even auditory stimuli [Leigh and Zee 2015]. The latency of voluntary saccades is typically around 200 milliseconds from the onset of a stimulus [Leigh and Zee 2015]. This delay is attributed to the time required for neural processing of the stimulus, both in the retina and the brain.

Smooth Pursuit

Smooth pursuit eye movements enable the maintenance of an almost static retinal image for objects that are in smooth motion within the FOV [Komogortsev and Karpov 2013]. While saccadic movements allow to bring a region of interest to the high acuity foveal vision, smooth pursuit is necessary to continuously keep track of a moving target.

The importance of smooth pursuit becomes evident when considering the rapid decline in visual acuity away from the macula (cf. Section 2.1.1). Without smooth pursuit, tracking of objects would result in a significant reduction in the perceptual quality of spatial details as the object would not be consistently maintained in foveal vision.

Ideal pursuit movements would instantly adjust to the speed changes of the target. However, such immediate action is hindered by the delay of the visual processing, with approximately 80–120 milliseconds, being way too slow for direct response [Leigh and Zee 2015]. To counteract the delay, the HVS employs predictive smooth pursuit movements, as visualized in Figure 2.6. Predictive processes in human vision can initiate pursuit movements based on the anticipation of target motion [Barnes 2008]. Even when the target temporarily gets out of sight, such as during occlusions, the predictive pursuit movements can reliably follow the target.

Smooth pursuit movements likely evolved to sustain fixation on targets during self-motion [Miles 1998]. The efficiency of smooth pursuit, however, decreases with age, visual impairments like cerebellar disease, and the influence of drugs affecting the nervous system [Miles 1998].

Moreover, the input signals for smooth pursuit are subject to considerable noise, which makes its control less precise. To enhance performance, the visual cortex appears to employ Bayesian filtering techniques to handle imprecise visual cues [Yang et al.

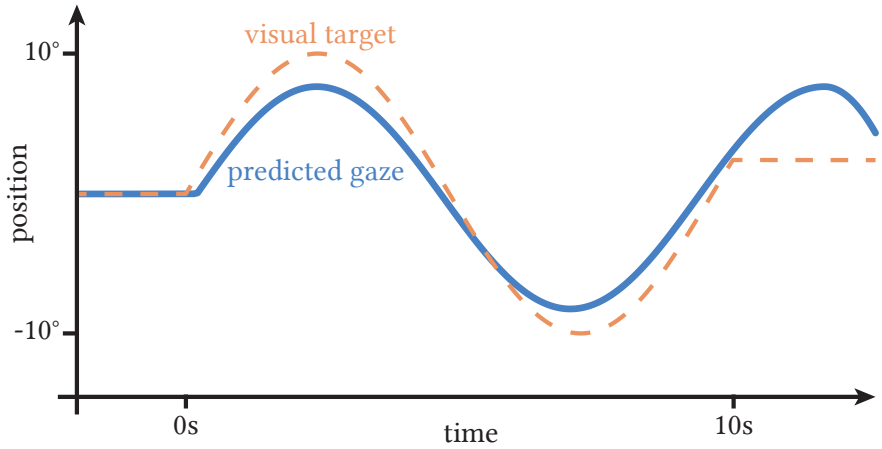

Figure 2.6: Simplified illustration of predictive pursuit movements based on the results of Leigh and Zee [2015]. The smooth tracking of the gaze movement (blue line) first lacks behind the visual target that is followed (dotted orange line). However, after the first peak of the sine wave, the smooth tracking starts to lead the target which indicates predictive movement tracking. After the target stops abruptly after 10 seconds the eye continues the expected sinusoidal motion for a short time, indicating a memory of prior motion.

2012]. These filtering processes allow for more accurate tracking despite the inherent noise and imprecision in the visual input signals.

Vestibulo-Ocular Reflex

The vestibulo-ocular reflex allows humans to maintain steady vision during head movements [Viirre et al. 1986]. The effect is based on signals from the vestibular system which assists in the stabilization of the retinal image during brief head rotations and translations.

Eye movements triggered by vestibular signals, generated based on the acceleration data of the balance sensors in the inner ear, are notably faster than those based on visual information. The difference in response time is crucial for fast response and compensation of head rotations with the vestibulo-ocular reflex is performed with latencies of less than 15 milliseconds [Halmagyi et al. 1990, Ramat and Zee 2005]. In contrast, movements that are guided by visual signals typically have a latency of more than 70 milliseconds [Leigh and Zee 2015]. The vestibulo-ocular reflex operates independently of visually mediated eye movements.

There are two distinct types of the vestibulo-ocular reflex, for rotational and translational self-motion signals [Liao et al. 2009]. The translational vestibulo-ocular reflex

is particularly relevant when viewing objects in close proximity.

It is important to note that even during pure head rotations, the eyes experience slight translation since they are not located precisely at the center of rotation. When focusing on near objects, this additional translation necessitates increased eye rotation to maintain the focus and account for the additional translation [Han et al. 2001, Viirre et al. 1986].

2.1.7 Change Blindness

Change blindness is a phenomenon in human visual perception where significant changes in a visual scene go undetected when accompanied by a visual disruption, such as during eye blinks or saccadic movements. While this effect occurs naturally based on HVS limitations in attention and memory, it can also be artificially induced.

In the work of Rensink et al. [Rensink et al. 1997], the flicker paradigm was used to show that when an altered image is presented with an 80ms gray screen break in-between, the immediate detection of changes gets significantly disrupted. Similarly, in virtual reality environments, changes that are made while the viewport is briefly moved away often remain undetected when the object is in sight again [Martin et al. 2023].

Change blindness further extends to scenarios where the observer's attention is focused elsewhere in the scene. This effect has been observed in both video settings [Levin and Simons 1997] and real-world situations [Simons and Levin 1998b]. As one explanation for change blindness, the limited capacity of the visual memory to retain details of a scene is put forward [O'regan and Noë 2001, Rensink 2002, Horowitz and Wolfe 1998]. This argumentation suggests that the overall meaning of a scene can be preserved. However, specific visual details are lost when attention is disrupted. This explanation follows the assumption that the visual memory is transient.

However, other studies, such as those by Hollingworth et al. [2008] and Mitroff et al. [2004] reject this assumption and aim to demonstrate that robust visual memories for objects can persist under certain conditions. For example, in the studies of Hollingworth et al., an observer's gaze was redirected to the location of a change after it occurred, suggesting the presence of some form of persistent visual memory.

These newer findings suggest that the issue in change blindness may not be the absence of a visual representation in memory, but rather the difficulty in comparing this stored visual information with the post-change scene. Visual perception and memory are complex and while robust memories may exist, their accessibility and utility in detecting changes depend on attentional focus and the context of the visual scene.

Change blindness presents unique opportunities in computer graphics applications.

Background

Strategically updating areas in a scene where the user is most likely not to notice changes allows to efficiently allocate resources to enhance the performance without compromising perceptual quality. Furthermore, in interactive media like VR, change blindness allows for dynamic alterations of the scene without disrupting the user experience.

2.1.8 Binocularity and Depth Perception

Binocular vision, achieved through the frontal position of the eyes and their overlapping FOV, enhances visual perception with depth cues and spatial relationships in the environment. The overlapping segment of the visual field that is seen by both eyes is called the binocular field [Bruce et al. 2003]. Stereoscopic vision significantly enhances the accuracy in processing spatial and temporal patterns [Jones and Lee 1981]. A fundamental function of stereoscopic vision is the ability to discern close-by object distances more precisely.

Each eye possesses a FOV of approximately 160 degrees horizontally and 135 degrees vertically. The combined horizontal FOV of both eyes is approximately 200 degrees, with a 120-degree overlapping region due to binocularity [Thompson et al. 2011, Wandell 1995]. This wide FOV is advantageous for a variety of reasons: it facilitates the perception of a greater amount of environmental information, simplifies the estimation of self-motion, and improves the assessment of geometric perspective [Thompson et al. 2011].

However, an expansive FOV also imposes high demands on the visual processing system. Complex attentional mechanisms and visual search strategies are necessary to effectively process and interpret the detailed information available in a vast optic array [Thompson et al. 2011].

Accommodation

The human eye's ability to focus on objects at different distances is primarily determined by the combined properties of the lens-cornea system and its distance from the retina, comparable to a camera lens [Bruce et al. 2003]. The optics arrangement defines a focus plane in space, where objects at a specific distance from the eye are sharply focused on the retina. Despite this discrete plane, objects that are slightly closer or farther from the optimal focus distance can still appear sharp, forming a distance range that is referred to as *depth of field* [Thompson et al. 2011]. The phenomenon of the depth of field is determined by physiological factors such as the resolution limit of retinal receptors and ambient lighting conditions. The depth of field varies significantly depending on

the focal point. For instance, when focusing at infinity, the depth of field spans from infinity to about six meters [Bruce et al. 2003].

Accommodation is the process through which the human eye adjusts its focus for objects at different distances [Bruce et al. 2003]. This adjustment involves the contraction and relaxation of the ciliary muscles attached to the lens [Bruce et al. 2003]. When the muscles contract, the lens thickens, increasing its curvature to focus on near objects. Conversely, when the ciliaries relax, the lens flattens and its curvature gets reduced, allowing the eye to focus on objects at infinity.

However, even under optimal focusing conditions, a certain degree of blur persists due to optical imperfections in the eye's lens system. These imperfections, inherent in the human optical system, mean that perfect focus is never truly achieved, but constrained by the complex and adaptive nature of the HVS.

Vergence

During evolution, the HVS has adapted to binocularity and foveation, which necessitates the ability to direct the line of sight of both eyes towards a common point of interest. Vergence eye movements allow the alignment of both eyes to a common focal point by simultaneous movement of both eyeballs in opposite directions [Leigh and Zee 2015]. In comparison, during versional movements, from smooth pursuit or the vestibulo-ocular reflex, both eyes move in the same direction.

Vergence movements are primarily driven by two factors, image disparity and image blur [Leigh and Zee 2015]. Disparity refers to the difference in the two-dimensional location of a target on the retinal images of the two eyes. When disparity is detected, the visual system initiates fusional vergence movements to align the images to create a single, coherent view [Leigh and Zee 2015]. On the other hand, when focus decreases, indicated by increased image blur, the visual system stimulates accommodative vergence movements. These movements are designed to regain clear vision of the target. Accommodative vergence movements occur along with the accommodation of the lens and pupillary constriction [Leigh and Zee 2015].

Intriguingly, vergence movements can be stimulated even if one eye is covered and the other eye accommodates linearly to a different focal plane [Leigh and Zee 2015]. Furthermore, during visual search tasks, vergence movements are typically accompanied by saccades [Leigh and Zee 2015]. This combination is necessary because objects in a visual scene often differ in both position and depth, which requires the eyes to adjust in both dimensions simultaneously.

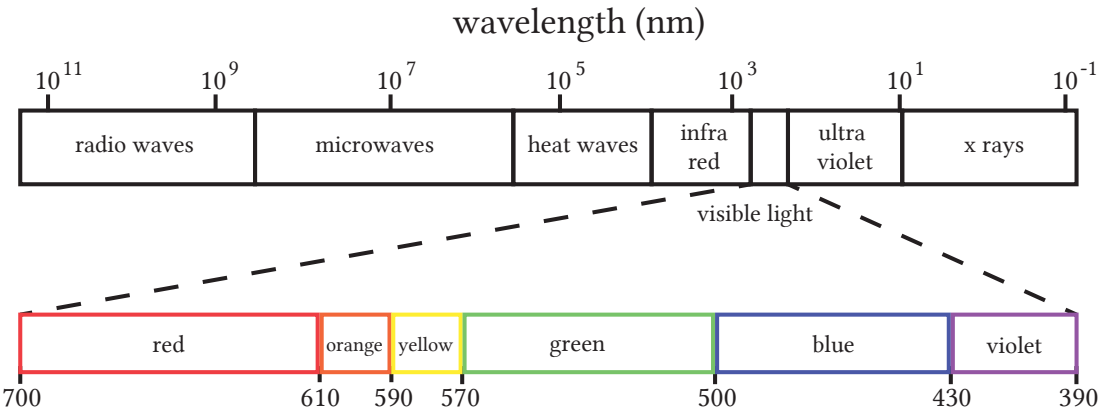

Figure 2.7: Illustration of the color distribution over different wavelengths (following [Bruce et al. 2003]). The individual color sections are not discrete, like in this abstract presentation, but merge seamlessly into one another. As demonstrated, visible light makes up only a small part of the existing wavelengths.

2.1.9 Color Perception

Light, in essence, can be thought of as a stream of photon particles that travel at the speed of light, approximately 300,000 kilometers per second [Conference Générale des Poids et Mesures 1983]. These photons transfer their energy when being absorbed by the photoreceptors in the retina. Monochromatic light consists of a single wavelength and corresponds to a specific color when perceived by the human eye (see Figure 2.7).

The human eye not only detects light intensities but also colors with the retinal detectors. The spectral composition of light conveys color information about the surface from which the light is reflected. In some cases, certain objects that might be challenging to detect based on light intensity alone become clearly visible with color perception, such as red fruits in a dense forest of leaves.

Color vision requires different types of receptor cells with pigments that have different absorption spectra [Bruce et al. 2003]. Rods always contain the same type of pigment and do not contribute to color vision. The cones, however, exist as three types of receptors, the S, M, and L cones, each with different pigments [Bruce et al. 2003]. This characteristic is named trichromatic vision, which is an ability only shared with apes. The peak absorption wavelengths for the three pigments are 419 nm (S cones), 531 nm (M cones), and 558 nm (L cones) [Bruce et al. 2003]. In dark conditions where cones are inactive and vision relies solely on rods, humans cannot perceive color.

Although theoretically, the number of different spectral compositions of light reaching the eye is infinite, humans can only distinguish a few thousand different colors [Bruce

et al. 2003]. This limitation arises because any light composition that generates the same response in the three cone types will be perceived as the same color. Hence, any color can be replicated by a specific combination of intensities of three monochromatic light rays.

From an evolutionary perspective, trichromacy provides an advantage, especially in food search, over animals with only two color receptors [Osorio and Vorobyev 1996]. Unfortunately, about 8% of humans, predominantly men, are affected by color perception deficiencies [Bruce et al. 2003]. A common form of color deficiency, commonly referred to as color blindness, occurs in individuals lacking either the M or L cones, resulting in dichromatic vision. This condition primarily impairs the ability to distinguish colors with red or green components.

2.1.10 Sensory Adaptation

Sensory adaptation is a phenomenon observed across all sensory systems, characterized by the nervous system's adjustment to constant stimuli [Thompson et al. 2011]. This adaptation in perception manifests itself as a reduction in response to continuous stimulation. For instance, the initial sensation of extreme temperature when stepping into a hot bath or cold pool diminishes over time as the body's temperature sensors adjust to the new condition, undergoing sensory adaptation.

In the brain, sensory adaptation involves the selective activation of specific neurons, which adjust their firing rate in response to stimulus features over time [Thompson et al. 2011]. After stimulation stops, these neurons continue to fire at the adapted rate for a brief period. The result is an aftereffect that mirrors the original visual features.

Furthermore, when multiple sensory systems conflict, human perception can also adapt to discrepancies between senses. This high-level change in perception over time is called perceptual adaptation [Thompson et al. 2011]. Cybersickness, an example of perceptual adaptation in VR, is caused by a conflict between visual and vestibular sensory information. With repeated exposure, users typically adapt to the VR experience and the intensity of the cybersickness symptoms decline slowly.

2.2 The Vestibular System

The vestibular system is a sensory organ located in the inner ear, casually referred to as the balance system [Dobie 2019]. As part of the inner ear labyrinth, the main function of the vestibular system is to maintain orientation in space and to register accelerations of the head, both lateral and circular [Baloh and Kerber 2011]. The vestibular system consists of three semicircular canals and the otolithic organs, as illustrated in Figure 2.8, which are mirrored in pairs on each side of the head. The semicircular canals detect rotational movements, while the otoliths perceive linear accelerations [Klinke 1993, Birbaumer and Schmidt 2010]. The inside of the organ is filled with a fluid, called the endolymph [Klinke 1993].

As a human sense for the perception of self-motion, the vestibular system is indispensable for the interpretation of the surrounding world and our ego-motion in it. The conception of what is real arises by combining all sensory impressions. Since VR systems aim to immerse the user in an alternative virtual world, the full acceptance of this simulated reality requires coinciding information from all sensory organs. The human visual perception can be successfully deceived by the presentation of the displayed content in VR glasses. The vestibular system, however, continues to perceive the signals of the real world. This discrepancy not only hinders the user to fully immersing in the virtual world but can also result in negative physiological effects as a reaction of the body to the sensory mismatch. Cybersickness is a natural defensive mechanism of the body against inconsistent sensory signals and is common in moving VR simulations (compare Section 2.3.5). The efficient mitigation of cybersickness requires a fundamental understanding of the functioning of the vestibular system, which is to be established in this chapter.

2.2.1 Sensation of Rotation

Rotational movements of the head are detected by the semicircular canals. In each inner ear labyrinth, there are three semicircular canal organs arranged approximately orthogonally to each other. These canals are responsible for perceiving rotational accelerations. The horizontal canal registers yaw rotations, while both vertical canals register roll and pitch rotations [Fitzpatrick and Day 2004]. The end of each canal has an ampulla, a sac that contains many hair sensory cells enclosed by a gelatinous mass, the cupula [Birbaumer and Schmidt 2010]. The hair cells are the vestibular system's detector cells for stimulation through acceleration.

When the head rotates, the endolymph fluid initially remains at rest due to inertia,

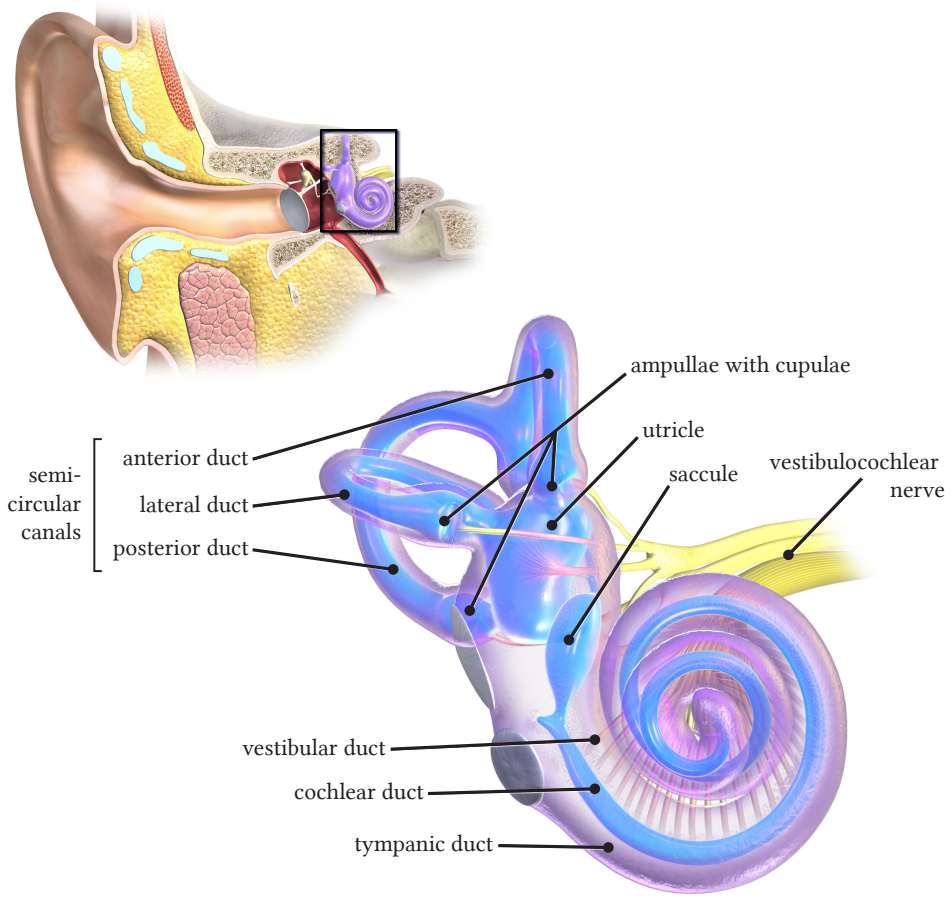

Figure 2.8: Visualization of the inner ear with the semicircular canals and otolithic organs, the utricle and saccule (Original version of [Blausen Medical 2014], CC BY 3.0)

while the cupula, firmly attached to the edge of the canal, follows the rotation. Because of the resting endolymph, the cupula is bent in the opposite direction of the rotation and the hair sensory cells are stimulated [Klinke 1993, Birbaumer and Schmidt 2010]. This stimulation is transformed into electrical activity which is transmitted to the brain, where the rotation is registered.

The functioning of the semicircular canals follows a push-pull principle, meaning that corresponding canals on opposite sides register opposite movements [LaViola 2000]. For instance, during head rotations, the hair cells in the horizontal canal in the left ear are deflected in the opposite direction to those in the horizontal canal in the right ear. The perceived movement results from the difference between the signals of both sides [Fitzpatrick and Day 2004].

2.2.2 Detection of Linear Acceleration

Linear acceleration is perceived by the otolithic organs of the vestibular system [Purves et al. 2004]. Humans have two different otolithic organs in each labyrinth: the utricle organs, which are positioned vertically in the inner ear, and the saccule organs, situated horizontally. Similar to the cupula of the semicircular canals, the otoliths have a gelatinous membrane with embedded hair sensory cells. Unlike the cupula, the membrane of the otoliths contains small crystals, the otoconia [Sheykholeslami and Kaga 2002]. Due to these otoconia, the membrane has a higher density and inertia than the surrounding endolymph. During accelerations, the dense otolith membrane slides over the hair sensory cells. As a consequence, the hair cells bend and stimulate the emission of electrical activity to the brain, resulting in the perception of motion [Purves et al. 2004]. Once a constant velocity is achieved, the membrane stabilizes and the hair sensory cells no longer get bent. Lateral movements, which are sideways, are detected by the utricle while upward and downward movements are detected by the saccule. For forward and backward movements, the motion detection is realized by a combination of both, utricle and saccule signals [Purves et al. 2004].

2.2.3 Galvanic Vestibular Stimulation

Galvanic vestibular stimulation (GVS) involves the safe stimulation of the vestibular system by low electrical currents [Utz et al. 2010]. As a subset of transcranial direct current stimulation (tDCS), the technique is extensively used in medicine, for instance in neurological therapy or tumor treatment [Brunoni et al. 2011, Nitsche et al. 2009, Song et al. 2015].

The name *galvanic stimulation* is based on the research of Luigi Galvani from 1791 who conducted early experiments on animal electricity. Around the same time, Alessandro Volta, a rival and critic of Galvani's work, performed the first human GVS experiments on himself. He felt his head spinning and reported hearing a noise, which is likely related to the high voltage he employed. Today's applications of GVS use only very weak currents of no more than 2.5 mA. Applications of GVS are often found in medical research and treatment. Positive effects of electric stimulations have previously been reported for motor, visual, somatosensory, attentional, vestibular, and cognitive functions as well as multiple neurological and psychiatric disorders [Utz et al. 2010].

Multiple studies by Wardman et al. [2002, 2003] suggest that the vestibular afferents from the otoliths and semicircular canals are similarly affected by GVS then by physical acceleration. Instead of acceleration forces, in GVS the electrical stimulation triggers

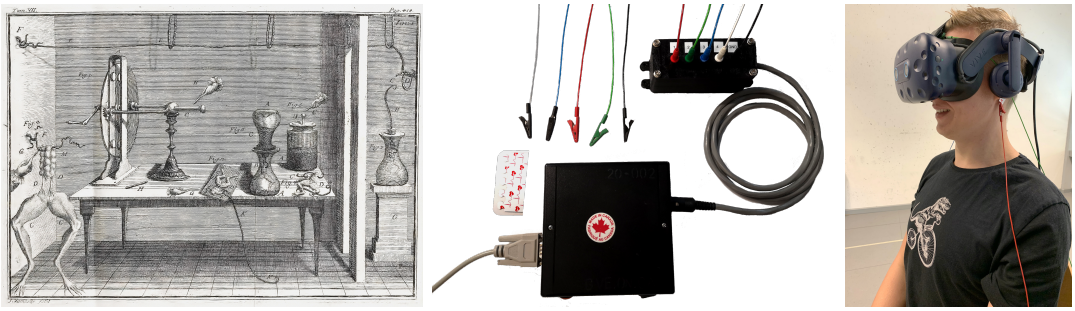

Figure 2.9: Luigi Galvani's early experiments dealt mostly with frog legs and high voltages (left). Modern GVS devices use low currents and allow for safe stimulation over multiple electrodes (middle). One promising application of GVS is VR where the feeling of motion can be induced in users that are not actually moving (right).

the automated response for head stabilization. Consequently, the cause of the signal remains hidden when interpreting the vestibular signals in the brain.

GVS is typically applied via electrodes attached to the skin above the mastoids behind each ear. Extensive research has been conducted to investigate the safety of the technique (see for example Brunoni et al. [2011]).

The most common form of GVS is bilateral, bipolar stimulation where two electrodes are attached to the mastoids. With this stimulation model, the sensation of a strong roll rotation and a weak yaw rotation towards the cathode is induced [Fitzpatrick et al. 1994, Day et al. 1997].

More sophisticated models also allow controlled stimulation around several axes. The ODAS model from Aoyama et al. [2015] and the OVR model from Cevette et al. [2012] are worth mentioning here. Both models use four stimulation electrodes. The ODAS model uses three independent circuits, whereas OVR can integrate all electrodes into one coherent circuit. Both models claim to be able to provide independent stimulation of roll, pitch, and yaw.

2.3 Virtual Reality

Virtual reality (VR) offers enhanced, engaging, and comprehensible experiences by deeply immersing the user in a virtual scene. The deep immersion enhances the understanding of visual presentations over traditional display media. The entertainment industry is a key driving force behind the advancement of virtual reality technology. Driven by technological progress, VR applications are developed also for various other purposes including complex data visualization, flight simulation, military training, and architectural planning. In the automotive industry, VR replaces physical prototypes with virtual models, reducing cost and time. For tasks that would be hazardous in reality, VR can create safe and realistic learning environments. This section provides an overview of essential VR concepts which are relevant to the understanding of the subsequent content of this thesis.

2.3.1 Terminology

For the interpretation of the term *virtual reality*, it is essential to initially explore the meaning of the components 'virtual' and 'real'. At first glance, these two terms appear contradictory, rendering "virtual reality" an oxymoron.

According to Paul Milgram and Fumio Kishino, the distinction between virtual and real depends on three aspects [Milgram and Kishino 1994]. The first aspect relates to the existence of an entity. While virtual objects possess an appearance or effect, real objects have an actual existence. The second aspect involves the mode of observation. An object is directly observed when only air and possibly a transparent material, like a window, lie between the object and the eye. Indirect observation occurs through a screen. Directly observed objects are considered real, whereas indirectly observed ones are virtual, regardless of whether they appear real. The third aspect pertains to the light emitted by an object. In real representations, luminosity is always at the location of the observed object. In virtual representations, however, the light emanates from the screen, therefore originating consistently from the same place regardless of the apparent distance of the observed object.

Consequently, VR unites the concepts of virtual and real rather than presenting a contradiction. This concept acknowledges the existence of virtual objects through appearance and effects and creates an interactive experience even though the objects have no physical form. VR relies on indirect observation via screens. The consistent light source from the screen, irrespective of the virtual object's apparent distance, facilitates a unique blend of real sensory perception with virtual environments. Therefore, VR

represents a synthesis of virtual and real elements, providing an immersive, reality-like experience within a non-physical space.

In the work of Milgram et al. in 1995, the authors describe VR as a fluctuating, highly dynamic concept of a computer-generated replication of a real or fictional environment. They state:

> *The commonly held view of a VR environment is one in which the participant observer is totally immersed in a completely synthetic world, which may or may not mimic the properties of a real-world environment, either existing or fictional, but which may also exceed the bounds of physical reality by creating a world in which the physical laws governing gravity, time and material properties no longer hold.*

In this definition, a virtual environment is characterized by the observer's immersion, which will be discussed more extensively in Section 2.3.2, and the independence from physical laws. Contrary to a uniform definition, the authors relate VR to reality to resolve the apparent contradiction between the real and the virtual.

Other common definitions of VR in the scientific literature focus on the interaction of the user with the computer-generated, three-dimensional environment [Burdea and Coiffet 2003, Gutiérrez Alonso et al. 2008]. The interaction often occurs through devices such as head-mounted displays or other hardware, allowing for a simulation of real-life environments. Immersive technology enables users to navigate and interact within a computer-simulated environment, sometimes incorporating multiple senses for more realistic experiences. Interaction, indeed, plays a key role in virtual reality, which further increases with the introduction of devices that allow for natural interaction with the environment, e.g., by camera-based hand tracking. The key aspects of interaction and its implications for VR are discussed in detail in Section 2.3.6.

VR can also be described in terms of the virtual environment. Accordingly, it is defined as a computer-generated, digital environment that can be experienced and interacted with. In this context, the environment appears real [Jerald 2015].

One of the earliest definitions of virtual reality was provided in 1965 by Ivan Sutherland, the developer of one of the world's first VR systems, through the analogy of the ultimate display [Sutherland et al. 1965]:

> *The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal.*

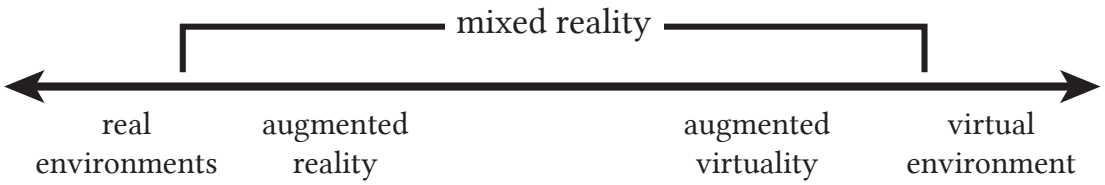

Figure 2.10: Reality-Virtuality Continuum based on Milgram et al. [1995]

This idealized version of virtual reality would occur in a space where matter is controlled by a computer. All objects would affect the observer not just virtually, but physically. Sutherland critically notes that in such a room, a bullet could be lethal to a user. Therefore, the realization of a perfect VR system raises not only technical challenges but also ethical concerns.

Reality-Virtuality Continuum

Milgram et al. [1995] describe virtual environments to be contrary to real environments, with a continuous transition in-between, formulated in the reality-virtuality continuum (see Figure 2.10). The reality-virtuality continuum differentiates between various forms of virtuality and reality. It describes a concept of classes of objects in specific observation situations. The reality-virtuality continuum also segments the term *mixed reality* as combinations of these classes. The extremes of the continuum are constituted by fully real environments on one side and completely virtual environments on the other side. Following the former definition, real environments describe spaces that consist exclusively of real objects, observed without the use of any electronic display. Virtual environments, on the other hand, comprise environments of virtual objects, which are observed through an electronic display only. An example of a virtual environment is a computer graphics simulation or video game.

Environments situated between the two extremes of the continuum are collectively considered as *mixed reality*. Mixed reality environments cover and represent both real and virtual objects in a single display. Further classification of mixed reality considers *augmented reality* and *augmented virtuality*.

An environment is classified as augmented reality when the display of a real environment is enhanced with virtual or computer-generated objects. Devices like the Microsoft HoloLens are examples of Augmented Reality displays [Microsoft].

Conversely, augmented virtuality describes virtual representations augmented with real objects. Environments of augmented virtuality are primarily computer-generated. An example is a head-mounted display (HMD) that senses real objects from the envi-

ronment and integrates them into the virtual display, like representing the user's hands in VR.

This dissertation focuses mostly on the virtual-dominated part of the reality-virtuality continuum, considering fully and augmented virtual environments.

2.3.2 Immersion

Immersion refers to the objective level to which a virtual system maps stimuli onto a user's sensory receptors [Slater and Wilbur 1997]. It characterizes a technology, describing how effectively computer displays can create an illusion for the human senses [Jerald 2015]. Slater and Wilbur define the degree of immersion to rely on six properties: Extensiveness, Matching, Surroundness, Vividness, Interactability, and Plot.

Extensiveness relates to the variety of sensory modalities presented to the user, for instance, an application that incorporates seeing, hearing, and feeling.

Matching describes the consistency among the represented sensory modalities. A fitting visual presentation that accurately matches the user's head movements with the display exhibits high congruence between sight and vestibular signals.

Surroundness refers to the physical extent to which signals are represented. Examples of high surroundness are wide FOVs or a 360-degree tracking area.

Vividness expresses the quality of stimulation. This includes factors such as resolution, lighting, and frame rate.

Interactability denotes the user's ability to make changes in the virtual world, receive feedback on those changes, and influence future events.

Plot represents the consistent portrayal of a message or experience. It also describes dynamic, exposed sequences of events and the behavior of the world and its objects.

Immersion has the potential to deeply engage users in a virtual experience. Since it always requires a person to receive and interpret the presented stimuli, immersion is a part of the VR experience itself [Jerald 2015]. This means that immersion can guide but not control the mind. The subjective experience of immersion is referred to as presence.

2.3.3 Presence

Presence describes a sense of feeling physically present in a room, even when the actual physical space differs from the experienced one [Jerald 2015]. Thus, the presence in an experience is not necessarily linked to the reality of the experience. Presence is an inner psychological state that represents a form of instinctive communication and therefore challenging to define [Jerald 2015]. Instinctive communication refers to a form of direct

communication, which relates to automatic emotions and fundamental behaviors [Jerald 2015]. An example of this in virtual technologies is the sensation of discomfort when looking down from a high building.

Lombard and Ditton describe presence as the perceptual illusion of non-mediation, where users of a media technology overlook or misconceive the presence of the medium itself, feeling directly immersed in the experience it provides [1997]. Based on this commonly accepted definition, presence can be viewed as a psychological state or subjective perception that fails to recognize an artificial world as technically generated. This means that, unlike immersion, presence is subjectively perceived. Immersion deals with the characteristics of technology, while presence represents the user's inner psychological and physiological condition. When presence is achieved, the user no longer feels like they are operating technology and perceiving its projections; instead, they interact with and perceive objects, events, and characters in the virtual world. Thus, presence is both a result of the user's perception and immersion. Immersion can create a sense of presence, but it does not always invoke it. On the other hand, presence is limited by the degree of immersion of a VR system. The better a system's immersion, the greater the potential for a user to feel present in the virtual world.

VR experiences that create a stronger sense of presence also increase the task performance of users, allowing for more effective simulations [Kim et al. 2020, Girod et al. 2016, Loup-Escande et al. 2017].

2.3.4 User Experience

The term *user experience* combines the entirety of all actions, thoughts, and feelings a person has when engaging with a product or service over time [Schumacher 2010]. It is a broad term and, in the context of VR experiences, the virtual world is the product of engagement.

For VR applications, it is crucial to evoke a good user experience, leaving the user with positive thoughts and feelings during and after the immersive encounter. When users are presented with a negative experience in their journey to the virtual world, the likeliness of their return to this virtual world at a later point will naturally decrease. Carefully designing VR products to evoke positive user experiences is, therefore, a key aspect of their general acceptance and enjoyment.

Increasing presence, immersion, and engagement in VR will lead to an increase in user experience [Kim et al. 2020]. Thereby, the sense of presence is a representative component of user experience, and increasing presence is likely to strengthen the user experience [Schuemie et al. 2001, Takatalo et al. 2008]. A stronger sensation of

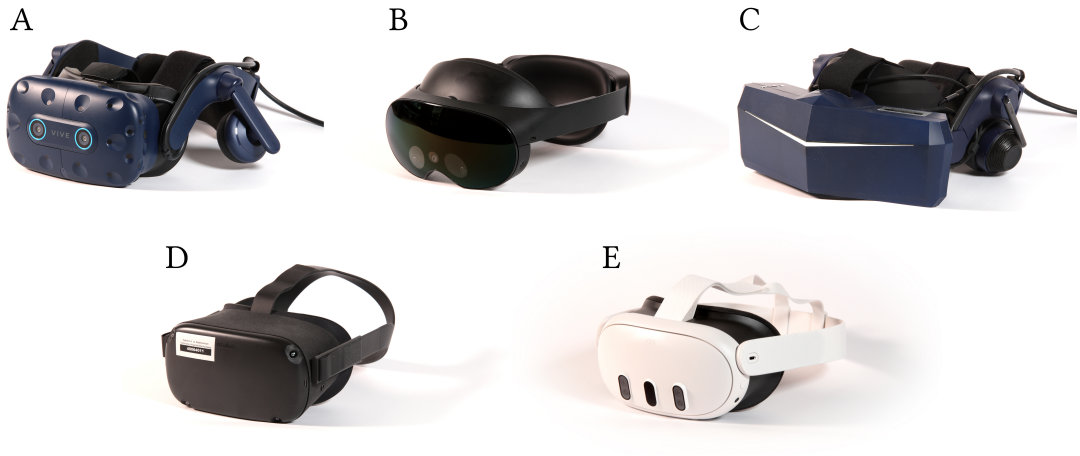

Figure 2.11: Examples of consumer-grade HMDs that were used in the scope of this dissertation: A) HTC Vive Eye, B) Meta Quest Pro, C) Pimax 8K X, D) Oculus Quest, E) Meta Quest 3.

self-motion induced by visual stimuli, referred to as vection, was shown to improve immersion and user experience in VR [Riecke 2011]. In addition, the strength of the user experience of a VR environment is influenced by the personality trait of the observer, such as empathy [Shin 2018].

On the other hand, side-effects of VR experiences can have a negative effect on the general user experience [Kim et al. 2020]. Cobb et al. [1999] categorize the influencing negative factors to be cybersickness, postural instability, psychomotor control, perceptual judgment, concentration, stress, and ergonomics effect. However, multiple studies conclude cybersickness to be the most influential negative side-effect of decreasing user experience [Nichols and Patel 2002, Stanney et al. 2003, Cobb et al. 1999]. Consequently, minimizing cybersickness is essential to sustain high comfort and a pleasant user experience. Therefore, this thesis has a clear focus on effective techniques that mitigate cybersickness to improve the user experience of VR applications.

The major factors that influence the user experience on a system level are the user, the device, and the interaction [Thüring and Mahlke 2007]. While the users and their devices can hardly be changed or predicted, we can consider the interaction as a factor to achieve engaging experiences. Core parts of this dissertation focus on the creation of immersive interactions that allow for positive user experiences in interactive scenarios.

2.3.5 Cybersickness

While virtual reality applications can offer tremendous freedom and possibilities, they also have the potential to induce negative feelings along with the virtual experience. Cybersickness describes a set of negative body symptoms that can result from visual movement inside an immersive virtual environment. Cybersickness is the most common and most severe negative effect of VR experiences [LaViola Jr 2000]. Common symptoms of cybersickness include fatigue, headache, eyestrain, discomfort, difficulty focusing, increased salivation, sweating, nausea, difficulty concentrating, blurred vision, or dizziness with eyes closed and open [Kennedy et al. 1993]. Even when the virtual experience has ended, cybersickness can prolong for some time. In some occasions, it was found that users still suffer from some negative effects of the virtual experience up to 24 hours later. Thereby, the intensity and duration vary highly between individuals. Generally, females appear to be more sensitive to cybersickness [Narciso et al. 2019].

Terminology

Any unwell physical condition caused by some form of motion is generally described by the term motion sickness [Irwin 1881, Lawther and Griffin 1988]. This motion can happen in physical as well as only in visual form [Walker et al. 2010]. Motion sickness that is experienced visually in a virtual environment is often referred to as visually induced motion sickness (VIMS), cybersickness, or simulator sickness. Cybersickness and simulator sickness describe an effect that evokes symptoms similar to those of motion sickness [Mazloumi Gavgani et al. 2018], caused by a visual simulation instead of physical movement [Dużmańska et al. 2018].

According to Stanney et al. [1997], cybersickness and simulator sickness are something different as they differ in the intensity and frequency of the symptoms. The authors base their claims on their analysis of self-reports from military flight simulator users and those of experiments with virtual environments. Cluster analysis of the symptoms revealed three clusters: oculomotor effects, nausea, and disorientation. Oculomotor effects refer to physiological impacts on the eye muscles by visual activity. While users suffering from cybersickness show a higher prevalence of disorientation effects, oculomotor effects emerged as the main side-effect in simulators. They argue that cybersickness in virtual environments is approximately three times more severe than simulator sickness. They conclude different underlying causes for the two forms of sickness and highlight the need for targeted strategies.

Contradictory, more recent research by Gavgani et al. [2018] has shown that phys-

ically induced motion sickness and cybersickness by immersion in VR are identical in terms of symptoms and physiological representation. They assume that the reason why many earlier investigations hinted at a distinction between different forms of motion sickness results from the experimental design. They argue that the intensity of cybersickness and simulator sickness varies between different studies, like in the work of Stanney et al. [1997], because a different group of participants was used: When simulator sickness is to be triggered, typically flight simulators with trained pilots are chosen. In contrast, experiments about cybersickness are conducted with arbitrary people. Naturally, the first group is less susceptible to motion sickness than the other group due to their training [Mazloumi Gavgani et al. 2018]. Furthermore, for the distinction between occurring symptoms regarding physical motion sickness and simulator sickness, respectively, they claim that participants were not stimulated to their tolerance limit in simulator sickness experiments and therefore the symptoms differ. When strongly experiencing simulator sickness, nausea symptoms are likely to appear stronger than oculomotor effects [Mazloumi Gavgani et al. 2018]. In comparison to former works, Gavgani and colleagues conducted two experiments where they tried to cause sickness from purely visual and purely vestibular stimuli in the same group of participants. Their results indicate that motion sickness, simulator sickness, and cybersickness are indistinguishable from a clinical point of view.

This dissertation refers to cybersickness as a special form of motion sickness where the perceived motion is induced by visual stimuli of a VR environment. It follows the assumption that the symptoms and severity of cybersickness and motion sickness are equivalent.

Motion Sickness Theories

There are several theories about the causes of motion sickness and it is still controversial which theory is most likely to apply [Dobie 2019].

The most common reason for the cause of motion sickness above all theories describes a conflict between different modalities, particularly between vision and the vestibular system [Reason and Brand 1975, Reason 1978, Treisman 1977]. This discrepancy, referred to as the sensory conflict theory, is described to be the foundation of the negative symptoms of the body [Reason and Brand 1975]. Reason and Brand describe sensory rearrangement as a core principle of this theory. Sensory rearrangement arises when the perceptions of one modality, in a given situation, do not match with the perception of functionally related modalities. According to them, the vestibular system has to be involved in any sensory conflict causing motion sickness. This explanation matches the

assumptions of Golding et al. [2006] who state that any person with an intact vestibular system will experience motion sickness eventually when the simulation gets increasingly stronger.

The sensory conflict theory defines two categories for situations that are likely to cause motion sickness: visual-inertial rearrangements, i.e., conflicts between modalities, and canal-otolith rearrangements, i.e., conflicts within the vestibular system [Reason and Brand 1975]. Furthermore, these categories can be divided into three subcategories each differing in what system sends contradicting information (Type 1: both systems; Type 2 and 3: one of the systems, respectively). From this follows that cybersickness is a visual-inertial rearrangement of type 2 with the visual system sending contradicting information and the vestibular system remaining passive.

Later, Reason revised his work and published the neural mismatch theory because the sensory conflict theory did not sufficiently explain why humans adapt to motion sickness [1978]. Their theory states that motion sickness results from a mismatch between the perceived and expected perception in a given situation and not between the perceptual modalities themselves. The theory's name is derived from the mismatch in the brain between the neural activations from the different perceptual information and the stored neural connections from learned experiences. With the neural mismatch theory, the adaptation to motion sickness can be sufficiently explained, as the mismatch signal can also be stored with repeated occurrence.

For Treisman [1977], motion sickness results from evolutionary reasons. In his toxin detector theory, he argues that motion sickness arises from problems with motor coordination caused by conflicts between the spatial orientation systems of the body. In his point of view, motion sickness is adaptive for humans in an evolutionary sense and the symptoms arising from motion sickness are a warning sign of neurotoxin poisoning which is tried to be removed.

Riccio and Stoffregen [1991] argued that sensory conflicts not only occur in situations that induce sickness but also in non-sickness-inducing situations. Therefore, the sensory conflict cannot be the only reason for motion sickness. Furthermore, because the sensory conflict theory is based on the unmeasurable expectations of the organism it cannot be falsified. They suggest that motion sickness is rather based on prolonged postural instability causing the symptoms to occur.

2.3.6 Interaction

> *Tell me and I will forget, show me and I may remember, involve me and I will understand.*
> — Confucius

Interaction is considered as the communication between a user and the VR application, mediated by the input and output of the involved technical devices [Jerald 2015]. Based on this definition, interaction is a subset of general communication. The three-dimensionality of virtual worlds together with the stereoscopic perception in VR suggest that interaction in this media is most natural and intuitive when it occurs in the same form as interaction in the real world. However, interfaces in VR are usually designed in a considerably different way from their real-world comparisons. First, this deviation is attributed to the lack of proper recreation of the real-world in its entirety. More importantly, however, abstract representations of interaction are often much more practical than the same interaction in the real-world. For example, to obtain information of a book in virtual reality, it would be impractical to first walk in the virtual world to a representation of a library.

While extensive interaction methods have been designed for traditional desktop interactions, they usually do not provide intuitive interaction in VR. When interactions are designed well, however, they allow for high performance and comfort.

Usual input for interaction in VR requires tracking the pose and motion of the head, body, and hands of users. Tracking is performed either with tracking hardware based on sensors, such as accelerometers and gyroscopes, or using optical localization using cameras.

Types of Interaction

Different types of interaction can be represented on a continuum from direct to indirect interaction [Hutchins 1986]. Depending on the task, both direct and indirect interactions are useful for VR applications.

Direct interaction refers to the impression of direct communication with an object, without any intermediary in the interaction process [Hutchins 1986]. The most direct form of interaction occurs when the user physically holds the object they are interacting with in their hands. Interactions where objects are virtually held in the hands and follow their movements, like a virtual pen, are considered slightly less direct.

Indirect interaction, on the other hand, requires more cognition and a mental conversion between input and output [Hutchins 1986]. For an example of indirect interaction

consider a text-based image search task: The users must infer their objective, articulate a precise search query, and finally interpret the generated results for relevance. Although indirect interaction demands more cognitive effort, direct communication is not inherently superior. Indirect interactions are significantly more effective for their intended purpose. For instance, if users were to manually browse through all available images, their search would be more direct compared to using a search function, but not necessarily more efficient.

In the continuum between direct and indirect interaction, semidirect interactions occupy the middle ground. One example of semidirect interactions is a dimmer switch for lighting. The user directly manipulates the switch as a mediator but interacts less directly with the light itself. However, after several uses, the interaction feels much more direct as a mental representation is formed linking the position of the dimmer to the intensity of the light.

Interaction Realism

The realism of interactions in virtual environments can be defined on a continuum ranging from highly representative, realistic interactions to non-realistic interactions. The degree to which the physical actions of virtual tasks correspond to the physical behavior for the same tasks in the real world is termed *interaction fidelity* [McMahan et al. 2012].

Realistic interactions closely mimic reality and carry high interaction fidelity [Jerald 2015]. An example of a realistic interaction is holding a virtual bat and swinging it to hit a virtual ball. Realistic interactions are particularly relevant in training simulations, such as surgical simulation applications. Realistic interactions require the user to already possess or be familiar with the corresponding real-world action.

Non-realistic interactions, on the other hand, are interactions that do not exist in the real world. Moving in VR through teleportation is an example of a non-realistic interaction with low interaction fidelity. However, lower interaction fidelity does not necessarily imply a disadvantage. It can enhance performance and enjoyment, and reduce user fatigue [Jerald 2015].

The optimal level of realism depends on the specific application, with most interactions falling in the middle of the spectrum.

Visual-Haptic Conflict

Virtual reality systems typically lack or provide only limited haptic feedback. In contrast, human touch in the real world provides a nuanced perception of objects and

surfaces. The phenomenon known as the *visual-haptic conflict* arises from the discrepancy between the visual occlusion information processed by the HVS and the tactile feedback of haptic perception [Jerald 2015]. This conflict often originates from visually rich interactions lacking a corresponding physical counterpart of the objects and their properties.

The primary issue in VR haptics is not so much the absence of tactile sensation, but rather the lack of haptic feedback corresponding to interactions with objects and the perception-based information processed by the brain. This visual-haptic conflict, stemming from incongruent visual and tactile perceptions, can confuse users due to unmet expectations, thereby diminishing the sense of presence. Furthermore, insufficient tactile feedback can limit the perceivability of details and make interactions with the scene and objects more challenging.

In cases where physical representations of virtual objects exist in the real world, for example, through generic haptic proxies, redirection methods can be employed to correct for misalignments. These methods subtly guide the user toward the intended target to resolve the visual-haptic conflict through appropriate representations.

In situations where virtual objects have no real-world representation and physical stimulation from interactions cannot be induced, sensory substitution allows to partially mask the visual-haptic conflict. Sensory substitution refers to a phenomenon where one sense is replaced by one or more other senses. For example, the absence of haptic feedback can be adequately compensated by enhanced visual information or auditory signals. There are several approaches for sensory substitution of missing haptic cues [Jerald 2015]:

Highlighting involves visually outlining or coloring objects when a hand is sufficiently close. This technique is often used to indicate the possibility of interacting with an object.

Auditory signals can be very effective in alerting the user to a collision between the hand and an object.

Controller vibrations can also be an effective way of alerting the user to a collision between their hand and an object.

2.3.7 Further Conflicts of Perceptual Modalities

In the former sections, the negative effects on the VR experience primarily originated from movement and interaction with the virtual environment. However, visual stimuli from a static scene can also lead to adverse user experience impacts. Known issues include the Vergence-Accommodation Conflict and the Binocular Occlusion Conflict.

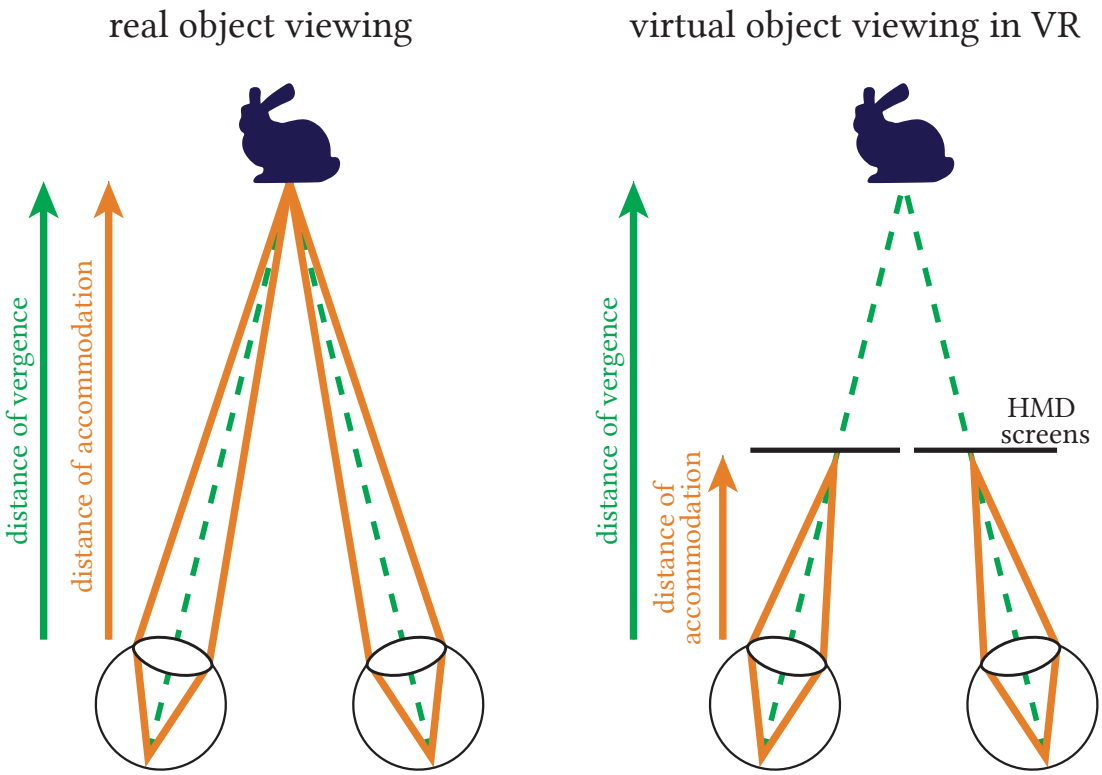

Figure 2.12: Illustration of the vergence-accommodation conflict: The left side depicts natural viewing, where the eyes' vergence and accommodation distances are in sync. In contrast, the right side demonstrates a VR scenario, highlighting the mismatch as the eyes accommodate to the close proximity of the screen, while still converging as if focusing on a distant object.

Vergence-Accommodation Conflict

The vergence-accommodation conflict occurs when the distance to which the eyes converge does not match the distance to which they focus [Kramida 2015]. This adverse effect that can occur in HMDs is illustrated in Figure 2.12.

As discussed in Section 2.1.8, vergence refers to the simultaneous movement of both eyes to focus on a single point to match the positions of objects on the retinal images of both eyes. Accommodation is the dynamic adjustment of the eye's lens to see objects in focus at different distances. In the real world, vergence and accommodation work closely together to provide clear vision. However, with HMDs, a conflict between vergence and accommodation arises, as the perception of the virtual world does not align with the perception of the real world. In VR, the stereoscopic images of both eyes create the illusion of depth and distance, while being projected on a screen at a fixed optical

distance. Therefore, while the eyes apply vergence movements to focus on a seemingly near or far object in the virtual space, the accommodation response is always fixed to the screen's constant distance. This unnatural scenario forces the visual system to operate in a way it is not typically used to, leading to negative effects like eye strain and discomfort [Kramida 2015].

Binocular Occlusion Conflict

The binocular occlusion conflict arises from discrepancies in the stereoscopic information presented to each eye [Au et al. 2022]. The effect particularly concerns the occlusion of objects in a way that does not align with the expectations of binocular vision. The consequences of the binocular occlusion conflict for the user can range from confusion to discomfort. In VR, the conflict often occurs when text is overlaid in the rendered image without consideration of the 3D geometry of the scene.

To prevent the binocular occlusion conflict in VR applications, all overlays should be provided with a three-dimensional position, allowing to render occlusions by surrounding geometry. This approach ensures that the stereoscopic presentation is consistent with the spatial and depth cues inherent in binocular vision, thus reducing potential visual conflicts and enhancing the overall user experience [Maus et al. 2016].

3 Related Work

This chapter examines existing research on user experience in VR, a key aspect of applications for their overall acceptance and enjoyment.

In static virtual environments where the user does not interact with the environment, the user experience and the feeling of presence in the virtual world depends solely on the quality of the stimuli presented to the user. This stimuli quality relies on several factors, such as the hardware, the fidelity of the stimuli, but also storytelling. These basic factors are considered in the **first part** of this chapter.

Static VR environments, however, are not overly exciting, which is why some form of user interaction with the world is included in most applications. The user's interaction with the world adds a new layer of complexity to the experience. When individuals interact with a virtual object or the world itself, they always have implicit expectations about the reaction or outcome of the interaction. When these expectations are not met it strongly influences the immersion in VR and the quality of the user experience. The **second part** of the chapter addresses techniques to reduce breaks in immersion caused by incorrect interactions.

The final component that VR experiences usually include is movement. VR offers designers the opportunity to create endless worlds that push the boundaries of human imagination. To explore these virtual worlds some form of virtual movement is required. Unfortunately, this movement often leads to negative physical effects for the user. This phenomenon, known as cybersickness, is one of the strongest factors influencing the user experience and can even lead to complete rejection of the technology by consumers. Accordingly, many scientists have studied cybersickness and potential mitigation strategies. The results of this research are outlined in the **third part** of this chapter.

3.1 General Factors for User Experience

The overall user experience is influenced by a manifold of factors, both on the technological side and from the applications themselves. This section explores a variety of important aspects that can have a lasting effect on the experience of the user in VR.

Related Work

Stimuli and Realism In general, high frame-rate renderings, high-quality tracking, and reduced latency systems reduce the occurrence of cybersickness symptoms and increase the overall user experience [LaViola Jr 2000, DiZio and Lackner 1997, Sherman 2002]. The virtual experiences can further be enhanced by providing a multimodal experience. When the perception of multiple senses, such as haptics, vision, and audio, are in line with the virtual experience, users are generally more immersed in the virtual world [Danieau et al. 2012]. While games are a prominent content type in immersive media and therefore receive a lot of attention, pre-recorded content like 360° videos can also provoke cybersickness and have a significant influence on the user experience [Elwardy et al. 2020, Kim et al. 2018a]. The video format of panorama videos, i.e., monoscopic versus stereoscopic, was found to not influence the users' feeling of presence and cybersickness [Narciso et al. 2019]. Elwardy et al. [2020] evaluated cybersickness for 360° videos in VR. The authors were particularly interested in how the level of previous VR experience influences the outcome. As a result, mainly participants with little experience in immersive media suffered from negative symptoms. Other research uses neural networks to predict the severity of pre-recorded content to evoke negative symptoms that result in a negative VR experience [Kim et al. 2018a, Islam et al. 2021].

Hardware and Devices In recent years, VR hardware has experienced significant advancements that greatly improve user experience, characterized by enhanced performance and hardware capabilities. Modern VR headsets offer higher resolution displays, which provide sharper and more detailed visual stimuli. This increase in resolution, alongside more accurate color reproduction and recent advancements in high-dynamic-range (HDR) displays, has substantially reduced the screen-door effect, which was a common issue in earlier VR models [Wu et al. 2023]. Reduced latency and higher refresh rates of modern hardware ensure smoother motion and interaction, thereby helping to prevent cybersickness and enhance the realism of the simulation [LaViola Jr 2000, DiZio and Lackner 1997, Sherman 2002].

Moreover, modern headsets increasingly integrate eye-tracking technology. Eye trackers allow for more intuitive control and interaction within the virtual environment and enable foveated rendering, where only the foveal area on the display is rendered at full resolution for high-performance gains without loss in perceptual quality [Patney et al. 2016]. Additionally, advancements in ergonomics and weight distribution of VR headsets have enhanced comfort and allow users to experience virtual worlds for an extended period of time [Zhang et al. 2023].

Audio The quality and type of audio in virtual worlds were found to have a significant effect on the user's presence and spatial awareness [Brinkman et al. 2015]. Adding auditory information to VR generally increases the user experience through a higher immersion into the virtual space. In the studies of Brinkman et al. [2015] this effect was reflected by an increased heart rate when including audio for the display of virtual insects. The audio format, comparing stereo and 3D audio, was found not to influence users' feeling of presence in VR scenarios [Narciso et al. 2019, Brinkman et al. 2015].

Performance and Storytelling While high realism in VR experiences, often characterized by the integration of many sensory modalities, can enhance the user experience, it does not necessarily lead to better task performance of users. In fact, in some cases, it can even worsen user's performance [Veit et al. 2009]. Possible explanations are the cognitive overload of the experience due to the high freedom, and the complexity in the task execution. Post-processing the rendered image with semi-transparent FOV modulators does not seem to influence the performance of gamers [Hillaire et al. 2008]. In the study of Hillaire et al. [2008], participants even reported an increase in realism with the peripheral blur.

Storytelling in VR significantly influences the user experience by enhancing immersion, empathy, and embodiment [Shin 2018, Dooley 2017, Ladeira et al. 2011]. However, the effectiveness of VR storytelling is highly dependent on individual user traits and their active engagement with the content [Shin 2018]. This means that while VR has the potential to create powerful storytelling experiences, its impact can vary significantly among different users.

3.2 Interaction

Interaction is a key aspect of engaging, immersive virtual reality experiences. The form in which interaction is incorporated into virtual worlds can range from natural exploration through body motion to touching objects with the virtual representations of the user's hands. However, every interaction can cause breaks in the user's sense of presence as soon as the tactile and visual cues do not match. For instance, mismatches can arise when the physical movement space ends before the virtual one, or as soon as a virtual object does not generate haptic feedback. The following sections explore techniques that aim at avoiding sensory mismatches for different interaction scenarios.

3.2.1 Natural Walking in VR

The most natural way to explore virtual spaces is through the direct display of physical body movements in the virtual world. Such natural locomotion incorporates the vestibular sensation into the virtual experience and allows for a high level of immersion. However, virtual spaces are typically vast and cannot be compared directly to the dimensions of the (typically) small physical space available.

Solutions to enable a natural exploration of virtual spaces of bigger dimensions than available in the real environment involve redirecting the real-world movements of the user to allow for a bigger virtual interaction space [Sun et al. 2018, Langbehn et al. 2018, Nguyen and Kunz 2018, Grechkin et al. 2016, Thomas and Rosenberg 2020], or walk-in-place methods where the user remains on the same physical position [Wang et al. 2023, Feasel et al. 2011, Kassler et al. 2010, Iwata et al. 2005, Avila and Bailey 2014, Cakmak and Hager 2014].

Classical redirection approaches involve scaling the translational and rotational movements of the user with a constant factor that is below the threshold of detectability [Grechkin et al. 2016]. Ideally, VR users are tricked into walking in circles in the real world while they believe in steering straight based on the virtual presentation.

Aside from constant scaling of movements, eye blinks were leveraged to mask instantaneous changes in the scene [Langbehn et al. 2018, Nguyen and Kunz 2018]. In the work of Langbehn et al. [2018], the phenomenon of change blindness was exploited for real-time adjustment of position and rotation redirection.

Change blindness for redirected walking can further be leveraged based on saccadic suppression [Sun et al. 2018]. In the work of Sun et al. [2018], redirection was achieved by a real-time algorithm for path planning in rooms with dynamic objects. Thereby, the saccadic-based redirection was further enhanced with gaze direction methods tailored to VR perception. Based on their results, the rotation gains by saccade redirection can be increased without introducing visual distortions or cybersickness.

With additional hardware, a natural walking behavior of the users can also be achieved in-place. In-place walking techniques in VR synchronize physical movement with the visual input which significantly reduces the vestibular sensory mismatch and the associated risk of cybersickness compared to continuous controller navigation. A simple example of active repositioning in a linear direction are traditional treadmills [Feasel et al. 2011, Kassler et al. 2010, Powell et al. 2011]. These locomotion devices allow to maintain a natural walking sensation while remaining in the same physical location. However, traditional linear treadmills only support simple forward movements. With omnidirectional treadmills, users are enabled to walk more freely in a direction

of their choice [Wang et al. 2023, Darken et al. 1997, Iwata 1999, Noma and Miyasato 1998, Souman et al. 2011].

Other repositioning systems encompass motorized floor tiles that move the walker back in place [Iwata et al. 2005], step-canceling via shoe-attached strings [Iwata et al. 2007], and large-scale hamster balls [Medina et al. 2008]. Passive repositioning can be achieved with friction-free platforms that keep the VR user in place by preventing the forward forces that are usually created during footsteps [Avila and Bailey 2014, Cakmak and Hager 2014, Huang 2003, Iwata and Fujii 1996, Swapp et al. 2010, Walther-Franks et al. 2013].

3.2.2 Hand Redirection

Similar to redirected walking, hand redirection involves deliberate manipulations in the mapping between the real and virtual environment. In contrast, hand redirection does not aim to increase the interaction space but tries to gain control over the movement of the user's hand. With hand redirection in VR, the user's real hand movement trajectory is subtly steered towards a target different from the one the user perceives from the virtual representation. The redirection of hands is useful in various VR applications, most notably to enhance the scalability of passive haptic feedback [Hinckley et al. 1994].

For a reasonable use of hand redirection, some form of physical presentation in the real environment is required, for example, a haptic proxy [Cheng et al. 2017]. In this scenario, hand redirection can be applied to map some physical target to another virtual target when they are spatially apart [Kohli et al. 2012, Azmandian et al. 2016, Cheng et al. 2017], or to avoid collision with targets that do not have a representation in the virtual world [Murillo et al. 2017]. In most cases, hand warping techniques involve either a constant offset [Benda et al. 2020, Han et al. 2018] or incremental relocation of the real-to-virtual mapping [Cheng et al. 2017, Azmandian et al. 2016, Burns et al. 2006, Esmaeili et al. 2020, Kohli et al. 2012, Spillmann et al. 2013, Zenner et al. 2021].

In the work of Kohli et al. [2012], redirected touching is introduced, a technique that combines distortions in the virtual scene with hand redirection to convey the perception of differently shaped virtual objects using only a single haptic proxy. Azmandian et al. proposed using hand redirection for haptic retargeting, allowing users to interact with spatially dislocated virtual cubes mapped onto one single physical proxy [Azmandian et al. 2016]. This proxy in the real environment allowed users to have proper haptic feedback during their interactions with the virtual objects.

Further approaches demonstrate that sparse haptic proxies allow reliable hand redirection in VR for simulating touch feedback [Cheng et al. 2017]. Cheng and col-

leagues [2017] demonstrated the effectiveness of continuous hand redirection and discovered that touch intentions can be predicted based on users' eye gaze. Their work emphasized the importance of considering both physical and cognitive factors in designing immersive and realistic VR experiences with enhanced touch sensations.

Furthermore, hand redirection has been applied to enhance the perceived resolution and speed of shape displays [Abtahi and Follmer 2018], to overcome limitations in interactive touch feedback systems [Abtahi et al. 2019, Gonzalez et al. 2020], to extend the range of haptic effects [Zenner and Krüger 2017, 2020], and to enable more ergonomic interactions with virtual user interfaces and scenes [Murillo et al. 2017]. Additionally, redirection techniques have been explored in the context of 3D interaction techniques and pseudo-haptic effects, simulating drag or weight sensations [Dominjon et al. 2005, Rietzler et al. 2018, Samad et al. 2019].

Previous research utilizes the change blindness effect during spontaneous blinks to offset the virtual hand [Zenner et al. 2021]. While the results showed sufficient success and notable detection thresholds, relying solely on spontaneous blinks occurring on average every three seconds may not be sufficient for rapid hand movements [Nakano et al. 2013].

3.2.3 Haptic Feedback

While the former approaches try to subtly adjust the virtual space and presentations to match the real environment, several techniques allow to directly create tactile sensations for haptic feedback on virtual touch.

Haptic feedback devices allow the direct stimulation of haptic feedback in VR [Jung et al. 2024, Maereg et al. 2017, Preechayasomboon and Rombokas 2021]. Examples include wearable vibrotactile haptic units for stiffness discrimination [Maereg et al. 2017] and finger-worn, wireless feedback devices for enhancing tactile experiences [Preechayasomboon and Rombokas 2021]. These devices allow for more immersive VR interactions by enabling users to feel subtle differences in virtual environments through refined, tactile cues.

VR interactions can further be enhanced by providing the user with direct force feedback over hand-held devices [Ban and Hyun 2019, See et al. 2022]. Also, special gloves can create a strong tactile sensation [HaptX, SenseGlove]. However, this hardware is usually quite costly and not very compact.

Recent research in ultrasonic haptic feedback for VR enables to induce tactile sensations without direct contact with the user. The use of piezoelectric micromachined ultrasonic transducers [Liu et al. 2022] and robotically actuated ultrasound transduc-

ers [Faridan et al. 2022] is applied to create more compact, efficient, and versatile systems to simulate tactile sensations in virtual environments.

3.3 Cybersickness

For immersive content, conflicts between the visual and vestibular signals can result in cybersickness and quickly decrease the user experience [Reason and Brand 1975, Danieau et al. 2012]. Accordingly, many techniques have been proposed to mitigate cybersickness. Efforts towards this goal differ in the knowledge of the scene they require.

3.3.1 Screen-Space Modifications

Various techniques have been explored that modify the visual stimuli for cybersickness mitigation. These techniques do not require any knowledge of the three-dimensional virtual scene but are applied as a post-process of the rendered images. Most commonly, post-processing techniques for cybersickness reduction consider some form of filtering on the image content.

A common and unobtrusive technique involves the gaze-contingent blurring of peripheral image content with lower resolution representations, like Gaussian filtering [Patney et al. 2016, Hillaire et al. 2008]. Alternatively, the blurring is applied in fixed outer regions [Lin et al. 2020b] or non-salient areas [Nie et al. 2017, 2019].

Also opaque occlusions in the visual field were shown to reduce vection and cybersickness. This content masking was implemented with fixed central regions [Bos et al. 2010, Lin et al. 2002, Seay et al. 2001] or based on the eye gaze [Adhanom et al. 2020]. However, as intuitively assumed, full occlusions of the peripheral content are significantly more intrusive for users compared to smooth blurring and can not always maintain a positive user experience.

Occlusion approaches, either opaque or by blurring, act as a low-pass filter on the content and reduce contrast and information perception, thereby diminishing cybersickness. However, the information reduction in the periphery can be an issue in applications that require high-frequency information for fast reactions like in VR shooter games.

Based on the results of Zayer et al. [2019], methods that modify the FOV are equally efficient for men and women, and do not influence the spatial navigation performance. Combining the reduced FOV with a content-independent background grid can increase the positive effects of the technique with the static reference of the grid [Bala et al. 2020]. However, the static reference also negatively affects the user's immersion in the scene.

Recent methods include integrating reverse optical flow visualizations [Park et al. 2022, Kim and Kim 2022]. The idea is based on the larger pooling of the ganglion cells in the peripheral vision [Anderson et al. 1991]. Park et al. 's [2022] experiment with reverse optical flow arrows reduced cybersickness but significantly affected participants' experience due to its intrusiveness and application in both peripheral and foveal regions. Another visual method to reduce cybersickness involves the discretization of rotational animation frames [Farmani and Teather 2018].

While these techniques promise to potentially minimize cybersickness, they represent a compromise and are often connected to a negative user experience or reduced feeling of presence [Lin et al. 2020b, Moghadam et al. 2018, Weißker et al. 2018].

3.3.2 Scene Manipulations

While the former techniques aim to be more generally applicable, knowledge of the 3D scene can be used to avoid cybersickness by design. Designing a virtual world in a way that does not induce cybersickness requires avoiding the induction of vection by visual stimuli. Vection is evoked by accelerations of the scene camera that is not experienced in reality. Therefore, the safe way (in terms of cybersickness) to move through virtual worlds is by a direct mapping of the body movement in reality. Given the space limitations in reality, however, this type of locomotion is not applicable in many applications. Instead, virtual movement is used. A common and simple technique for vection avoidance during virtual locomotion is teleportation movement, where the user navigates through the scene in a jump-like manner. Since each jump to a new location is instantaneous, cybersickness symptoms can be kept minimal [Weißker et al. 2018, Moghadam et al. 2018]. However, virtual teleportation can also adversely affect the spatial orientation of users [Weißker et al. 2018].

For a known camera path, alterations in velocity along the path can also reduce the severity of cybersickness [Hu et al. 2019]. In the work of Hu et al. [2019], the intensity of cybersickness for a predefined path and scene is predicted based on the characteristics of the virtual motion and the composition of the scene. The authors show that optimizing the predefined camera trajectory based on a perceptual model can significantly reduce cybersickness.

The relationship between geometry appearance, motion perception, and cybersickness is less strongly investigated. In a simple prototype of Lou et al. [2022], the geometry of a building is squeezed towards the viewport's edge during forward movements. While their project was not generally applied or experimentally validated the authors' approach motivates the use of geometrical deformations for vection reduction.

3.3.3 Enhancement of Vestibular Signals

Former methods focus on the reduction of vection by the visual stimuli to keep the sensory mismatch at a minimum. However, the mismatch can also be resolved by enhancing vestibular signals.

The work of McGill et al. [2017] highlights how the adjustment of vestibular stimuli affects cybersickness and immersion. In an in-car experiment, the movements of a 360° video were synchronized with the movements of an in-motion vehicle. No single method was able to consistently mitigate cybersickness across all users. However, peripheral motion cues helped most users to enhance vestibular signals and reduce cybersickness. The efficiency of re-adjusting the misaligned sensory impressions is also used in commercial products such as the SEETROËN glasses [Citroën 2024]. By visualizing the impressions of the vestibular system via liquid-filled rings, a visual reference to the perception of acceleration is created, which prevents motion sickness.

For static participants, visually induced self-motion can be enhanced with galvanic vestibular stimulation (GVS) [Maeda et al. 2005, Aoyama et al. 2015, Reed-Jones et al. 2007, Sra et al. 2019]. Maeda et al. [2005] applied GVS while showing white dots move on a screen. Even with this simple setup, a strong feeling of vection by GVS could be induced.

In the following years, several works were published that use bilateral GVS without a clear focus on cybersickness [Fujimoto et al. 2016, Sra et al. 2017, Moore et al. 2011, Maeda et al. 2005]. For example, GVS was used as a game design tool for vertigo games [Byrne et al. 2016].

More sophisticated stimulation models with multiple electrodes promise a feeling of motion around multiple axes [Aoyama et al. 2015, Cevette et al. 2012]. Aoyama et al. [2015] investigated the enhancement effect of GVS with counter-currents on human acceleration perception. In their experiment, they used a 4-pole GVS with a capacitor-resistors circuit model in VR. They found that counter-current stimulation enhances the strength of the acceleration sensation whereby the duration of the stimulation correlates with the perceived strength.

The oculo-vestibular recoupling (OVR) stimulation model uses five electrodes to stimulate the three principle rotation axes (yaw, pitch, roll) individually [Cevette et al. 2012]. This model was designed as a result of the aerospace research of Cevette et al. [2012] using flight simulators.

In 2007, GVS was applied to counter cybersickness in VR for the first time [Reed-Jones et al. 2007]. While in these early VR experiments the GVS was still activated by hand, positive indications for cybersickness reduction with GVS emerged.

Related Work

Sra et al. [2019] presented an extensive experiment with bilateral GVS, a portable stimulation device and custom app control. The vestibular feedback, which has to be predefined, significantly lowered levels of cybersickness for the tested scenes while the immersion is higher when GVS is applied.

While the vestibular stimulation is mainly performed with electrical currents, also bone-conducted vibrations can be used as a stimulus [Weech et al. 2018]. The vibrations achieve the same effect when they are randomly induced compared to time-coupled stimulations [Weech et al. 2018]. Still, this technique remains uncommon in VR research and has a lower impact as electrical stimulation [Weech et al. 2020].

Noisy GVS is a promising technique to stimulate the vestibular system when the visual movements in the scene are unknown [Weech et al. 2020]. In the work of Weech et al. [2020], the researchers focus on sickness severity in VR during and immediately following noisy GVS. They state that noisy stimulation can reduce cybersickness severity for intense VR content, while it has no effect for moderate content. After GVS exposure, the vestibular system quickly re-adopted to its normal state.

This concludes the discussion of relevant previous work on user experience in VR. The subsequent chapters present the publications that form the primary contributions of this dissertation. These publications are the main part of this thesis.

A Omnidirectional Galvanic Vestibular Stimulation in Virtual Reality

In this paper we propose omnidirectional galvanic vestibular stimulation (GVS) to mitigate cybersickness in virtual reality applications. One of the most accepted theories indicates that Cybersickness is caused by the visually induced impression of ego motion while physically remaining at rest. As a result of this sensory mismatch, people associate negative symptoms with VR and sometimes avoid the technology altogether. To reconcile the two contradicting sensory perceptions, we investigate GVS to stimulate the vestibular canals behind our ears with low-current electrical signals that are specifically attuned to the visually displayed camera motion. We describe how to calibrate and generate the appropriate GVS signals in real-time for pre-recorded omnidirectional videos exhibiting ego-motion in all three spatial directions. For validation, we conduct an experiment presenting real-world 360° videos shot from a moving first-person perspective in a VR head-mounted display. Our findings indicate that GVS is able to significantly reduce discomfort for cybersickness-susceptible VR users, creating a deeper and more enjoyable immersive experience for many people.

Colin Groth and Jan-Philipp Tauscher and Nikkel Heesen and Max Hattenbach and Susana Castillo and Marcus Magnor
Institute for Computer Graphics, TU Braunschweig, Germany

Journal article published in *IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG) 2022,* Volume 28, Article 5, Pages 2234–2244. Presented at *IEEE VR 2022, ACM SIGGRAPH 2022, IEEE VIS 2022.*
DOI: 10.1109/TVCG.2022.3150506

Omnidirectional Galvanic Vestibular Stimulation in Virtual Reality

Figure A.1: To reduce cybersickness for moving-camera sequences in VR, we evaluate the effectiveness of galvanic vestibular stimulation. We stimulate the VR user's vestibular sense in all three spatial directions taking the motion of the camera in the 360° video as well as the user's current viewing direction into account. This way, we aim to reconcile visually induced and felt self-motion.

A.1 Introduction

While virtual reality (VR) is not a new technology, it is only in recent years that it has started to win more and more support and acceptance in society [Rheingold 1991]. This trend is driven by new VR devices as well as an increasing number of games, videos and even movies for VR. However, this encouraging progress also raises the bar for expectations and the acceptance of the general public.

A main reason that curbs the spread of immersive content are feelings of discomfort caused by the VR experience. The most commonly known adverse effect is cybersickness (CS). This term describes any physical discomfort evoked by visually perceived motion that is not actually experienced [Irwin 1881, Lawther and Griffin 1988]. The mismatch between the visual and vestibular channel is the origin of CS as described by the *sensory conflict theory* [Reason and Brand 1975], one of the most accepted theories. CS covers an extensive collection of symptoms which include, in a low state, oculomotor effects (e.g., blurry vision), headaches, and dry eyes. In severe cases, the user can even experience disorientation and nausea [Kennedy et al. 2010, Stanney et al. 1997]. Although CS is widely known for VR games and virtual worlds, it is not explored much and sometimes underestimated for real-world content in VR. Still, CS can have huge consequences and strong implications for immersive video material [Groth et al. 2021b, Elwardy et al. 2020, Kim et al. 2018a].

For scenes were ego-motion is visually perceived, a common way to reduce CS is to minimize the optical flow, e.g., with a reduced field of view (FOV) or blurred outer

regions [Groth et al. 2021b, Bala et al. 2020, Adhanom et al. 2020, Groth et al. 2021a]. These techniques reduce the visual motion and readjust it to the vestibular sensation. Thereby, the compensation of rotational movements is particularly important as they contribute most to CS [Groth et al. 2021b, Kim et al. 2021]. While these methods are effective against CS, the observer's perception is moved away from the virtual experience back to the sensations of the real world. As a result, the feeling of presence of the viewers can suffer. For VR scenes it is most important to create experiences that feel as real as possible and, thus, evoke high presence. Accordingly, we require the vestibular system to perceive the same movements like those visually observed from the VR scene. Using galvanic vestibular stimulation (GVS), the vestibular system can be influenced with small currents inducing a feeling of motion that is different from the real-world sensation. As the currents are kept very small (typically below 2.5mA) this process is safe and has no lasting effects [Utz et al. 2010].

Former research on GVS in the field of computer graphics typically employs systems with two electrodes and consequentially only one axis that can be stimulated. This bilinear model is sufficient as long as the stimulation of only one axis is necessary, e.g. for simple racing games. However, as soon as the axis changes, e.g. by turning the viewer's head, a more complex stimulation model is needed. Especially the complex, unpredictable trajectories of real-world 360° videos would need a precise stimulation to truly adapt the vestibular system to the perceived virtual movements. Surrounding video material can provide sophisticated viewing experiences in VR. Such experiences allow the spectator to deeply immerse in the content and see the world through the eyes of the narrator. While the linear story unfolds around them, viewers can turn their heads in arbitrary directions. Similar to traditional linear video formats like TV shows or movies, the recording takes place from the position of the camera but the viewers are in control of their own gaze. Unfortunately, it is not yet possible to precisely stimulate the multidimensional movements that are presented in such experiences.

In this work we introduce omnidirectional GVS to VR. This paper is particularly focused on immersive videos and VR applications without body movements of the user. However, for simplification, the term VR will be used throughout the paper. In an experiment, we show the capability of 3-dimensional GVS to effectively mitigate CS while increasing the user experience. The applied GVS is implemented via an oculo-vestibular recoupling (OVR) stimulation model with two variations and compared to a control condition (CC). With the first variation, we followed related work and stimulate only the strongest visual main rotation axis per frame. For the second variation, we stimulate the exact rotation axes as experienced by the viewer by interpolating the

electrical currents. Our GVS is dynamically computed during runtime and considers the user's head position in real-time additionally to the scene rotation. We performed an experiment with real-world moving-camera 360° videos with depth information. For the experiment we developed our own custom-built video player that allows for fast and high-resolution video playback.

The contributions of the paper are as follows:
(1) Implementation of 3D GVS for VR by OVR stimulation.
(2) First time interpolation of electrical currents and evaluation of the effects.
(3) Comprehensive experiment to investigate the effectiveness of 3D GVS on CS and discomfort mitigation for immersive videos.

A.2 Related Work

A.2.1 Visual Mitigation Techniques for Cybersickness

Given the ubiquity of CS, a lot of works study the mitigation of CS in user-controllable virtual environments like games that usually give the user 6 degree of freedom (DoF) to explore the scene.

General concepts to reduce CS in VR scenes cover high frame-rate renderings, high quality tracking and reduced latency systems [LaViola Jr 2000, DiZio and Lackner 1997, Sherman 2002].

Recently, the use of techniques that manipulate the visual stimulus before it is presented to the VR user has increased. These techniques focus on the reduction of optical flow in the peripheral viewing area. Typically, either a complete masking of the outer viewing areas is achieved [Adhanom et al. 2020, Groth et al. 2021a, Fernandes and Feiner 2016, Lim et al. 2020] or the peripheral viewing area is covered semi-transparently [Groth et al. 2021a, Lin et al. 2020b, Budhiraja et al. 2017]. On the downside these methods often lead to adverse effect on the feeling of presence in the scene.

One important but least explored field in VR includes 360° videos and VR movies. This field of immersive presentation comprises a wide range from personal holiday videos to car configurators and blockbusters. Due to the sophisticated form of immersive presentations to view a scene, users are able to perceive real-world content more intense and realistic as with classic displays. Yet, as with other VR content, 360° videos provoke CS when movements are shown in the VR glasses that are not perceived by the user [Elwardy et al. 2020, Kim et al. 2018a]. The multimodal perception of stimuli, such as haptics, to enhance virtual experiences was also shown by Danieau et al. [2012]. Elwardy et al. evaluated CS for 360° videos in VR [2020]. The authors were particularly

interested in how the level of VR experience influences the outcome. As a result mainly participants with low experience in immersive media suffered from CS.

The high risk of CS for viewers exposed to 360° videos was also recognized by Kim et al. [2018a]. Their solution was a neural network that predicts a sickness score for VR videos. Consequentially, their idea is to warn about videos that are most likely to make users sick rather than mitigate CS in these videos.

Bala et al. [2018] started investigations on CS mitigation for real-world 360° videos. They used an independent background grid, a fixed FOV reduction and a combination to reduce CS in an experiment. According to their own statement, their results did not show any significant difference in CS due to the small number of participants [Bala et al. 2018]. In a later work with 360° videos they included more participants and only focused on the combined method (independent background grid and reduced FOV) and were able to show a decrease of CS [Bala et al. 2020].

Recently, Groth et al. [2021b] demonstrated that CS in 360° videos is mitigated by unobtrusive modulations of the visual presentation. Their idea was to reduce motion in the peripheral area and thus minimize the sensory conflict. In their experiment they gaze-contigently modified the peripheral visual field by either blurring or opaque occluding eccentric view areas. The experimental results show that both techniques are effective to mitigate CS in pre-recorded VR content with the opaque occlusion delivering the best results.

For immersive videos, minimizing the conflict between visual and vestibular sensations is crucial to prevent CS and increase the overall well-being [Reason and Brand 1975, Danieau et al. 2012]. Particularly interesting are the efforts that adapt the vestibular stimulus without removing any information from the visuals.

The work of McGill et al. [2017] highlights how the adjustment of vestibular stimuli affects CS and immersion. In an in-car experiment the movements of a 360° video were synchronised with the real vehicle movements. Unfortunately, no best practice for driving simulator-based presentations were found. Still, as one result people got more sick in a moving car with VR applied, which underlines the sensory conflict theory.

The video format (monoscopic, stereoscopic) and audio format (stereo, spatialized) was found to not influence the users' feeling of presence and CS for 360° videos [Narciso et al. 2019]. In contrast, gender is a significant variable that should be considered for VR scenes that display 360° videos [Narciso et al. 2019].

A.2.2 Electric Stimulation

GVS is the direct and safe stimulation of the vestibular system by electrical currents [Utz et al. 2010]. It is a subset of transcranial direct current stimulation (tDCS), a method often used in medicine, e.g. for therapy or tumor treatment [Brunoni et al. 2011, Nitsche et al. 2009, Song et al. 2015]. While tDCS is applied by electrodes attached to the subject's scalp, the electrodes for GVS are usually attached to the mastoids behind the ears to specifically stimulate the vestibular canals. When standard procedures are followed, tDCS and GVS are safe, noninvasive and low-cost techniques that have been extensively studied and applied in practice [Utz et al. 2010].

The most commonly used form of GVS is bilateral, bipolar stimulation where two electrodes are placed on the mastoids. With this GVS, the sensation of a strong roll rotation and a weak yaw rotation towards the cathode is elicited [Fitzpatrick et al. 1994, Day et al. 1997].

The name *galvanic stimulation* is based on the research of Luigi Galvani from 1791 who conducted early experiments on animal electricity [Galvani 1791]. Around the same time, Alessandro Volta, a rival and critic of Galvani's work, performed the first human GVS experiments on himself. He felt his head spinning and reported to hear a noise, which is probably based on the high voltage he used. Today's applications of GVS use only very weak currents of no more than 2.5 mA. These applications of GVS are often found in medical research and treatment. Positive effects of electric stimulations have previously been reported for motor, visual, somatosensory, attentional, vestibular and cognitive functions as well as multiple neurological and psychiatric disorders [Utz et al. 2010].

However, medicine is not the only field where GVS has been applied. In the work of Byrne et al. [2016], GVS is explored as a design tool for vertigo games. In their application *Balance Ninja* two players try to unbalance each other by controlling a GVS device connected to the other person. Based on the positive response of the participants, the results suggest considerable potential of GVS as a game design tool.

The insights of further research support the hypothesis that GVS can be a successful tool not only in medicine, but also for entertainment and VR.

Already in 2005, Maeda et al. [2005] revealed that visually induced vection, i.e., the illusion of self-motion, can be enhanced with GVS. In their experiment participants were stimulated with GVS while watching white dots move on a screen. Thereby, the intensity of the movements and the GVS were matched. As a result participants reported a strong feeling of vection by GVS as long as the sensory conflict was low.

Aoyama et al. [2015] investigated the enhancement effect of GVS with counter-

currents on human acceleration perception. In their experiment they used a 4-pole GVS with a capacitor-resistors circuit model in VR. They found that counter-current stimulation enhances the strength of the acceleration sensation whereby the duration of the stimulation correlates with the perceived strength.

A.2.3 GVS for Cybersickness Reduction

In 2007, GVS was applied to counter CS in VR for the first time [Reed-Jones et al. 2007]. In multiple sessions the visual and vestibular stimulus was either turned on or off while participants performed a driving simulation. In these early VR experiments the GVS was activated by hand by the researchers. In conclusion, the results already gave a positive outlook for CS reduction when GVS is applied.

Over the years several manuscripts using bilateral GVS were published but mostly do not address cybersickness[Fujimoto et al. 2016, Sra et al. 2017, Moore et al. 2011, Maeda et al. 2005].

One of the most extensive experiments with a bilateral GVS was presented in 2019 by Sra et al. [2019]. They constructed a 2-electrode GVS device that is hung around the neck. With the device and a mobile phone companion app, the vestibular feedback can be predefined on one stimulation axis for a given scene. The results of their experiment show significantly lower levels of CS while the immersion is higher when the GVS is applied.

Typically, the stimulation of the vestibular system is done with electrical currents. Also, bone-conducted vibration can be used as a stimulus as shown by Weech et al. [2018] who reduced CS with this GVS alternative. In an experiment with a VR head-mounted display (HMD) and a projection cave, they got positive results for sickness reduction. The vibrations achieved the same effect when they were randomly induced in comparison with time-coupled stimulations to angular accelerations. Still, this technique remains uncommon in VR research and has a lower impact as electrical stimulation [Weech et al. 2020].

Later, the same researchers moved their focus to noisy GVS as a promising technique to stimulate the vestibular system when the visual movements in the scene are unknown. In a VR experiment they focused on sickness severity during and immediately following noisy GVS. They state that noisy stimulation is able to reduce CS severity for intense VR content, while it has no effect for moderate content. Even more interesting for GVS research is the result that a rapid re-adoption of the vestibular system to its normal state happened after GVS exposure. As no effect persisted after the experiment, this is a positive outlook for the overall safety of GVS.

In the work of Cevette et al. [2012], the oculo-vestibular recoupling (OVR) stimulation model is presented. By using a 5-electrode GVS device they were able to stimulate the three main rotation axes (yaw, pitch, roll) individually. In their experiment the visual stimuli were presented on a screen in front of the participants who controlled a flight simulator with a joystick. In their approach they synchronized the rotational speed and main direction of the visual movements with the GVS induced vestibular sensation of the participants. Their results show that OVR based stimulation significantly reduces simulator sickness in a cockpit flight simulator.

In our work, we build upon the work of Cevette et al. and use OVR-based GVS to re-adjust the vestibular perception of VR users to their visual impression. With a 5-electrode GVS device we apply two different methods. First, we stimulate only the strongest visual rotation axis comparable to former work. Second, we stimulate the exact visually perceived movements by interpolating GVS currents between electrodes.

A.3 Methods

In a VR experience, we elicit a feeling of self-motion that corresponds to the visual movements of the virtual scene. The movements are visually induced by a 360° video and vestibularly stimulated with a GVS device. In the following we describe the implementation necessary to achieve this particular experience.

A.3.1 GVS

For the GVS we apply the OVR stimulation model proposed by Cevette et al. [2012]. With the five electrode OVR model all single Euler rotations (yaw, pitch, roll) can be stimulated by one negative and one positive pole on defined electrode positions [Cevette et al. 2012].

As proposed, we attach two electrodes to the left and right mastoid (LM and RM), one electrode to the forehead (F), one electrode to the neck (N) below the hair line, and one ground electrode further down the neck (GND). Yaw is stimulated with the two electrodes on the mastoids. Pitch and roll on the other hand are defined by two different electrode pairs (see Table A.1). The coordinate system of the camera is left-handed and the Euler angles are defined in the order: yaw, pitch, and roll. All recorded rotations are given in world coordinates and are further transformed with the transformation of the VR headset (see details below).

In theory, any intermediate axis is represented by a linear combination of the principle components of the Euler rotations. To interpolate main rotations with OVR, the

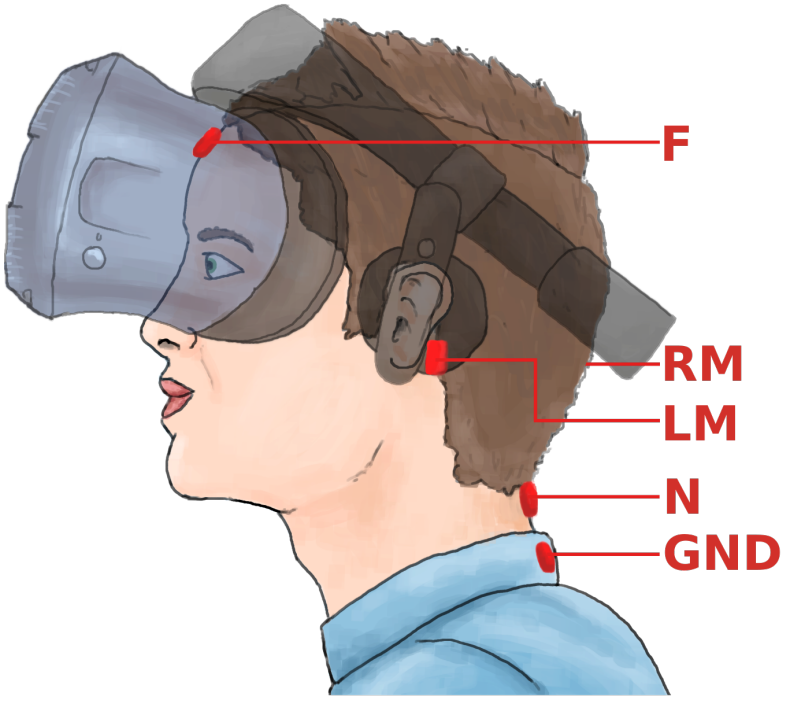

Figure A.2: Placement and naming of the five electrodes (red) for the GVS. RM is placed on the right mastoid similar to LM.

relevant electrode pairs have to be carefully chosen to prevent negative influences on the current flows. Accordingly, one electrode should not be considered as a negative pole for the stimulation of one rotation axis while at the same time it serves as a positive pole for another rotation. To date, such a linear combination of the rotation axes does not exist in the literature. Therefore, we conducted an internal study among the authors to empirically verify the validity of a interpolated stimulation. In the internal study we tested all possible combinations of the single main rotation axes with OVR stimulation. Note that the combinations especially consider all potential electrode pairs to stimulate one particular axis (cf. Table A.1). In this pre-experiment the participants were standing with their eyes shut and did not know what stimulation they were receiving. We evaluated the combinations through body tilt and perceived stimulation direction via oral feedback. Our results show that a linear combination of the single rotation axes is capable of representing all intermediate angles when the electrode pairs for pitch and roll are chosen correctly. Following our theory, the intermediate angles are stimulated by combined currents of the respective Euler angle components (yaw, pitch, roll).

Table A.1: Directional stimulation used by OVR to cause a particular motion perception [Cevette et al. 2012]. The electrode naming follows Fig. A.2. Annotation for current stimulation: anode to cathode.

Intended Motion	Current Flow Direction
yaw right	LM to RM
yaw left	RM to LM
pitch forward	RM to F || LM to F
pitch backward	F to RM || F to LM
roll right	N to LM || RM to N
roll left	LM to N || N to RM

This interpolated currents per electrode are defined by function f (see Equ. A.1–A.4). Here, the name of the electrodes follows the previous convention.

$$f_F = \beta \tag{A.1}$$
$$f_{LM} = k + H(\gamma)sgn(k)\gamma \tag{A.2}$$
$$f_{RM} = l + H(-\gamma)sgn(l)\gamma \tag{A.3}$$
$$f_N = -H(\gamma)sgn(k)\gamma - H(-\gamma)sgn(l)\gamma \tag{A.4}$$

As input, all functions take the three components of the Euler rotation defined by α for yaw, β for pitch and γ for roll.

The variables k and l are given by:

$$k = \alpha - H(-\beta) * \beta$$
$$l = -\alpha - H(\beta) * \beta$$

The Heaviside step function $H(x)$ is defined as usual. The sign function $sgn(x)$ is defined by:

$$sgn(x) = \begin{cases} 1, & x > 0 \\ -1, & x \leq 0 \end{cases}$$

For the stimulation of the intermediate angles we consider every axis by its weight and in a clear order. We assign a maximum current per user to consider individual

tolerances. This maximum current value per electrode per participant is determined during the calibration phase (described in more detail in Sec. A.4.5). In the experiment we take care that the currents never exceed this value. Also, the stimulation per frame depends on the angular velocity per axis in relation to the pre-defined maximum speed of the video.

In the experiment, we only stimulate the movements that are actually seen by the VR users in their FOV. To achieve this dynamic stimulation, we consider the pre-recorded movement of the recording device and the real-time head transformation of the VR glasses. The camera rotations were recorded simultaneously with the video by a gyroscope that is built into the 360° camera. During the presentation, these rotational movements are evaluated and transformed in real-time with respect to the head movements of the observer.

A.3.2 Video Display Framework

For the experiment, we custom-built a video player based on OpenVR which is able to decode and present the single 6k stereo frames (one 6k image per eye) at the recorded video frame rate (30 FPS) and render the output based on the FOV with the frequency of the HMD (90Hz). We did not experience any frame drops throughout the experiment. Furthermore, we required real-time metadata of the content for the GVS, namely the current camera motion properties and the head transformation.

Our video player uses the full potential of the GPU for the video decoding as well as the rendering of the video.

A.4 Experimental Design

We conducted an within-subjects experiment with three sessions per participant to explore the effect of GVS on the perception of 360° videos. While the visual stimuli were the same for all sessions, the GVS was altered. In the control condition, the GVS was inactive at zero current. In the other two conditions the GVS either stimulated the strongest main rotation axis present in the FOV or the actual rotational velocities as seen by the observer. We counterbalanced the order in which the three conditions were shown to the participants. Furthermore, a 48 hours recovery time was maintained between sessions to avoid carry-over effects. The experiment was reviewed and approved by the corresponding ethics committee under the identification number D_2021-06.

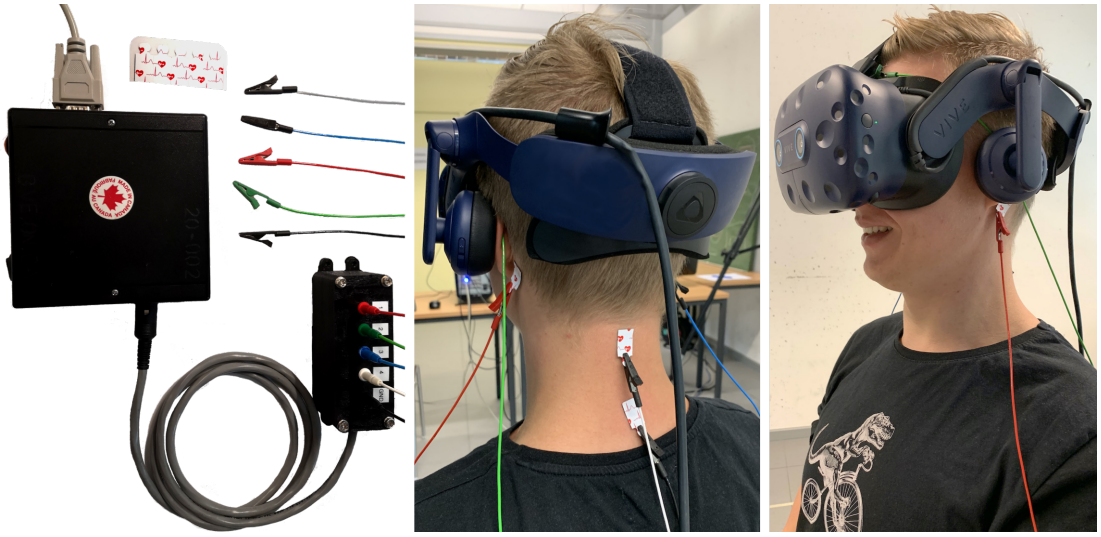

Figure A.3: The GVS device, adhesive electrodes, connector box and electrode cables. On the right, we show its connection with the VR glasses.

A.4.1 Stimuli

Visual Stimuli

In every session, we presented the same 360° video with camera motion in a HMD (see Figure A.3). The total runtime of the video is 10 minutes. It shows scenes of walking through a park (moderate movements) followed by a bicycle ride on an uneven road through a forest (fast movements). We chose two different scenarios to determine how GVS affects different modalities. Both scenarios have a presentation duration of five minutes and were played without a gap. In our experiment, participants had no control over the video except that they could change their viewing direction by head movement.

Cinematic panorama videos are known for their high probability to cause CS [Groth et al. 2021b, Tauscher et al. 2020]. In our videos, the sickness induction is supported by the fast camera movements. These movements increase the sensory mismatch that the seated participants perceive [Reason and Brand 1975].

Vestibular Stimulation

We investigated three different conditions: two conditions with GVS to investigate their effects on the perception of a 360° video in VR and a control condition as a comparison:

– CC: control condition without stimulation. The apparatus is still properly connected (electrodes, GVS), but the stimulation is kept at 0 current. We do not

inform the participants that one session has no stimulation to not influence their opinion.

– SA: strongest axis condition with GVS. In this condition only the strongest main rotation axis in the FOV is stimulated. This technique is inspired by the work of Cevette et al. [2012] and also serves as a comparison to the GVS interpolated condition.

– IN: interpolated condition with GVS that stimulates the exact rotation axis visible in the FOV. This condition uses the axis combination mapping from Sec. A.3.1. Our goal is to evaluate whether a precise stimulation significantly differs from the strongest axis condition.

The stimulation in both GVS conditions depends on the pre-recorded rotational velocities of the camera in the video. We transform these movements at run-time with the HMD transformation to truly stimulate the rotations as seen by the VR user. Therefore, although the movements of the recording device are pre-captured, the stimulations in our experiment were entirely dynamic and defined at run-time by the participants' posture. We obtained the angular velocities of the camera in the scene by the gyroscope that is built into our recording device. The movements were therefore recorded along with the video.

A.4.2 Apparatus

We recorded the 360° videos with an Insta360 Pro camera [Insta360 Pro] at $6k$ stereo resolution (6400 x 6400 px) and 30 FPS. The videos are encoded with the HEVC codec (H265). Gyroscopic data was recorded every 3ms. The 360° camera was mounted on a snowboard helmet to record the scenes since we required fast and flexible movements with hands-free control. The aperture was custom-built and strengthened with extra padding and stabilizers to control the weight of the camera. The final apparatus can be seen in Figure A.4.

For the experiment we used a commodity HTC Vive Pro HMD with a FOV of 110° and a frame rate of 90 Hz. The resolution of that HMD is 2880 x 1600 px and therefore renders the full resolution of the video in the respective FOV. All video footage is played with the corresponding audio.

Our GVS device from Good Vibrations Engineering [Vestibulator] (Canada) has four current outputs and one ground electrode (see Figure A.3). The device offers a maximum current of 2.5 mA per electrode. The latency between sending a signal in

the application and stimulation to the electrode is less than 15ms. In the experiment, the GVS was controlled by our experiment presentation software. For safety reasons, the device is protected against any unanticipated power transmission by design using battery power and information transmission via air-gapped fiber optics. We use adhesive electrodes with a size of 13x16 mm to attach the GVS to the participants.

A.4.3 Participants

A total of 47 participants completed all three sessions (18 females, Age range = 19-52, Avg age = 24.79, SD = 5.72). Participants were compensated with 30€. The order in which each of the three sessions took place was counterbalanced, with each participant receiving a different order. The experiment followed a full within-subjects design.

For the analysis, we divided the participants into two disjoint groups based on whether they were negatively affected by the virtual experience. The group with individuals perceiving the 360° video as unpleasant consists of 30 participants (13 females, Age range = 19-52, Avg age = 24.8, SD = 6.41). 17 participants (5 females, Age range = 21-36, Avg age = 24.76, SD = 4.24) were not affected by the virtual scene.

A.4.4 Measurement

We used the simulator sickness questionnaire (SSQ) [Kennedy et al. 1993] and Slater-Usoh-Steed (SUS) presence questionnaire [Slater et al. 1998, Usoh et al. 1999, 2000] for participant feedback. Both questionnaires require self-assessment of the participants based on how they feel and how they perceived the scene, respectively. The SSQ is an effective tool to measure CS for 360° videos in VR as demonstrated by Singla et al. [2021]. Following common procedure, we let participants fill in the SSQ twice, before and after each session of the experiment, to counteract different daily conditions. The total sickness score as well as the corresponding subscores of the SSQ are calculated according to the original procedure of Kennedy et al. [1993]. The SUS presence questionnaire was filled out once per session, immediately after the experience.

During the experiment we also asked the participants to press the trigger button on the Vive controller every time their comfort feeling got worse, and to press the touchpad when their well-being increased. Based on these responses the participants' individual level of discomfort is calculated. We specifically consider discomfort to not only refer to symptoms of CS but also to any negative effects of the GVS like scratching from the electrodes. Furthermore, the head movements of the participants were recorded.

Effective GVS can be verified by the physical tilt of the body in the direction of

Figure A.4: Left: Helmet carrying the 360° camera to capture the recordings for the experiment. Right: Camera in action.

the stimulation. However, since in the brain the visual information overwhelms the vestibular information for inconsistencies, a body leaning only occurs with closed eyes or matching information. For evaluation, our participants were asked to rate the strength of the perceived curve lean in the video after each session. The methodology follows the SUS presence questionnaire.

A.4.5 Procedure

The experiment was divided into three sessions with one experimental condition each (CC, SA, IN). The sessions for each participant were conducted on different days with a pause of two days between sessions to avoid any carry over effects [Groth et al. 2021b, Lin et al. 2020b, Ebrahimi et al. 2014, 2015].

The experiment started with an informed consent and a demographic questionnaire including factors influencing susceptibility to CS. Also, at the beginning of every session the first SSQ [Kennedy et al. 1993] was filled. In all sessions (including the sham session) we then attached the five electrodes. With the electrodes attached we performed a calibration of the participant's individual galvanic stimulation level in the first session. The calibration is necessary to find the maximum current per person that is still comfortable. We found that it is crucial to specify suitable maximum currents as the sensitivity to electric currents highly varies. We performed the calibration in two steps: First, the current was gradually increased with a yaw stimulation at the two electrodes

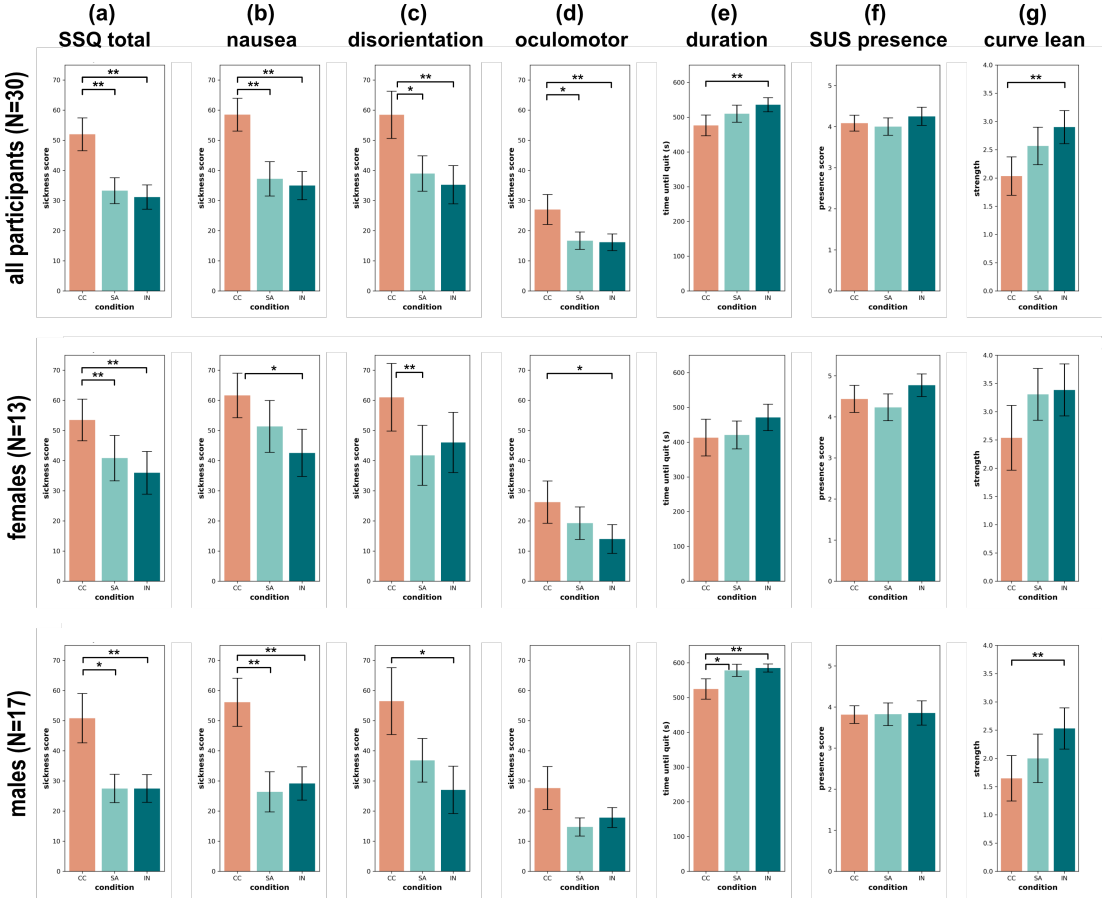

Figure A.5: Averaged SSQ scores, durations and presence results for the control condition (CC), strongest axis condition (SA) and interpolated condition (IN). Error bars represent the standard error of the mean (SEM). **(a)** SSQ results for the total score. **(b-d)** results for each of the SSQ subscales. **(f)** SUS presence questionnaire results. **(e)** duration people were willing to spend in the virtual environment. **(g)** self-indicated score of how much participants felt they were leaning into the curves during the video. Significant results are denoted by '**' ($\mathbf{p \leq 0.016}$, Bonferroni-corrected for multiple comparisons) and '*' ($\mathbf{p \leq 0.05}$).

behind the ears. In our experience, these are the most strongly perceived electrodes and with the mapping of the interpolated condition the highest currents will occur here (cf. Equ. A.1–A.4). The participants were asked to tell when they noticed any unpleasant feeling from the electrodes. As soon as we found a suitable maximum current, we cross checked the value in a second step where we toggled the current from zero. This sudden current change is typically perceived stronger than a slow change. When the participants still found the chosen current to be tolerable, the maximum was found. Otherwise, we repeated the second step with a lower current. The maximum current served as a scaling factor and was used for both GVS sessions to ensure comparability.

In the experiment, the participants were asked to watch a 360° video in VR and informed that they can quit the experiment at any time in case of severe negative feelings. They sat on a chair and were allowed to freely explore the scene but to remain still with their bodies to avoid negative effects on their immersion. They were asked to press certain VR controller buttons when they noticed any positive or negative change of their well-being. The participants watched a 360° video that was divided into two stages graded by the severity of motion in the video (walking, biking), each with a duration of five minutes (cf. Sec. A.4.1). In total, the participants spent up to ten minutes in the virtual environment but could voluntarily quit the experiment at an earlier point in time. After the experiment, participants filled in the second SSQ, the SUS presence questionnaire and, in case it was the last session, were asked if they noticed any difference between the sessions and received information on the actual goal of the study.

A.5 Results and Analysis

For the analysis of the experimental results we used factorial mixed repeated-measures ANOVAs with condition as within-subject and gender as between-subjects factor. As post-hoc tests we performed pairwise two-sided dependent t-tests for repeated measures with Bonferroni-correction. The time-series data was analyzed with cluster-level permutation tests.

We found that in most cases the participants could be categorized into two groups: people for whom the video had no effect (SSQ score and discomfort close to 0) and those who were strongly affected, at least without GVS. We separated these two groups for the analysis in order to make more concrete statistical statements and to understand the impact of GVS on these two different groups. The total SSQ score served as the classification factor for the groups. When this score was below 20 without GVS, the

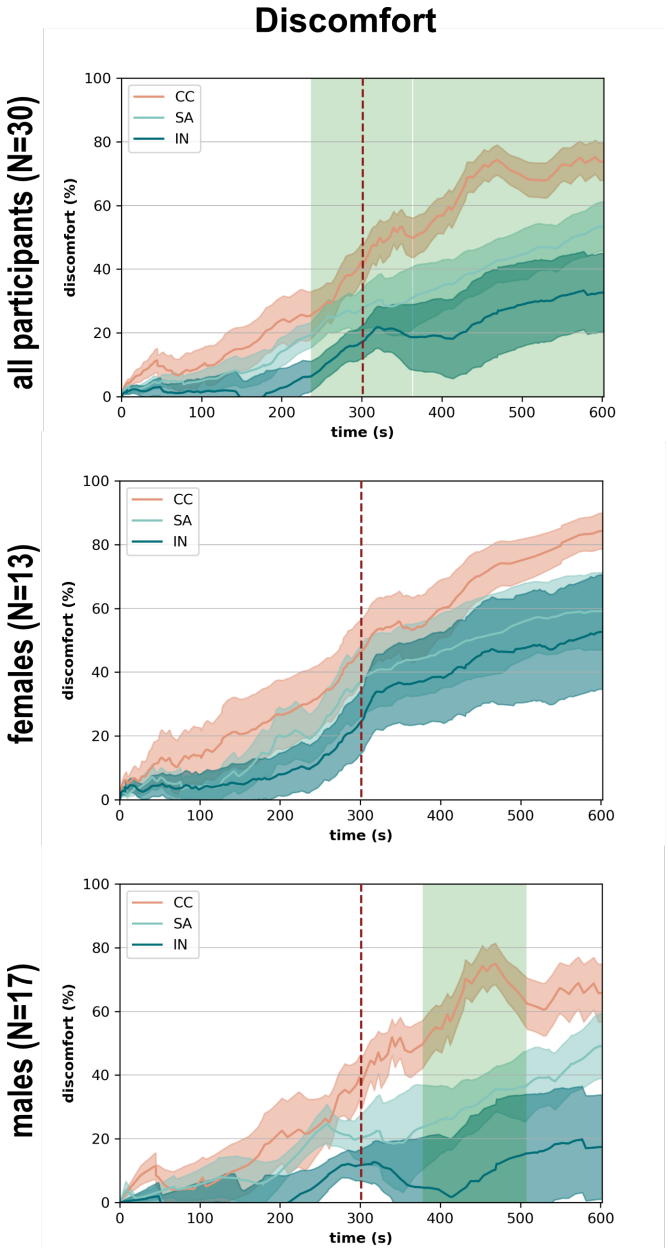

Figure A.6: Relative discomfort over the time of the video ($N_{all} = 30$, $N_f = 13$, $N_m = 17$). Normalized per participant by their highest score in all sessions under consideration of the session times. The shaded areas around the mean line denote the SEM. Sections with significant differences are highlighted with a green background ($\mathbf{p \leq 0.05}$). The red dashed line marks the scenario switch from walking to biking.

participant was assigned to the first group (no effect) and vice versa. For the analysis, we particularly focus on the group of participants for whom the video had a strong effect, since the effectiveness of GVS can only be considered when the video caused CS in the first place. For all statistics that account for the unaffected participants this is explicitly stated. However, we also contemplate that GVS can still have a negative effect when people do not get CS from the VR scene. The results of this analysis are described below.

Gender has a strong influence on the susceptibility for CS according to previous research [Narciso et al. 2019]. Consequently, our analysis includes a separate analysis of gender.

During the experiment, 20 participants chose to end one or more of the sessions early because of severe sickness symptoms (42.6%). Most of the terminations were observed during the sham condition with the GVS device deactivated (control: 31.9%, strongest axis: 25.5%, interpolated: 25.5%), which is an early indicator that the participants experienced the strongest symptoms in CC. As expected, most participants opted for a termination already during the first session (S1: 38.3%, S2: 25.5%, S3: 19.1%). Therefore, it is most likely that a learning effect occurred. We used a counterbalanced design to act against this learning effect.

Figure A.5 illustrates the results of the SSQs, the SUS presence questionnaire results and the average times for participants to end a session. The SSQ was analyzed for its total score and the three subscores of nausea, disorientation and oculomotor [Kennedy et al. 1993].

Overall, the results show the same trend for almost all questionnaires as well as the discomfort data: CC is perceived as the most sickness inducing. Both GVS conditions were equally able to reduce the sickness score. For the interpolated 3D GVS we can also see some positive side effects unique to this session: an increased time until participants terminated the session and the highest feeling of comfort.

The analysis shows a significant main effect on the SSQ total score (Figure A.5a) for condition ($F(2, 87) = 6.12$, $p = 0.0033$). Pair-wise dependent t-tests confirm the significant difference to appear for both GVS conditions when all participants are considered. The difference to CC for the interpolated 3D stimulation was most significant ($T = 3.76$, $p = 0.0008$), but also the strongest axis condition achieved notable results ($T = 3.1$, $p = 0.0043$). A comparable trend as for the overall results can also be found for the single subscores of the SSQ. The nausea subscale of the SSQ (Figure A.5b) presents a significant main effect for condition ($F(2, 87) = 6.0$, $p = 0.0036$). While the interpolated condition lowered nausea for both genders ($T > 2.2$, $p < 0.05$), the

strongest axis condition varied highly with gender and was only for males significant in difference ($T = 3.62$, $p = 0.0023$). For disorientation a significant reduction of this SSQ cluster was achieved by GVS ($F(2, 87) = 3.42$, $p = 0.0371$). Pair-wise dependent t-tests confirm this difference for both, the strongest axis condition ($T = 2.4$, $p = 0.0233$) and the interpolated condition ($T = 2.76$, $p = 0.0099$). However, when the genders are considered separately, we see a substantial variation in the effect. While for men, only the interpolated 3D GVS was able to significantly reduce disorientation ($T = 2.31$, $p = 0.0343$), it was the opposite case for women. Females were least affected by disorientation effects with the strongest axis condition applied ($T = 3.01$, $p = 0.0109$). For the oculomotor effects (Figure A.5d) a significant main effect for condition is shown by the factorial mixed ANOVA for all participants ($F(2, 87) = 2.77$, $p = 0.048$). Again, this SSQ subscale is highly affected by gender. For men, the interference of the oculomotor effects is not significantly changed by the use of GVS. On the other hand, for women the interpolated stimulation (IN) was able to reduce these effects by a meaningful amount ($T = 2.27$, $p = 0.0425$). The strongest axis condition had no effect for both genders. Note, that in general the oculomotor scores are quite low compared with the other subscales of the SSQ.

The results of the duration participants were willing to spend in the 360° video before they chose to end the experiment are shown in Figure A.5e. The two conditions had very different effects on the duration participants' stayed in the scene. While the interpolated condition made the VR users stay significantly longer in the scene ($T = -3.2$, $p = 0.0033$), the strongest axis condition had no such effect. However, the gender is of great importance here: men were willing to spend more time in the virtual scene with either GVS method applied (strongest axis: $T = -2.23$, $p = 0.0408$; interpolated: $T = -2.78$, $p = 0.0134$). Female participants on the other hand were less affected and without a statistically significant effect for duration. Still, on average, women spent 59 seconds (14%) longer in the 360° video with interpolated GVS. For the results of the SUS presence questionnaire, all sessions led to comparable perceptions of the scene. The use of GVS did not notably change the users' feeling of presence ($F(2, 87) = 0.38$, $p = 0.6882$). Also gender showed no effect here. While these presence results remained on the same level, the participants' score for curve leaning did vary. The participants reported scores for physically leaning into the curves they visually observed was significantly improved with interpolated 3D GVS ($T = -2.7$, $p = 0.0114$). These general results were mostly due to the male participants ($T = -2.76$, $p = 0.0139$) and insignificant for women. The strongest axis condition was not able to achieve a statistical improvement of the curve leaning.

Figure A.6 shows the results of the overall discomfort over time of the experiment as measured by the controller feedback. The discomfort data is normalized per person by the keystroke responses during the experiment. This self-indicated feeling of discomfort decreases or increases by one unit each time the corresponding keys are pressed. The highest value reached in one of the three sessions weighted by the duration spent in the session is considered to be the maximum global discomfort level for this person across all sessions. The normalization is therefore relative and the maximum value may be perceived differently by each individual. As expected, the discomfort of the participants increases over the time of the experiment. However, the results of the GVS sessions are distinctly different from SHAM and increase at a much slower rate. Especially in the second part of the video with strong movements, a differentiation of the results is noticeable and thus the scenes trigger considerably less discomfort when GVS is applied. The statistical analysis confirms these results and shows a significant difference (shaded areas denote standard error of the mean in Figure A.6) for more than half of the video duration and especially for the second part (strong movement).

Although participants felt more comfortable in the GVS session, no difference in their exploratory intention could be determined by their head movement intensity.

For the group of participants that was unaffected by the video, GVS induced no negative effect. For this group all statistical results are insignificant. This finding suggests that VR users unaffected by CS also do not experience adverse effects by GVS. As a result, the applicability of GVS to the general public is greatly increased.

A.6 Discussion

A common way to describe 3D rotations is via the three main rotations of the axes, known as roll, pitch and yaw. Any rotation can be achieved by a concatenation of three rotations around the principal axes (Euler angles). The commonly used bilateral bipolar GVS stimulates the horizontal and vertical canal of the vestibular system simultaneously. As a result a yaw rotation towards the cathode and a simultaneous roll on the same side are perceived. In this stimulation model the canals are always simultaneously stimulated and cannot be disconnected. A pitch rotation is not induced in this model.

Cevette et al. [2012] proposed the OVR model which allows for a separate stimulation of roll, pitch and yaw. With five electrodes and well-defined stimulation pairs, the rotations around the principle axes are stimulated one at a time. We adapted this type of stimulation (cf. strongest axis) for VR use.

However, rotations rarely occur exactly around one principal axis, but rather in the

broad spectrum in between. If these rotations are projected onto the nearest principal axis, as in SA, the remaining information is lost. This could also have an effect on the perception by the vestibular system. In theory, the optimal case is a stimulation that relays all information to the vestibular system in exactly the same way as it appears visually. With the second GVS condition (cf. interpolated condition) we present an OVR mapping that we hypothesized to be capable of stimulating all rotational movements of the visual field. The electrode pairs responsible for the stimulation of the principal axes are combined in the mapping, taking into account the correct polarization. Thereby, the influence of the rotational components is weighted by their strength. Hence, if the visual field moves in yaw direction while pitching twice as fast, the mapping stimulates one part of the yaw electrode pair and two parts of the pitch electrode pair.

A.6.1 GVS vs. Sham Stimulation

The experimental results show a clear indication that both GVS techniques successfully reduce CS. In the following, we focus on the participant group that suffered CS symptoms during the experiment. In all sessions the SSQ score approximately dropped by half when GVS was applied. This positive effect occurs not only for the total SSQ score but also for the SSQ subscores of nausea, disorientation and oculomotor effects. But CS is not the only negative effect that was reduced by galvanic stimulation of the vestibular system. In the experiment, the participants indicated all changes of their feeling of comfort via button presses, which allows us to derive a general progression of discomfort over the time of the experiment. Based on this data, a significantly higher feeling of comfort was achieved with the use of GVS. This substantial improvement was present during almost the entire scene with fast movements, but also in large portions of the walking scene. While it stands to reason that the mitigation of CS also contributed to an improvement in the overall comfort feeling, the results also show that no other negative effects arose from the GVS itself. Such negative effects would include, e.g., itching at the electrodes.

These general results show VR to be a valid field of application for the GVS technology. A significant benefit for the reduction of CS is achieved regardless of which of both GVS method is used. The integration of GVS technology into future VR glasses could permanently counteract CS in VR environments. Thereby, the visual stimulus remains unchanged and the vestibular sensation adapts to the virtual experience. GVS is unobtrusive and is effective even with current levels that are only slightly or not at all noticeable to the user. However, this unobtrusiveness requires individual calibration. Unfortunately, calibration has often been missing in previous work. This work

has shown the importance of customizing the current intensity to the user's personal preference.

A.6.2 Strongest Axis Stimulation vs. Interpolated Currents

While both GVS conditions outperformed CC, they were not equal in their overall effect. Although CS was at the same level in both GVS sessions, the participants chose to stay longer in the scene with interpolated stimulation. This significant increase of the session duration is unique to the interpolated condition in our experiment. From the results of the SUS questionnaires we could not infer a change in the feeling presence with GVS. Interestingly, most of the participants mentioned a different feeling between sessions in the open interview after the experiment. Without any knowledge of the actual procedure, they mentioned to "feel the movements in the video, even when the head is static" and that "[the GVS sessions] felt like a realistic dream, instead of a bad video". A difference is also evident in the results for perceived curve leaning. For the IN session, participants perceived a significantly stronger desire to physically lean into curves they watched in the HMD. Given these subjective user reports, it should not be rejected that GVS can, in a certain way, intensify the presence of VR users. However, further research is needed to understand this connection in detail.

As described above, we also investigated the progression of participants' overall comfort levels during the three VR sessions. Both 3D GVS methods were shown to be effective techniques to reduce discomfort. In a direct comparison of SA and IN, we can see that IN is always slightly below SA on average. However, a significant difference between the two sessions is only present in the last third of the walking sequence. Furthermore, for male participants, the flatter progression of IN shows a diverging development towards the end of the video.

Altogether, the presented interpolated mapping based on OVR is a promising method to compensate CS and improve the overall user experience. With this method, users were willing to stay longer in the scene while feeling more comfortable. With interpolated 3D stimulation, we aimed to represent all rotational movements of the visual field equally on the vestibular canals, thus significantly minimizing the visual-vestibular mismatch. We were particularly interested in how precise stimulation of intermediate axes differs from the method of strongest main axis stimulation. With the aforementioned differences, using interpolated stimulation is particularly useful whenever CS improvement is not the only objective. However, since all aspects of user experience are a fundamental part of almost every VR scene, we are convinced that the presented interpolated 3D mapping for precise stimulation of 3D rotations could find general ap-

plicability.

Nevertheless, CS and discomfort were not completely prevented. On the one hand, this may be attributed to the uncompensated linear accelerations. Linear accelerations cause CS to a lesser extent than rotational movements and the movements in our video had mostly a constant velocity [Groth et al. 2021b, Kim et al. 2021]. Still, residual accelerations may have caused some CS in the participants. Furthermore, it is likely that due to a general tendency to motion sickness in real life, some participants also experienced CS when the scenarios felt real to them.

A.6.3 Influence of Gender

VR experiences are subject to a general gender bias as noted in previous research [Munafo et al. 2017, Grassini and Laumann 2020]. In line with these findings, we found major differences in the impact of our GVS methods when comparing men and women. While GVS is certainly able to reduce CS independent of gender, the scores for men exposed to GVS are significantly lower. This is particularly true for SA. Without GVS, both genders experienced the same level of CS. The differentiation is even stronger for the nausea subscores of the SSQ. GVS was able to drastically reduce nausea, especially in men, after similar baseline results in SHAM. It is also worth mentioning that the CS triggered by the video in our experiment was perceived to be equally strong for men and women. The time the participants were willing to spend in the virtual environment until they became too uncomfortable is again affected by gender. Male participants were willing to stay significantly longer in the scene as soon as GVS was applied. For females, there was no significant change in duration. These results are especially interesting considering that the males perceived lower CS despite the longer duration of the experiment. The comfort levels also evolved more positively for male participants than for females using the GVS device. In particular the second part of the video (with strong movements) made male participants feel significantly more comfortable with GVS. It is likely that these results are related to the already lower CS scores of the males.

Consequently, when comparing men and women for moving camera 360° videos, a stimulation of the galvanic system may be most effective for male VR users. Also, for female participants, we found a significantly positive impact of the overall experience, levelling the field in terms of gender-based CS susceptibility. However, these results are given under consideration of the relatively small groups of participants per gender and may need further verification.

All of the results above consider the group of people for whom the video caused negative effects. When people use VR without any signs of CS, no mitigation is necessary.

However, we investigated whether GVS would have a negative effect in this case. With the use of GVS, our results show no increase in CS severity for these individuals and also the comfort levels remain the same. With these results, we conclude that regardless of the user, GVS has either a positive effect when needed, or no effect otherwise.

A.6.4 Impact of the Scenarios and Generalizability

In the stimulus video, two scenarios are shown. The first sequence is a walk through a park in summer. In this scenario the camera moves around a few curves and there are also slight head turns. The speed is moderate but constant. The second part of the video is shot from the perspective of a mountain biker who is riding at high speed down a forest path. The path itself contains a lot of curves and fast head turns of the rider in pitch and yaw direction. Roll is represented mainly by the curve leaning.

Unfortunately, it is hard to determine how much each scenario contributed to the resulting CS. For this separation, another SSQ would be required after the first sequence. However, any interruption of the video breaks the presence of the participants. Consequently, we compare the two scenarios mainly on the basis of the discomfort measurements. From the oral feedback of the participants, biking was, as expected, perceived to be significantly more sickness-inducing than walking. When considering that the participants were already affected by the walking sequence at the beginning of the second scenario, the discomfort level increases comparably in both parts of the video. The main difference is in the increase of the values: in the walking scenario, the discomfort score first grows slowly and gradually increases from halftime onward. In the last section of the walking scenario, where the GVS achieves significant improvements, the video shows a sequence of walking down stairs. This part was perceived as very unpleasant by most of the participants, but could be compensated effectively with GVS. In the biking scenario, the discomfort level increases constantly, but with distinct peaks where the movements in the video were very fast. These peaks are absent in the GVS sessions and the discomfort level increases only slightly in the second part.

A generalization for usability of GVS for other 360° videos is expected, given that our video shows many natural rotations in all directions and is presented for different scenarios. However, such a generalization does not necessarily apply to computer generated scenes. In principle, a mitigation is accomplished by reducing the mismatch between perceived motion from different sources (visual, vestibular). The motion in the visual is reflected by the optical flow of the frames. In computer-generated content, the optical flow is comparable to that in real-world footage, but with fewer disturbances, as the influencing factors remain under full control. Accordingly, an effective use of

GVS to mitigate CS in computer generated video-like content (little linear motion of the user) is very probable. In this sense, we would like to reinforce the importance of real-world videos in our experiment. Compared to fully controllable generated content, they incorporate some *real world complexity* (imperfect pixels, unpredictable events). Once this complexity can be controlled with the GVS, it is trivial to achieve an effect with generated content. But this is not necessarily applicable the other way around. For virtual interactive worlds, however, a generalization is more difficult, as these additionally involve linear motions of the user. Unexpected interactions may occur between the two motion sources (linear and rotational motion). Still, other works in the field of redirected walking with GVS already show that the use of GVS can be an effective and useful extension in interactive scenes [Langbehn et al. 2019a, Sra 2017].

A.6.5 Experimental Modalities

The duration of the video in the experiment was ten minutes in total, which is probably shorter than a typical VR session. We chose this rather short time because our 360° video is sickness inducing by design and therefore has a high probability to cause CS. Any longer duration was considered to become unethical. We chose strong scenarios to have a better control over the experimental variables, i.e., to increase the group of participants that are affected by CS. With the two scenarios, people are more likely to be affected by the video at some point during the exposure time, regardless of their susceptibility to CS.The capability of our videos to cause CS is further demonstrated by the number of participants leaving the experiment early (20).

For our video we used a gyroscope to capture the rotational movements of the video recordings. Most 360° cameras are equipped with a built-in gyroscope, mainly used for image stabilization. The prerequisite for this kind of metadata is therefore usually given for own recordings. It is more difficult when videos from the Internet are used, as there is usually no motion data available for these recordings. Analytical methods or artificial neural networks can be used to infer the background movements of arbitrary video. In particular, feature detection and quaternion matching or smoothed optical flow promise a reliable calculation of the rotations on existing material [Doyle et al. 2014, Horn 1987, Muja and Lowe 2009]. However, this is outside the scope of this work.

A.6.6 Ethical Considerations

According to current scientific knowledge GVS is considered safe given a moderate current (2.5mA or smaller) and healthy participants [Utz et al. 2010, Brunoni et al.

2011]. Most interesting, regarding adverse effects of current stimulation of the brain is the work of Brunoni et al. [2011]. They analysed over 200 studies for tDCS, a superset of GVS, with a total of 3836 participants. They found that "type of adverse events is mild and frequency of them in tDCS studies is low". No serious adverse event occurred. However, this primarily concerns short-term stimulation. Throughout our work, we could not observe any contrary effects. All effects were temporary (minor localized irritation of the skin, initial low disorientation symptoms) and resolved quickly after the stimulation. Nevertheless, GVS should be used with caution. All of our experiments involved only a short stimulation ($\leq$ 10 min) and were conducted in strict accordance with ethical guidelines under the supervision of the ethics committee and were previously approved by a physicist with proven expertise on the field. The effect of stimulation over longer periods (several hours) and frequent use may need further investigation.

A.7 Conclusion

In this work, we investigate 3D GVS to induce the sensation of rotation in arbitrary direction. Our galvanic stimulation is based on the OVR model and implemented to stimulate either only the strongest visual motion axis of roll, pitch and yaw, or precise intermediate rotations. Removing the visual-vestibular mismatch for rotational motions is essential, since they are the main factor responsible for inducing CS [Groth et al. 2021b]. This importance is confirmed by the results of our experiment, where over half of the initial CS was removed and the users' feeling of comfort significantly increased with the applied GVS methods.

Furthermore, we revealed that the application of GVS makes a difference to its effect. Whereas both methods, strongest axis stimulation and interpolated 3D GVS, successfully mitigated CS, the here introduced precise stimulation of intermediate axes has further positive effects. With this method, participants are willing to spend more time in the VR scene. At the same time they experienced the lowest level of discomfort with interpolated currents.

Comparing both genders, both GVS methods yield considerably better results for men while also significantly increasing the feeling of comfort for female participants. After the same initial values for the control session, GVS reduced the CS scores and the time spent in the session increased by a large margin. Nevertheless, both men and women took advantage of the stimulation.

Our results raise the confidence that 3D GVS offers a meaningful extension for VR applications that is able to achieve significant improvements of the virtual experience.

Nevertheless, this work does not necessarily prove that all rotations of the visual field are fully mapped onto the vestibular system by our interpolation as initially suggested by our small empirical pre-experiment. Since we now demonstrated the practicality of our approach, our future work will focus on the precision of this mapping and further improvements of the GVS application in virtual reality.

B Instant Hand Redirection in Virtual Reality Through Electrical Muscle Stimulation-Triggered Eye Blinks

In this paper we investigate the use of electrical muscle stimulation (EMS) to trigger eye blinks for instant hand redirection in virtual reality (VR). With the rapid development of VR technology and increasing user expectations for realistic experiences, maintaining a seamless match between real and virtual objects becomes crucial for immersive interactions. However, hand movements are fast and sometimes unpredictable, increasing the need for instantaneous redirection. We introduce EMS to the field of hand redirection in VR through precise stimulation of the eyelid muscles. By exploiting the phenomenon of change blindness through natural eye blinks, our novel stimulation model achieves instantaneous, imperceptible hand redirection without the need for eye tracking. We first empirically validate the efficiency of our EMS model in eliciting full eye closure. In a second experiment, we demonstrate the feasibility of using such a technique for seamless instantaneous displacement in VR and its particular impact for hand redirection. Among other factors, our analysis also delves into the under-explored domain of gender influence on hand redirection techniques, revealing significant gender-based performance disparities.

Colin Groth and Timon Scholz and Susana Castillo and Jan-Philipp Tauscher and Marcus Magnor
Institute for Computer Graphics, TU Braunschweig, Germany

Conference paper published in *ACM Symposium on Virtual Reality Software and Technology (ACM VRST) 2023,* Article 37, Pages 1–11. Presented at *ACM VRST 2023.*
DOI: 10.1145/3611659.3615717

Instant Hand Redirection in Virtual Reality Through Electrical Muscle Stimulation-Triggered Eye Blinks

Figure B.1: We explore the use of electrical muscle stimulation to elicit eye blinks in VR users for instantaneous hand redirection. In a first eye-tracking experiment, we verify the reliability of triggering full eye closure with our novel stimulation model (left). In a second comprehensive VR experiment, we exploit muscle stimulation in conjunction with hand tracking to achieve imperceptible redirection (middle and right). Thereby, our method does not require eye tracking.

B.1 Introduction

In recent years, VR technology has evolved at a remarkable speed, with modern devices offering new features such as eye tracking (ET), facial tracking and hand tracking. As the realism of VR presentations increases, so does the user's expectations for the experience. A mismatch between the visual and tactile perception of a scene can cause irritation and break the immersion of the user.

Redirection techniques can perform unnoticeable manipulations to the real-to-virtual mapping to maintain the user's sense of presence even when the physical environment does not match the virtual one. Two representative examples are hand redirection to adjust for an offset between real and virtual objects or redirected walking to allow exploration of larger virtual environments and avoidance of obstacles in the real room. hand redirection adjusts the position of the virtual hand to obtain correct haptic feedback even if the virtual and physical object positions or scene geometry do not coincide.

A practical example of the potential benefits of hand redirection comes from prototype production. Nowadays, when the interior of a new automobile is designed, usually virtual models are created and examined in VR. For compelling interaction and a true-to-life experience, a physical representation of the most important parts of the cabin is used as a haptic proxy. However, extensive physical models that are adjusted with each iteration of the virtual model are impractical. Here, a generic haptic proxy along with a hand redirection solution would save time and money while improving the overall user experience.

Hand redirection exploits the phenomenon of virtual dominance [Gibson 1933]. In a conflict between visual and proprioceptive sensations, the human brain tends to trust the visual information more. Conventional hand redirection applies an incremental offset

to the virtual hand, which is corrected by a natural compensatory movement in the opposite direction, thus reaching an offset target. Besides this continuous redirection, some works try to exploit the effect of change blindness for the purpose of redirection. Change blindness describes the phenomenon of people's inability to perceive changes in an object or scene [Simons and Levin 1998a]. Naturally, change blindness occurs as soon as the vision is temporarily interrupted, for example, because an eye blink is performed. Both, research on hand redirection [Zenner et al. 2021] and redirected walking [Langbehn et al. 2018, 2019b, Nguyen and Kunz 2018] try to detect and utilize natural blinks for subtle redirection. Compared to other redirection methods, the exploitation of change blindness allows for instant redirection. This unnoticeable, instantaneous displacement is particularly interesting for unexpected changes in the direction of movement or to enhance continuous methods for particularly strong redirection [Zenner et al. 2021]. Humans blink approximately 15-20 times per minute [Nakano et al. 2013]. Relying on users to have their eyes closed at the exact moment when redirection is needed seems impractical. Accordingly, it would be highly advantageous to subtly encourage the user to blink at precise moments to exploit the associated change blindness. The literature presents methods that systematically induce eye blink. The stimulation in these cases is performed by bright light flashes [Rushworth 1962], airpuff stimulation [Manning et al. 1983, VanderWerf et al. 2003], physical taps to the glabella [Rushworth 1962] or loud sounds [Säring and Von Cramon 1981]. Unfortunately, these techniques were found to be very intrusive and do not have a good success rate in most cases.

In this paper, we investigate the actuation of eye blinks by electrical stimulation of the eyelid muscle and evaluate the applicability of this method for instant redirection. We present the first electrical muscle stimulation (EMS)-based model for symmetric blink stimulation that was developed based on extensive preliminary studies and expert knowledge. Our model achieves eye closure by low electrical currents and causes the corresponding blinking response by intrinsic body signals. The efficiency of the blink stimulation is evaluated in a first psychophysical experiment. Secondly, in an experiment on hand redirection in VR, we demonstrate the capability of the blink stimulation for instant redirection. Although we chose a intentionally challenging scenario for the experiment, our EMS-based instant hand redirection achieves detection thresholds on par with those of prior publications. In comparison to other methods, eye tracking is not required for the implementation. Moreover, we investigate the role of gender on the efficiency of hand redirection techniques, which is unknown to date. We found that gender indeed is important for the efficiency of the redirection, with significantly better performance in females.

B.2 Related Work

B.2.1 Hand Redirection in Virtual Reality

Hand redirection involves deliberately manipulating the mapping between the real and virtual environment to gain control over the movement of the user's hand. In the scope of this research, hand redirection aims to enable the VR system to direct the user's real hand movement towards a different target than what the user perceives. This redirection can be compared to redirected walking where users are deceived into following a physical path that differs from their virtual path [Grechkin et al. 2016, Langbehn et al. 2018, Sun et al. 2018]. The redirection of hands is useful in various VR applications, most notably to enhance the scalability of passive haptic feedback [Hinckley et al. 1994].

Hartfill et al. [2021] explore VR avatars with fully articulated hands, fostering natural interactions in a VR environment. Their study investigates non-isomorphic techniques, particularly a hand retargeting approach for slower movements, relevant in therapeutic contexts. Psychophysical experiments reveal distinct detection thresholds of mid-air motion paths, with no significant difference between dominant and non-dominant hand. In the work of Kohli et al. [2012], redirected touching is introduced, a technique that combines distortions in the virtual scene with HR to convey the perception of differently shaped virtual objects using only a single haptic proxy. Azmandian et al. [2016] later proposed using HR for haptic retargeting, allowing users to interact with spatially dislocated virtual cubes mapped onto one single physical proxy. This proxy in the real environment allowed users to have proper haptic feedback during their interactions with the virtual objects. Building upon the former work, Cheng et al. [2017] presented research on hand redirection in VR for simulating touch feedback using sparse haptic proxies. They demonstrated the effectiveness of continuous hand redirection and discovered that touch intentions could be predicted based on users' eye gaze. Their work emphasized the importance of considering both physical and cognitive factors in designing immersive and realistic VR experiences with enhanced touch sensations. Furthermore, HR has been applied to enhance the perceived resolution and speed of shape displays [Abtahi and Follmer 2018], overcome limitations in encountered-type haptic systems [Abtahi et al. 2019, Gonzalez et al. 2020], extend the range of haptic effects [Zenner and Krüger 2017, 2020], and enable more ergonomic interactions with virtual user interfaces and scenes [Murillo et al. 2017]. Additionally, redirection techniques have been explored in the context of 3D interaction techniques and pseudo-haptic effects, simulating drag or weight sensations [Dominjon et al. 2005, Rietzler et al. 2018, Samad et al. 2019]. In most cases, hand warping techniques involve

either a constant offset [Benda et al. 2020, Han et al. 2018] or incremental relocation of the real-to-virtual mapping [Cheng et al. 2017, Azmandian et al. 2016, Burns et al. 2006, Esmaeili et al. 2020, Kohli et al. 2012, Spillmann et al. 2013, Zenner et al. 2021].

To address the time-critical nature of hand movements, this work investigates the use of stimulated blinks as a means of achieving instantaneous redirection in hand-related tasks. Previous research by Zenner et al. [2021] already utilizes the change blindness effect during spontaneous blinks to offset the virtual hand. In their study, blink-suppressed hand redirection (BSHR) was explored and found that combining BSHR with continuous redirection yielded better results than BSHR alone. The detection threshold for BSHR was around 8° (right, down) and 1.12 scale (towards). However, relying solely on spontaneous blinks occurring on average every 3 seconds may not be sufficient for fast hand movements [Nakano et al. 2013]. This work focuses on exploring the use of stimulated blinks to achieve time-critical redirection in hand-related tasks, building upon previous research by Zenner et al. [2021] and Cheng et al. [2017].

B.2.2 Electric Stimulation for Eye Blinks

The beginnings of modern blink research is associated with the work of Kugelberg, who first classified the blink reflex in 1952 [1952]. In this early days the reflex was triggered by physical taps. Electrical blink stimulation has been studied in recent years, mainly as a method of restoring the natural blink appearance in people with facial palsy and for facial pacing to counteract dry eye symptoms. McDonnall et al. [2009] were among the first to perform electrical blink stimulation. In patients with facial paralysis, they detected eyelid movements on the healthy side of the face to elicit a simultaneous blink response on the paretic side. In their approach, the electrodes were implanted into the eyelid of the patient. In a more recent study, Frigerio et al. [2015] used transcutaneous neural stimulation by surface electrodes to elicit eye blinks in individuals with acute facial paralysis. Their method had a 55% success rate, while the sensation of stimulation was rated as tolerable for daily use. In their experiments, an average current of 7.2 mA was required for full eye closure. Note, that this is well above our safety threshold of 2.5 mA. VanderWerf et al. [2003] investigated eyelid movements under different stimulus conditions, including electrical stimulation. They inserted a direct magnetic search coil into the eye and recorded its movements with EMG of the orbicularis oculi. They found that blinks induced by electrical stimulation have the shortest duration and are the most predictable. Lylykangas and colleagues have been major contributors to the field of constant-interval electrical blink stimulation [2018, 2020]. While their 2018 work, investigates the overall prevention of corneal damage due to the absence

of blinking [Lylykangas et al. 2018], in a more recent publication, the authors focus on the functionality and subjective experience of timer-triggered blinks [Lylykangas et al. 2020]. They showed that dry eye symptoms caused by chronic unilateral facial palsy could be significantly reduced by using predefined stimulation intervals. In their study of healthy participants, the stimulation was reported as not painful but mildly uncomfortable. Rantanen et al. [2016] investigated facial stimulation for both eye blinking and mouth movements in individuals with unilateral facial palsy. Whilst no experiments were conducted, the paper provided valuable insight into the appropriate size of the electrodes for healthy stimulation ($\geq 2mA/cm^2$). Several studies show that the human blink reflex is independent from the type of stimulation. Snow and Frith [1989] investigated the relationship between the blink reflex and eyelid movement. In their studies, both electrical and tapping stimulation were used with equal success. The relationship between the orbicularis oculi reflex and the type of stimulation has also been discussed by Cruccu and Deuschl [2000]. They argue that blinking is generally independent on the stimulus. Other papers delved more into the investigation of the movement of the eyelid during blinking. Evinger et al. [1991] characterized movements of the eyelid in a study that combined a search coil and EMG of the orbicularis oculi. Hammond et al. [1996] studied the early components of the blink, namely the cutaneous blink reflex, and how they relate to various properties of eyelid movement.

The former papers have demonstrated that electrical stimulation can be effectively used to restore a natural blink appearance and counteract dry eye symptoms in individuals with facial palsy. Since unilateral facial paralysis was considered the stimulation models developed so far are only suitable for unilateral blinking. Although the success rate of eliciting full eye closure varies between studies, it is clear that electrical stimulation can be an effective tool to trigger a blink response. Note that, in none of the works severe adverse effects were caused by electrical stimulation. However, it was reported to be slightly uncomfortable in some cases. Although we have focused so far on electrical stimulation, blink responses can be elicited in several other ways. In a recent work, Zenner et al. extensively investigated techniques for inducing the blink-reflex in immersive VR [2023]. Their work compared a verity of non-electric triggers, including visual, airpuff, mechanical, and auditory stimuli. The authors compared four promising techniques in a user study to reveal insights for efficacy and design recommendations. Notably, the VR-specific trigger through approaching virtual objects achieved best reliability. Thereby, the effectiveness of most methods contrasted with their potential disturbance, which should be taken into account in design decisions. In this paper, we propose a novel stimulation model that allows for the electrical stimulation of sym-

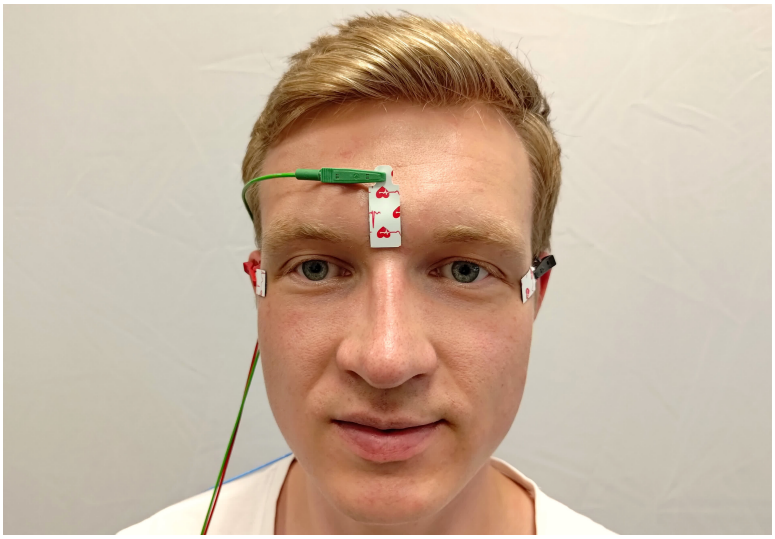

Figure B.2: Placement of the adhesive electrodes.

metrical bilateral eye blinks. We then use our stimulation model in an experiment to evaluate its efficient use for time-critical redirection tasks.

B.3 Stimulation Model

In this paper, we present a novel stimulation model for non-invasive and systematic actuation of eye blinks. Our model stimulates the orbicularis oculi muscles, which by contraction cause the eyelid movements. The stimulation requires only low electric currents that are safe for use in humans. The model was developed based on medical considerations and preliminary studies with multiple electrode placements tested. In order to effectively stimulate both eyes, three electrodes are attached to the wider area of the user's eye (see Fig. B.2). The first electrode is placed vertically on the glabella, the area just above the nose and between the eyebrows. The other two electrodes are applied in a horizontal position on either side of the eyes next to the lateral canthal region, 0.5 cm from the orbital rim towards the ears. During stimulation, the glabella electrode serves as the anode (current $= 2X$), while the other two electrodes make up the opposite pole, each with half the charge (current $= -X$ per electrode). The glabella electrode therefore operates as a current divider, distributing the stimulus equally to both eyelid muscles.

As the stimulation is applied, electric currents unfold over the orbicularis oculi, specifically the upper eyelid, causing irritation of the muscle, similar to a light physical

touch. As a result, a signal is sent to the brain requesting contraction of the eyelid muscle. The blink is thus triggered by a brain signal as a natural response of the human body. Alternatively, the zygomatic branch of the facial nerve, located on the lateral part of the orbital rim, can be directly stimulated to initiate eyelid movement [Frigerio et al. 2015, Lylykangas et al. 2020, 2018]. However, our pilot studies have shown that the effects of such stimulation are much smaller. In most cases, they are not sufficient to elicit a full eye closure with the required current safety threshold. Others proposed a stimulation of the supraorbital nerve with the cathode at the foramen, next to the eyebrows above the orbit, and the anode placed lateral on the forehead [VanderWerf et al. 2003]. However, our empirical findings suggest that with a safe threshold, this stimulation is much more likely to activate the eyebrow muscles than to stimulate the nerve sufficiently for eye closure.

The effect of stimulation seems to depend only on the alteration in direct current due to abrupt stimulation changes, rather than on constant stimulation. In preliminary experiments in which we gradually ramped up the current from zero to the maximum (and vice versa), blinking could not be elicited. However, this only applies to two-sided ramping with incremental steps at the beginning and end of the stimulation. The transient state that provokes eye contraction is initiated by two events: the current alteration from full current to zero and from zero to full stimulation. However, if the stimulation does not last longer than the combined time of the reaction delay and the blink duration (around 130 + 120 ms), the body treats the two current changes as one state and only one blink is elicited.

For stimulation, we apply countercurrents to Electrical Muscle Stimulation (EMS) to achieve the same stimulating effect with half the power. Countercurrents were proposed by Aoyama et al. [2016] for use in Galvanic Vestibular Stimulation (GVS), a special type of electrical stimulation of the vestibular system that activates the semicircular canals in the inner ear and produces a sense of motion. We have adopted the idea of countercurrent stimulation for eyelid EMS. Instead of stimulating from time 0 to N with the current X, we stimulate with half the charge but switch the polarity in between. Formally:

$$C(t) = \begin{cases} \frac{X}{2} & 0 \leq t \leq \frac{N}{2} \\ -\frac{X}{2} & \frac{N}{2} < t \leq N \\ 0 & \text{otherwise} \end{cases}$$

Hence, the amount of the electricity C used at any time t during the stimulation is halved, while the effect is sustained. In the experiments, the maximum current value X is determined per participant during the calibration phase.

The size of the electrodes has been chosen as a balance between the need for efficient stimulation and skin safety. Thereby we followed previous research on healthy stimulation to avoid skin irritation [Rantanen et al. 2016]. At the same time, we make sure that the electrode size is small enough to maximize efficiency. The electrodes next to the eyes have a size of 17 x 13 mm, while the electrode for the glabella position has a size of 25 x 13 mm. This larger size for the glabella electrode is appropriate as it handles twice the current of the opposite poles. We have found that the larger electrode achieves the same stimulation effect while minimizing the risk of skin irritation by covering a larger area of skin.

B.4 Experiments

All experiments are approved by the university's ethics committee under the identification number FV_2023-01.

We conducted two experiments to evaluate the practical use of EMS-based blink stimulation. The first experiment validates the effectiveness of the proposed stimulation model for eye blink success rate and properties of the elicited blinks. Based on these insights, the second experiment exposes users to a realistic hand redirection scenario in VR where hand redirection is performed instantaneously during the induced eye blinks as well as with continuous redirection.

B.4.1 Experiment 1: Blink stimulation and ET

In a first experiment we evaluated the efficiency of our proposed eye blink stimulation model. We also aimed to gain deeper knowledge about the nature of the eye blinks that are elicited by EMS. This experiment is performed without VR glasses, but instead participants looked at images on a monitor to utilize our stationary high-performance eye tracking (ET).

Stimuli The EMS of the eyelid muscles was performed as described in Sec. B.3. As a visual stimulus, we showed multiple images from an image retargeting dataset to the participants containing artifacts from the retargeting process [Castillo et al. 2011]. The images were not relevant for the evaluation of the stimulation method, but provided a visual landmark to take the participant's attention away from the electrical stimulation, as would be the case in a practical scenario with dynamic visual content. All images were displayed in random order for 18 seconds each. To provide a dummy task to participants, they were asked to explore the images and find artifacts.

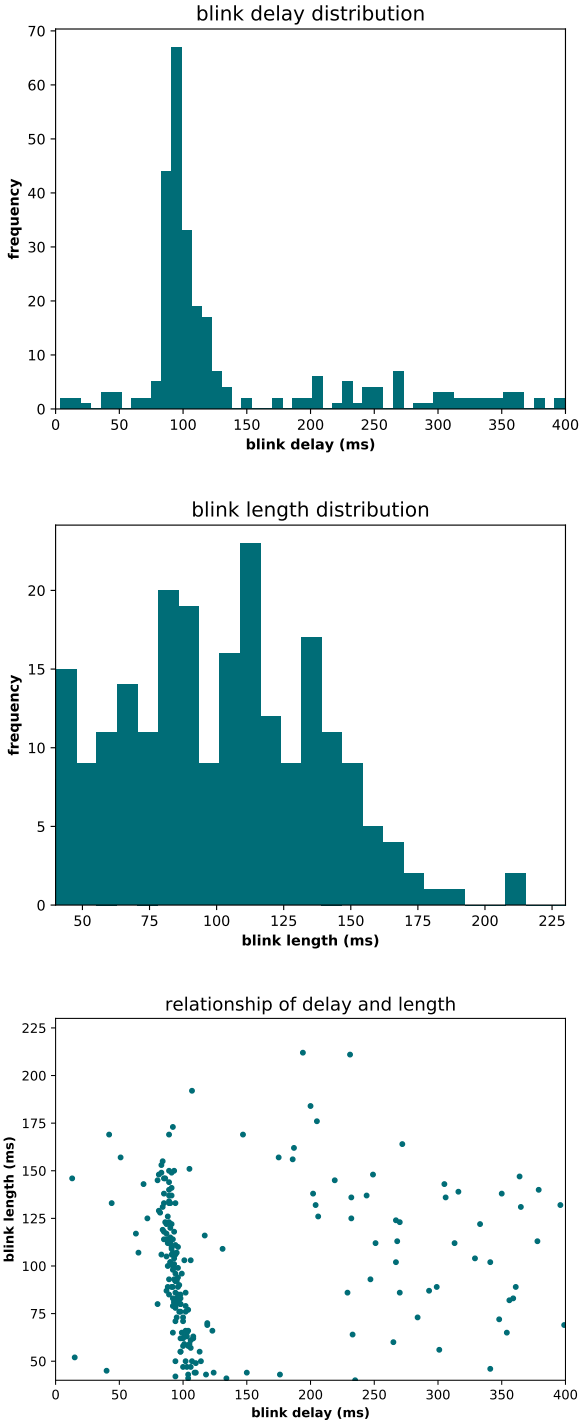

Figure B.3: Distribution of the delay between start of the stimulation and start of the blink (left), blink length (middle) and relation of both former parameters (right).

Apparatus For the EMS, we use a stimulation device from Good Vibrations Engineering [Vestibulator]. The device is restricted to a maximum current of 2.5 mA per electrode. The latency between sending a signal in the application and stimulation output to the electrodes is less than 15 ms. By using battery power and information transmission via air-gapped fiber optics, the device is protected against any unanticipated power transmission. For the eye tracking we used a desktop-mounted EyeLink 1000 eye tracker by SR Research that provides a 1000 Hz sample rate and reliable blink detection.

Participants A total of 15 participants took part in the experiment (6 females, Age range = 23-66, Avg age = 32.12, SD = 13.09). One participant terminated the experiment early by his own will. However, no participant experienced severe negative symptoms from the stimulation, e.g. severe discomfort, which would have resulted in an immediate termination of the experiment. The participation in this experiment was voluntarily and without compensation.

Procedure Prior to the experiment, all participants were provided with a consent form and information about the experimental procedure, as well as the functionality and safety of the EMS stimulation device. We told participants that the experiment aims to investigate eye tracking in combination with EMS near the orbicularis oculi, without disclosing the true objective of blink induction to avoid any biases. The placement of the three electrodes was carried out according to the methodology described in Sec. B.3, one at the glabella and two lateral to the eyes. Before attaching the electrodes, we meticulously cleaned the corresponding skin areas with an alcohol pad to enhance conductivity and electrode adhesion. Prior to the start of the experiment, the optimal stimulation current for each participant was determined. For the calibration, we gradually incremented the current until we observed induced full eye closure. This was our criterion for determining whether the blink induction was successful. Throughout the process, participants were asked to communicate any discomfort or general sensations they experience, allowing us to adjust the current intensity accordingly. The stimulation current chosen was the one that elicited the strongest blink reflex without causing unwanted negative feelings, which was double-checked by three additional stimulations to ensure consistent full eye closure. For the experimental setup, participants were positioned 95 cm in front of a screen and their heads were rested on a chin rest to maintain a consistent viewing distance (see Fig. B.1). To ensure accurate eye tracking, we executed a calibration routine that personalized the eye tracker to each individ-

ual participant, followed by a validation routine to ensure accuracy. The experiment consisted of 70 stimulation trials and 30 sham trials with no stimulation, which were randomly distributed to avoid anticipation bias. Each trial lasted 3000 ms while the stimulation had a fixed duration of 40 ms, determined by previous pilot experiments. Stimulation occurred once during a random time within the stimulation trial, with a safety gap of 500 ms to the start and end of each trial to account for the eye's relaxation time after blinking. Eye blinks induce rapid changes in pupil size characterized by a decrease and subsequent increase within 500 ms after eye closure [Nakano et al. 2013]. Participants did not wear glasses during the experiment to avoid interference with the eye tracking system. However, all participants reported clear visibility of the display. After the experiment, participants filled in demographic questions and rated the sensation experienced during stimulation.

Questionnaire We ask two successive questions to evaluate participants' feeling of the electrical stimulation. First, participants classified the stimulation with one of the following categories: unnoticeable, perceptible, discomfortable, pain. For all classifications except for unnoticeable stimulations a more detailed rating of the intensity of the feeling is asked for in the next step. To rate the sensation's intensity, we use a Borg CR10 scale. This scale rates the subjective sensation of participants between 0 and 12 and represents a common approach in medicine to provide perceptual ratings. Already in early research, the perceptual intensity was found to grow with the logarithm of the physical intensity [Fechner 1876]. The Borg CR10 scale is based on logarithmic growth functions, determined by internal psychophysical criteria, based on former results on ratio scaling methods [Borg 1998]. Multiple studies show the advantage of this measurement method over other methods like the visual analogue scale [Neely et al. 1992, Williamson and Hoggart 2005].

B.4.2 Results of Experiment 1

We found that for most people the blink induction worked well while no blinks were evoked in others. Overall, 60% of the participants showed a positive blink response. For the participants where blinks could be induced properly, the average blink rate is 72%. The other participants that did not respond on the electrical stimulation, their rate of blink responses in 8% of the time is on the same level of natural eye blink behavior [Nakano et al. 2013]. It is a surprising finding that for EMS-based blink stimulation two response groups seem to exist with either a consistent eye closure response or full resistance against any electrical blink stimulation. However, in our

experiment the classification of the participants into one of the two response groups was found to correlate with the blink response during the calibration. When full eye closure was achieved in the calibration, the success rate for eye blinks in the experiment was also high, while the opposite is true for participants where no full eye closure could be achieved during the calibration.

Figure B.3 shows the blink delay and length as well as the relation between the two parameters. For the analysis, eye blinks within 400 ms after the stimulation are considered. Later eye blinks are assumed to be spontaneous eye blinks that are not related to the electrical stimulation. The results indicate that the eye closure duration for EMS-induced eye blinks is varying significantly for most blinks lasting for 57 to 126 ms (25% and 75% percentile). The delay between start of stimulation and beginning of the eye blink, on the other hand, was found to be stable for all blinks of one individual and does also not vary considerably between participants. On average, it takes 136 ms for the blink to start after the stimulation was triggered. This duration already includes the 40 ms stimulation time. The majority of nine participants rated the feeling of the stimulation as noticeable, but did not feel uncomfortable by it. Six participants perceived the stimulation with mild to medium discomfort. One participant reported to not feel the stimulation at all in the experiment. The results showed no correlation between the feeling of the stimulation and the actual success of evoking eye closure. In the sham trials without blink stimulation, participants had their eyes closed for 10% of the time, which corresponds to normal human blink behavior [Nakano et al. 2013].

B.4.3 Experiment 2: Hand Redirection in VR

The second experiment explores the applicability of EMS-based blink stimulation for instant hand redirection in VR, based on the insights of the first experiment. Besides instant hand redirection, we apply continuous redirection as a second condition which adjusts the offset of the virtual hand over the entire reaching distance. Previous research found continuous redirection to be an effective and subtle method [Cheng et al. 2017, Zenner et al. 2021] that is used as a baseline in our experiments. The main objective is not to replace common redirection techniques but rather to provide an accompanying solution to enhance redirections with instantaneous displacements in complex situations. The implementation of the continuous redirection follows the design of Cheng et al. [2017].

In the experiment, we use the method of constant stimuli to determine the noticeability of the redirection for a certain parameter set [Kingdom and Prins 2016, Klein 2001, Grechkin et al. 2016]. For each of the two conditions (blink, continuous), we

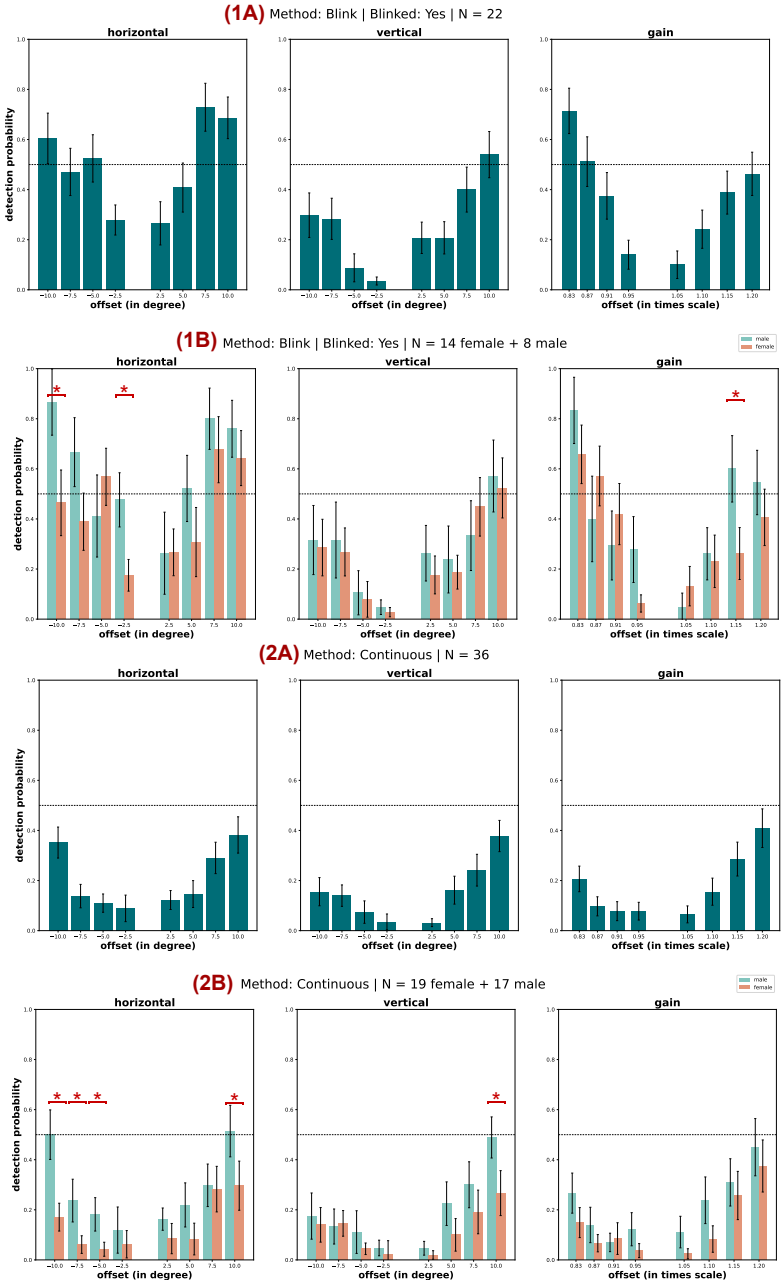

Figure B.4: Detection probabilities for the offsets of the physical hand and all three axes. The results are shown for the blink condition (1) and continuous condition (2). Aside from the overall results (A) we present a direct comparison of men and women (B). Error bars represent the standard error of the mean. Statistically significant results are denoted by '*' ($p_{\text{corrected}} < 0.05$).

considered a redirection along the three central axes (horizontal, vertical, gain). The redirection levels are chosen with four different offset magnitudes respectively, arranged around the detection thresholds of previous works (see Fig. B.4 for explicit level values) [Cheng et al. 2017, Zenner et al. 2021]. In addition, for each of these combinations, five repetitions were conducted, resulting in a total of 125 trials per participant (2 methods * 3 axes * 4 offset magnitudes * 5 repetitions + 5 control trials with no redirection). The detection probability for every redirection level is derived by the proportion of trials where the manipulation was noticed, that is the participants' amount of "No" answers over the total amount. We decided to distribute the trials equally across offset levels and presented them in random order. The frequently used 1up/1down method reported in the literature [Kingdom and Prins 2016, Zenner et al. 2021] could have introduced an expectation bias and was therefore not chosen for the experiment. We counteract the correspondingly larger number of trials of the method of constant stimuli with a streamlined experimental design.

Virtual Environment In the experiment, we presented a photorealistic virtual environment of a furnished room in a HMD (see Fig. B.1). The virtual room contains a table and a chair that are calibrated in every experiment to match the position of their physical counterparts. The hand models are rendered with a uniform diffuse material for clear visibility. We included animations for all boxes the user had to touch to provide proper feedback. The response control for the Yes/No question is designed so that the participant loses reference to the position of their physical hand, in order not to reveal the previous manipulation in the transition between trials. We achieve this loss of reference by hiding the real hand and transferring the control to a cursor with scaled movements that only moves on the axis between the answers. In contrast to our design, previous works required participants to move their hands entirely behind their backs in each trial to reset the manipulation [Zenner et al. 2021]. The described design choices allow for a fluid and fast experimental procedure, which prevents participants from becoming unmotivated or unfocused. Please refer to our supplementary video for further details. We implemented the application in *Unreal Engine 5.1.*

The EMS for eye blinks follows our method described in Sec. B.3. Experiment 1 showed the duration between EMS and the muscle response for an onset of the eye blink to be mostly constant. Based on these findings, we chose a fixed delay of 150 ms between start of the stimulation and manipulation of the position of the virtual hand model. This delay already accounts for variances in blink time and the latency of the rendering and visual display.

Apparatus We used an Oculus Quest Pro HMD with a 106° FOV and hand tracking capabilities. The resolution of that HMD is 1800 × 1920 px per eye at 90 Hz refresh rate. The processing was performed on a local workstation (RTX 3090) over a Link-Cable. For the stimulation, the same device was used in both experiments [Vestibulator].

Participants We recruited 40 participants, all of whom completed the experiment (20 females, Age range = 19-37, Avg age = 24.93, SD = 4.54). Post-hoc, the data of one person had to be excluded because he did not properly follow the experimental task. From the stimulation, most participants experienced no negative feelings. In rare cases, moderate discomfort was perceived. The majority of 36 participants were right-handed (3 left-handed, 1 two-sided). The participation was compensated with 15€. In total, 5000 trials were conducted.

Procedure Prior to the experiment, participants were provided with a detailed explanation of the functionality of the EMS device and the task they were about to perform. They were given a consent form to fill out if they wished to participate. During the setup phase, three electrodes are placed on the participants' eye region (cf. Sec. B.3), and the stimulation current was calibrated individually based on the procedure outlined in Experiment 1. Once the calibration was completed, the experiment proceeded with the calibration of the room location. Participants were instructed to place their right hand on a fixed centered position on the physical table to adjust the virtual counterpart accordingly. We also determined the maximum reach distance of the participants by asking them to extend their arm. The arm length served as a reference for the distant placement of the target boxes to account for physiological differences. In the experiment, we only let participants use the right hand to maximize the control parameters of the experiment. The initial five trials served as control trials, during which no redirection was applied. The objective of these trials was to gauge participants' understanding of the task and ensure their reliability. Note, that participants were not informed about the true objective of these validation trials.

At the beginning of each trial, participants were instructed to place their hand on a fixed virtually marked area on the table in front of them, providing a proper haptic presentation at the start. After holding the start position for a brief period of time, a box appeared at a random position but with a fixed distance from the start, which corresponded to the participant's calibrated reach distance. The random position of the target ranged from 10° to 40° vertically and -10° to 80° horizontally from the forward vector. While participants moved their hand towards the target, a certain redirection

was applied. The applied hand redirection is relative to the gaze vector towards the target. The offsets and methods for redirection were randomized across trials. However, the eye blink stimulation was applied in all trials regardless of the redirection method to avoid bias of the participants. The blink was triggered at a randomized position between 20 and 50 percent of the reach. This safety gap ensured the blink redirection to not occur when the hand is within the visual focal area. This assumption is founded in the natural behavior of humans to fixate their reach target with the eyes and use physiotactile perception rather than to follow the path of the hand with their gaze. As soon as the manipulation enters the user's focus area during the redirection, the remapping becomes easily noticeable. In the blink trials, the hand is instantaneously moved to the offset position during the estimated eye closure. Once the participant reached the target the background was blacked out and the participant's hand disappeared. Instead, participants controlled a cursor to answer a Yes/No question *"Did the movement trajectory of the physical and virtual hand coincide?"*. After their response, the virtual room environment and hand model reappeared and the next trial commenced. Following the completion of the experiment, participants filled in demographic questions and rated their perception of the electrical stimulation, as in Experiment 1 (see Sec. B.4.1).

B.4.4 Results of Experiment 2

For the analysis of the instant redirection during eye blinks we only considered those participants who showed a positive blink response to the stimulation during the calibration phase. These considerations are based on the results of the first experiment that has demonstrated the success of the blink stimulation to be anticipated by the level of eye closure during calibration. In our VR experiment, we achieved full eye closure by stimulation in 22 of the 36 considered participants (61.1%). The results of the continuous redirection include all participants as this method is independent of the blinking behavior. However, the subset of blink-susceptible participants showed comparable results for this condition.

The results do not include the participants who showed a poor understanding of the task or who already misperceived the trials with zero offset. For each participant, five control trials were performed without any redirection taking place. All data from participants who indicated in more than 20% of these control trials to see a manipulation that was not actually present was excluded from the analysis. In our experiment, this exclusion affects 3 out of 40 participants.

Fig. B.4 illustrates the results for instant redirection during EMS-based blink stimulation (1A). The axes represent the average likelihood that a redirection was detected,

Table B.1: Detection thresholds for the proposed blink-suppressed instant hand redirection technique.

	right / left	**up / down**	**forward / backward**
overall	5.81°/ -7.07°	9.43°/ -12.73°	1.2x / 0.88x
females	6.67°/ -10.04°	9.28°/ -12.91°	1.25x / 0.88x
males	4.89°/ -4.42°	9.61°/ -12.45°	1.16x / 0.88x

separated into the three base axes, the shift in positive and negative direction, and the discrete offset levels. In the following, we assume a 50% detection probability to define the detection threshold of the redirection technique similar to previous work [Zenner et al. 2021, Cheng et al. 2017]. To derive the detection threshold of the psychophysical experiments, we fit a psychometric function to the probability data. More precisely, a logistic function was applied which provides good properties for psychophysical data [Kingdom and Prins 2016]. We base the non-linear curve-fitting on linearly interpolated thresholds as initial guesses. Tab. B.1 shows the derived detection thresholds for all axes.

Furthermore, Fig. B.4 shows the results for the same scene but with continuous redirection applied (2A). The overall trend of the individual axes follows the results of the blink condition, with manipulations in vertical direction being the most subtle and manipulations to left/right coming more to users' attention. However, for the chosen scenario the detection probabilities of the continuous redirection condition are lower for all three axes. Even for the maximum level, which extensively exceeded the reported threshold from previous work, continuous redirection still achieved a detection rate of 31.4% on average. This surprising finding may underline that the fast scenario of our experiment was actually a good fit for this method. The time participants required to find and reach for the target from the start position does not deviate between conditions. The average time to perform the reach is 3.33 s.

VR experiences are subject to a general gender bias as noted in previous research [Grassini and Laumann 2020, Groth et al. 2022, Narciso et al. 2019, Groth et al. 2021b,a]. However, it is still unclear how gender influences the success of hand redirection in VR. In Tab. B.1 the separate detection thresholds of our results are given for men and women, respectively. To analyze differences in the results between genders, we perform pairwise two-sided independent t-tests. Prior to the pairwise tests, homogeneity of variances is tested by performing Levene's test. In case no equal population variance can be assumed, Welch's t-test is performed. To account for multiple comparisons,

we adjust the resulting p-values using the procedure of Benjamini and Hochberg to decrease the false discovery rate. Fig. B.4 (1B, 2B) demonstrates the results of the gender-based analysis. In line with former findings, we found major differences in the efficiency of hand redirection depending on the user's gender. In our experiment, given the mostly higher ratings for males, our data suggests that women are less likely to notice the redirection independent of the condition and axis. This gender difference is most pronounced for horizontal hand redirection. At the same time, female participants were slightly faster in performing the reach task with 3.17 s (SD = 1.97) on average compared with 3.52 s (SD = 1.74) for males. These times cover the full timespan between appearance of the target and the first hand contact with it, and therefore also include the timings participants required to locate the target.

Regardless of the efficiency of instant redirection achieved during eye blinks, practical applicability will not be obtained if the majority associates negative feelings along with the EMS-based stimulation. To evaluate participants' perception of the electrical stimulation, we perform a two step evaluation process immediately after the experiment. First, participants were asked to categorize their perception of the stimulation as either *unnoticed*, *perceptible*, *discomfortable* or *pain*. For every category (except for unnoticed stimulations) they were asked to rate the feeling of the perception on a Borg scale between 0 and 12 [Borg 1998]. In our experiment, the majority of participants noticed the stimulation but did not perceive it as uncomfortable (see Fig. B.5). A minority of 13% of the participants indicated to feel discomfort by the stimulation with mostly weak or moderate intensity. On the other hand, six participants did not notice the stimulation at all, including participants with good blink responses. There were no reports of more discomforting experiences. Overall, these results give a positive indication for the practical use of EMS-based stimulation in a broader audience.

B.5 Discussion and Limitations

Experimental Results Movements that are slower by nature, for example because they are performed with the whole body as in redirected walking, are known to perform much better with redirection methods, since less attention is paid to the manipulations and several displacements can be performed consecutively [Zenner et al. 2021]. In our experiment we chose a challenging scenario with fast and short distance hand movements to establish a reliable baseline. Despite the challenging conditions, our EMS-based blink redirection demonstrated convincing results, with detection thresholds on par with previous work [Zenner et al. 2021, Cheng et al. 2017].

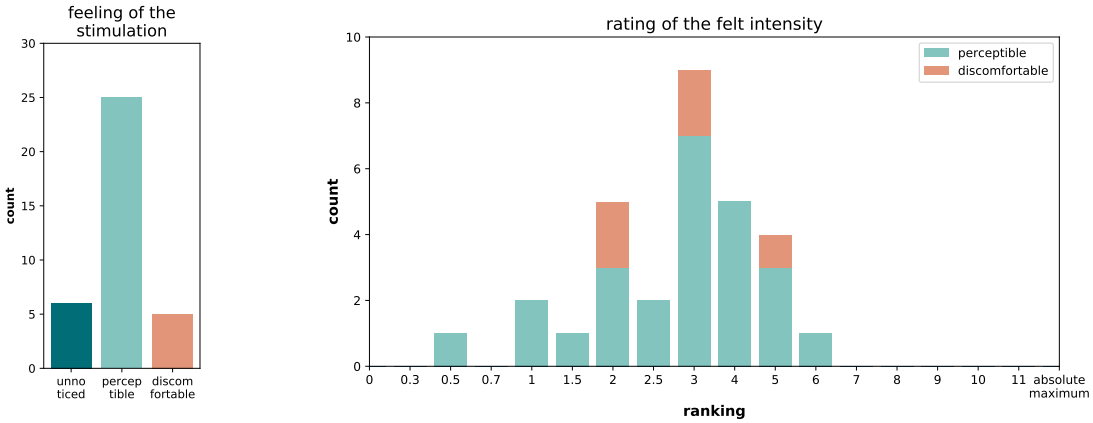

Figure B.5: Rating of the feeling of the stimulation. Participants first categorized their perceived feeling (left), to then rate its intensity (right). Unnoticed sensations were not rated.

With instantaneous redirection, a decisive advantage can be achieved in situations with complex conditions by repositioning the geometry of the hand unnoticed and without temporal delay. As already mentioned at the beginning, instant hand redirection is intended more as an extension of the well-functioning conventional redirection techniques. In continuous redirection, the intended target position has to be anticipated at the beginning of the hand movement in order to perform a suitable redirection. A version enhanced with instant hand redirection would be able to dynamically compensate for incorrectly anticipated directions by shifting up to 12.9° and scaling with a factor of 1.25x without the user even noticing the miscalculation. Previous work has already shown that an extension of continuous redirection with immediate displacements yields a significant improvement of the redirection threshold [Zenner et al. 2021]. In their work, the authors used spontaneous eye blinks and eye tracking to exploit the effect of change blindness. For our work eye tracking is not necessary since the blink is explicitly induced at the exact moment it is needed for the unnoticed manipulation of the scene.

Recently, Zenner et al. [2023] studied a variety of other techniques to trigger eye blinks. Comparing both works, electric stimulation seems to offer higher response rates than most visual or tactile techniques. However, they showed that VR-specific visual techniques, like the simulation of approaching objects, also achieve high blink response rates. Compared with visually induced blink triggers, the electrical stimulation of change blindness comes with the advantage that the visual field is not occluded and users could be less distracted. However, more future work is needed to truly compare

all relevant aspects of different eye blink stimulation models in VR scenarios.

Performance of Continuous Redirection Besides instant hand redirection through exploitation of eye blinks, we also applied continuous hand redirection in our VR experiment. The implementation followed the work of Cheng et al. [2017]. Unexpectedly, however, a significantly lower detection rate was obtained in our experiment than in previous work. Even the highest chosen offset value was not sufficient to determine a threshold at 50% detection likelihood. This outcome can have several reasons, which may be related to the method or virtual environment. Previous work usually performed a 1up/1down method to determine the detection threshold, which might expose issues discussed in Sec. B.4.3. In contrast, we chose a method of constant stimuli. The higher efficiency of the continuous redirection may also be caused by higher immersion. In our experiment, we utilized the Meta Quest Pro VR glasses, which by design leaves a narrow reference to the real world at the lower edge of the glasses. Through this reference, which also reveals parts of the upper arm, the virtual hand may have been adopted more strongly. Furthermore, a more precise representation of the virtual hand, e.g. via better hand tracking, is known to reduce the noticeability of hand redirection [Ogawa et al. 2020].

Effect of Gender VR was shown to be gender biased [Grassini and Laumann 2020, Munafo et al. 2017] while its effect on hand redirection was not yet studied. Based on our results, we found significant disparity in the effectiveness of hand redirection between males and females. While hand redirection techniques are certainly able to unnoticeably relocate the user's hand regardless of gender, the detection probabilities of women are significantly lower. This divergence appears more pronounced, with larger redirection offsets. Although both redirection methods showed different detectability in the experiment, the relative difference in results between men and women was comparable. Besides the stronger impact of the method on women, however, a difference in the feeling of stimulation is also discernible. While the classification of the stimulation as noticeable but not unpleasant were evenly distributed between the genders, categorizations as uncomfortable were solely made by women. Thereby, the average selected current intensity is comparable. An equal number of men and women classified the stimulation as imperceptible. In summary, especially for women, the VR experience can be greatly improved with redirection strategies, even when strong redirection is required. Future studies on redirection should account for gender effects, since a difference in the efficiency of the techniques is evident.

Positive vs. Negative Offsets Most previous redirection studies assume redirection offsets of the hand to be equal in positive and negative direction of an axis and, thus, only consider displacements in right, up and forward direction. However, some former works already gave an indication for disparities in detection threshold by the shift direction along an axis [Benda et al. 2020, Esmaeili et al. 2020, Hartfill et al. 2021]. Our results demonstrate a redirection along an axis to have a significantly different detectability depending on the direction of the shift. These differences in notability were most profound on the vertical axis and non-significant for the horizontal axis, which is true for both conditions. Please refer to Tab. B.1 and Fig. B.4 for detailed results. All our findings of varying effectiveness of the redirection methods based on the axis offset direction were independent of gender.

General Observations and Safety for EMS According to current state of the science, EMS of the eye muscle is considered safe given moderate currents ($\leq$ 3 mA), and is used in a variety of research works and medical applications [McDonnall et al. 2009, Frigerio et al. 2015, VanderWerf et al. 2003, Lylykangas et al. 2018, 2020, Rantanen et al. 2016, Snow and Frith 1989]. The eye closure serve as a protective mechanism of the body against the EMS, similar to a fly hitting the eyelid. While our experiments involved multiple stimulations on the same individual with consistent current and setup, stimulations were perceived with varying intensity, and the resulting blinks differed in strength. Therefore, a correlation between the perceived intensity of stimulation and the intensity of the blink is likely to be apparent. We also found the conductivity of participants' skin to be important. During initial experiments, conductivity decreased when the skin was not cleaned with an alcohol pad, necessitating higher currents to achieve the same blink effect. This higher resistance even pushed the bar of needed currents above the maximum 2.5 mA in some cases. Furthermore, we noticed that low currents ($<$ 0.2 mA) induced rapid light flashes in front of the participants' eyes. Most likely these flashes are the result of excitation of the retina through electrical flow. There appears to be an optimal stimulation current around 0.6 mA at which the stimulation works most effectively. Current values that surpass this threshold, no longer trigger eye blinks and become less noticeable. While no adverse effects were observed in this work, intense EMS should always be used with caution. All of our experiments involved $<$ 150 stimulations over a total of 10 min per participant. Also, this research was conducted in accordance with experienced ophthalmologists and the corresponding ethics committee. The effect of stimulation over longer periods (several hours) and frequent use may need further investigation.

B.6 Conclusion

In a comprehensive VR experiment we highlighted the potential of EMS-based blink stimulation to enhance hand redirection (hand redirection) and, thereby, increase users' VR experience. In this work, we proposed a novel stimulation model that can effectively and safely stimulate eye blinks for instantaneous displacements of virtual hand models that go unnoticed for VR users. The model requires only three electrodes and applies counter-currents for highest efficiency. Beside the main objective, our study revealed intriguing insights to a variety of unexplored factors in the domain of hand redirection including new methods, key considerations of EMS, and gender effects. In future designs, blink-controlled instant hand redirection can extend conventional redirection to provide a decisive advantage in complex conditions by repositioning the hand geometry unnoticed and without temporal delay.

C Wavelet-Based Fast Decoding of 360° Videos

In this paper, we propose a wavelet-based video codec specifically designed for VR displays that enables real-time playback of high-resolution 360° videos. Our codec exploits the fact that only a fraction of the full 360° video frame is visible on the display at any time. To load and decode the video viewport-dependently in real time, we make use of the wavelet transform for intra- as well as inter-frame coding. Thereby, the relevant content is directly streamed from the drive, without the need to hold the entire frames in memory. With an average of 193 frames per second at 8192 × 8192 -pixel full-frame resolution, the conducted evaluation demonstrates that our codec's decoding performance is up to 272% higher than that of the state-of-the-art video codecs H.265 and AV1 for typical VR displays. By means of a perceptual study, we further illustrate the necessity of high frame rates for a better VR experience. Finally, we demonstrate how our wavelet-based codec can also directly be used in conjunction with foveation for further performance increase.

Colin Groth and Sascha Fricke and Susana Castillo and Marcus Magnor
Institute for Computer Graphics, TU Braunschweig, Germany

Journal article published in *IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG) 2023,* Volume 29, Article 5, Pages 2508–2516. Presented at *IEEE VR 2023.*
DOI: 10.1109/TVCG.2023.3247080

Figure C.1: In this paper, we propose a wavelet-based video codec that is able to load and decode the data of 360° videos viewport-dependently by exploiting the properties of the wavelet transform. While the quality is on par with state-of-the-art video codecs, we can achieve significantly faster playback times. Here, a visual comparison of our codec is given with HEVC, AV1 and the uncompressed reference. The samples are out of the computer-generated *City* video (left), all with 8k full-frame resolution.

C.1 Introduction

Codecs provide efficient compression allowing to store hundreds of videos on a single drive. This efficiency results from a precise adaptation to their specific application. Video codecs like HEVC/H.265 or AV1 use the discrete cosine transform (DCT) and motion compensation for high compression rates at a reasonable perceptual quality. The high compression, however, requires complex coding procedures. The specific hardware decoders of modern graphics cards compensate for some of this decoding load.

360° videos are a sophisticated form of viewing experience which have become known with the spread of virtual reality (VR) technology. In 360° videos the user can change their view anywhere, since the information of the entire space is available (three degree of freedom (DoF)). This free exploration is not possible with traditional videos where the field of view (FOV) is limited. Accordingly, 360° videos are best suited for an immersive user experience for video playback in VR. For comparable quality, the resolution of 360° videos must significantly exceed those of videos with a discrete view, since the representation on the display device only corresponds to a fraction of the whole video frame. For a modern VR headset with 2000x2000-pixel resolution per eye and 90° FOV, a comparably resolved 360° video would require 8000x8000 pixels (stereo) with up to 120Hz temporal resolution. However, only the part of the frame that lies within the device's viewport at the time of decoding is relevant for rendering. Current DCT-based codecs do not allow to load or decode only a defined part of a frame due to their complex non-linear structure. Accordingly, with 360° videos the common practice is to load,

upload, and decode the entire 360° frame, even though only a small part of the frame is considered for the rendering. Furthermore, the recording quality of 360° cameras is limited. Consequently, the resolution and frame rate of 360° videos nowadays do not come close to the quality of the renderings of virtual environments. Although there are some existing ideas that already try to improve the compression – e.g. tiling the frame into separate regions – these approaches often go against the basic compression concept of DCT and are only a compromise at the expense of compression efficiency.

An alternative to the DCT for compression is the wavelet transform, which offers two decisive advantages over the former: (1) different parts of the image can be loaded and decoded individually, e.g. the viewport of an head-mounted display (HMD); (2) the encoding is performed in frequency layers, which are each halved in frequency. Decoding an area in fewer steps is equivalent to displaying the image area at a lower resolution. Early attempts to use wavelet transform for image compression were limited to traditional presentations with a discrete FOV and have not gained wide use.

In this paper, we propose a wavelet-based codec for the compression of 360° videos. We particularly aim for high display speeds of high-resolution videos. Our implementation of the wavelet-based codec uses the wavelet transform for inter- and intra-frame compression. To the best of our knowledge, this is the first codec for 360° videos based on wavelet transforms. In comparison with modern codecs (HEVC/H.265 and AV1) and related work, we show that our wavelet-based approach offers a significant speed advantage while providing a comparable video quality and reasonable compression rate. In addition, we introduce foveated decoding. With foveated decoding the properties of the wavelets are used to gradually decrease the resolution with the distance from the focal point. Such foveation is, so far, only known from virtual scenes and offers further opportunities to improve both decoding speed and image perception. The Code for our codec is available at https://github.com/ColinGroth/Wavelet_Codec_360.

The contributions of the paper are summarized as follows:

- a novel approach to encode and decode videos for fast viewport-dependent playback based on wavelet transforms
- wavelet-based inter-frame transform without keyframes
- a technique to implicitly apply foveated rendering for wavelet-encoded videos during run-time
- objective and perceptual evaluation and comparisons with state-of-the-art video codecs and related work

C.2 Related Work

Loading the entire frame for 360° video playback in VR is inefficient since only a fraction of the frame is actually rendered. However, this procedure is the most common practice, as it reduces the need to adjust the standard video pipeline. In the following, we first discuss previous work aiming at a more resource-efficient presentation by adapting existing codecs and, in the second part, we introduce former attempts for wavelet-based codecs.

C.2.1 Viewport-Adaptive Display Techniques for Videos

Zare et al. [2016] proposed to use a tiling scheme to increase the decoding speed of streamed 360° HEVC videos. Their experimental setup consisted of a pipeline with a dedicated server and client side. On the server side, the same video was encoded in high and low resolution. With the motion constrained tile sets (MCTS) extension of HEVC the tiling was enabled for both versions of the video. The client, on the other hand, requested the required tile sets from the server based on the viewport. The authors tested different tiling schemes. The scheme with the most tiles (18 tiles) showed the highest bitrate savings (-40% based on Bjontegaard Delta Bitrate BD-BR [Bjontegaard 2001]). However, the compression losses increase proportionally with the number of used tiles, since all tiles are saved independently.

The tiling approach was later officially formulated by the Moving Picture Experts Group (MPEG) into the omnidirectional media format (OMAF) standard for storage and distribution of 360° videos [Choi et al. 2018]. The idea of OMAF is comparable to the work of Zare et al. [2016] and is applied to HEVC or AVC video codecs. The viewport-dependent streaming also uses MCTS to split the frames into tiles, each encoded in different qualities [Hannuksela et al. 2019].

Sreedhar et al. [2016] also recognized the technical challenges of bandwidth associated with high-resolution 360° videos. The main focus of their work was the mapping techniques in which the recorded spherical scenes are packed in a rectangular frame. The most used mapping techniques are equirectangle and cubemap projection, which were also found as the most effective in their scenario. For the comparison, the authors presented a methodology of the rate-distortion performance of the schemes.

In the work of Corbillon et al. [2017] the 360° videos are separated in individual tiles and offered in different resolutions. Unlike former works, the single tiles are created in different versions with only a selected part of every tile in a better visual quality. While the 360° video is streamed, the client communicates its viewpoint to the server which

selects the tiles so that the viewpoint is in the higher quality region. The paper does not specify actual display speeds, but it should be clear that the technology can save bandwidth.

The performance of 360°videos can be improved not only on the software side. Recent works investigate how computation reduction and energy efficiency can be achieved through hardware design. Sun et al. [2020] designed special hardware to accelerate the perspective projection on a FPGA. The results show significant energy reduction without a loss of performance. Zhao et al. [2020] leverages redundancies across frames and tiles of left and right eye projection. Results indicate a computation reduction of 34% and energy saving of 17% for implementations on a GPU or FPGA. Leng et al. [2019] propose energy-efficient VR video processing by optimising the projective transformation with semantic-aware streaming on the server-side and hardware-accelerated rendering on the client. The authors demonstrate that up to 42% energy of the VR device can be saved for 360° video content using their system design.

C.2.2 Wavelet based codecs

Probably the best known use of wavelets for imagery is the JPEG2000 image compression standard [Taubman and Marcellin 2012, Marcellin et al. 2000]. At the turn of the millennium, it initiated a new form of image compression and was meant to replace DCT-based image compression formats. JPEG2000 supports lossless and lossy compression. The wavelet transform operates with the biorthogonal wavelets, either the Cohen–Daubechies–Feauveau (CDF) 9/7 wavelet [Cohen et al. 1992] for lossy compression or the LeGall-Tabatabai (LGT) 5/3 wavelet [Le Gall and Tabatabai 1988] for lossless compression. The standard has four levels of decomposition as a default since there is not significant improvement in using higher decomposition levels when compressing images [ISO 2019]. One general advantage of wavelet compressed imagery is the progressive decoding, so that the quality of the visualisation improves progressively when more information is received. This progressive decoding is also supported in JPEG2000.

The JPEG2000 image standard was later extended to include video files. The extension is known as *Motion JPEG2000* and is based on the MP4 format. This video standard uses the JPEG2000 coding for the compression of the individual frames. An inter-frame compression does not take place. Thus, Motion JPEG2000 is more of a container format for the joint wrapping of JPEG2000 compressed frames. Note that our video codec differs clearly to the Motion JPEG2000 standard. We apply a wavelet-based inter-frame transform to successive frames and use a special data structure that allows

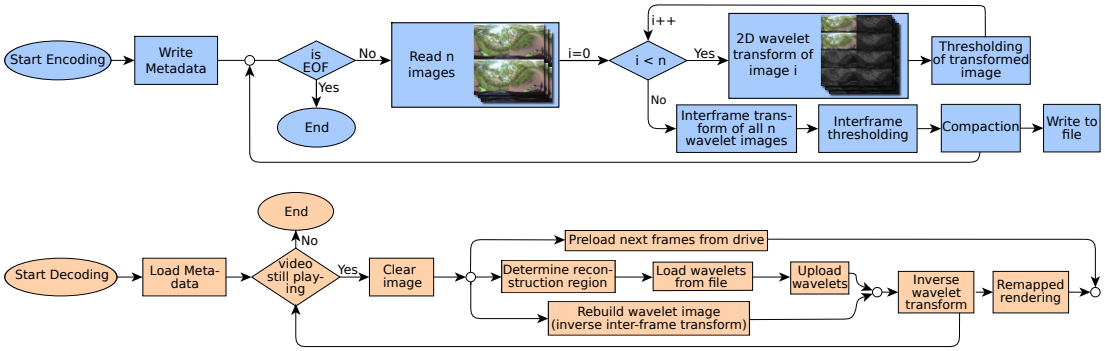

Figure C.2: Our program flow for encoding (blue) and decoding (orange) a video with our wavelet-based codec.

for viewport-adaptive data streaming. However, we took inspiration from JPEG2000 to design the frame-wise transform of our pipeline, e.g. we use the CDF 9/7 wavelet to obtain frequency information of single frames.

Efforts to create video codecs based on wavelet compression are rare and nowadays exclusively experimental. The most extensive attempt to create a wavelet-based video format to date was undertaken by BBC Research in 2008 [BBC Research 2008]. The resulting versions of the codec were named *Dirac* and *Schrödinger* in honour of the Nobel Prize-winning physicists. Dirac supports lossy and lossless coding for which it uses the same wavelets as JPEG2000 (CDF 9/7 wavelet or LGT 5/3 wavelet). The motion compensation is performed with the overlapped-block motion compensation (OBMC) logic for an effective inter-frame prediction [Orchard and Sullivan 1994]. Unfortunately, this overlap also prevents effective intra-prediction, since there are no unique separations for overlapping blocks, as is the case with common DCT-based codecs. The overlapping-block logic further prevents an efficient use of viewport-dependent decoding and is only designed for full frame data retrieval.

However, the codec could not gain wide popularity and further development was discontinued more than a decade ago. The reasons for the codecs limited success are not entirely clear, but may be related to an inability to provide significant improvement over established codecs like H264. Dirac and Schrödinger are now abandoned and no longer available.

C.3 Method

Two concepts that most video codecs apply for data compression are intra- and inter-frame coding. In practice, these methods are commonly applied with some information

loss to achieve better compression ratios. *Intra-frame coding* usually refers to the transformation of the data of one frame to a different representation that can be compressed more efficiently. *Inter-frame coding* utilizes redundancies between multiple frames to reduce the data size. In the following, we describe how we implement both concepts with wavelet transforms. Figure C.2 shows an overview of how this transformations are integrated in our program flow for the encoding and decoding of 360° videos.

C.3.1 Frame-wise Transform

The core of the frame-based compression of our codec is a 2D fast wavelet transform (FWT). Similar to other codecs, the transform of the frame data allows for a better compression, which in the raw state is too large to be stored. For example, a one minute 360° video in 8k resolution would contain around 300GB uncompressed data. We transform an input frame s of $(N \times M)$ pixels for a discrete 2D position (x, y) and frequency γ by the wavelet transform W with the mother wavelet ψ.

$$W_\psi s(\gamma, x, y) = \sum_{i=0}^{N} \sum_{j=0}^{M} \psi_{\gamma,x,y}(i, j) s(i, j) \tag{C.1}$$

To compress the transformed frame $W_\psi s$ all coefficients below a certain threshold T_s are set to zero, so that:

$$W'_\psi s \coloneqq \begin{cases} W_\psi s, & if |W_\psi s| > T_s, \\ 0, & \text{otherwise} \end{cases} \tag{C.2}$$

For a properly chosen T_s, this operation has only minor implications for the quality of the reconstructed image. This is especially true for high frequencies and is a general characteristic of the frequency domain. Usually, in natural images most of the information is contained in the low frequencies, which are represented by only a small number of coefficients [Unser and Blu 2003]. As usual for the FWT, with every step of the 2D transformation, the resolution of the image approximation is halved in both dimensions. In W_ψ this is addressed by the frequency layers over γ.

Former research has shown that the discrete wavelet transform can achieve better image reconstruction than a DCT-based method at high bit compression ratios [Boopathi and Arockiasamy 2012]. For the frame-wise transformation of the image we use the CDF 9/7 wavelet [Cohen et al. 1992]. The CDF wavelet is known to perform especially good on natural imagery and is also used for lossy compression in the JPEG2000 standard [ISO 2019].

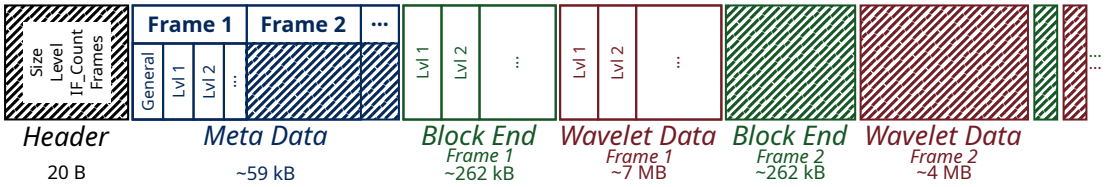

Figure C.3: Data arrangement of our video format. The sizes of each section are given by an example video in 8k resolution.

During playback, the compressed video information is decoded with an inverse fast wavelet transform (iFWT) to obtain the original images. This reconstruction is not conducted for the entire image, but only for the part of the 360° panorama that lies in the viewport of the display device. For a viewport-dependent reconstruction, we define the location of the viewport on a low resolution representation of the frame in binary form. This binary mask is uploaded together with the wavelet coefficients and is used for the inverse wavelet transform. The resolution of the binary representation is 256×256 pixels for a 8k stereo frame and can define arbitrary reconstruction shapes.

In theory, the transform can be performed until only one pixel defines the lowest frequency over the whole image. However, the low-frequency levels of the wavelet transform contain fewer discrete data points since the high-resolution in the frequency domain results in a low resolution in time due to the Heisenberg theorem [Heisenberg 1927]. Also, the wavelet coefficient values of these pixels can only be compressed inefficiently because the low-frequency information is significantly more important than high frequencies in natural image reconstruction. By default, we perform the wavelet transform for $l_{max} = log_2(\frac{N}{32}) - 2$ levels. For the number of wavelet levels, we took inspiration from the JPEG2000 standard, but also performed several pilot studies. While we found 6 levels to be the optimal default solution for 8k videos, the number of wavelet levels can be chosen individually per video.

C.3.2 Inter-Frame Coding

Inter-frame coding describes the compression of temporal information. In videos, the time component t is represented implicitly by a set of successive frames. In modern codecs the inter-frame compression is performed with keyframes and motion vectors where only the information differences are encoded. While this technique offers impressive compression rates, a compression with keyframes has the disadvantage that its speed depends on the linear information retrieval. When the video is skipped, all information since the last keyframe has to be reloaded first. With our codec we wanted to

get rid of this disadvantage and at the same time maintain a good compression rate between inter-frames. To achieve this purpose we apply a second one-dimensional wavelet transform that encodes the temporal pixel differences. The second wavelet transform is applied on a set of wavelet images resulting from the frame-wise wavelet transform $W_\psi s$. Here, we use a one-dimensional form of W_ψ with $s(\gamma, t)$ for the frequency γ of the temporal changes of every pixel over time. In other words, our wavelet based inter-frame transform encodes the frequency changes over time. Typically, even with movement in the frame the temporal information only changes on single frequencies. All frequencies that do not or only slightly change are compressed by our inter-frame transform. As result, the speed of the decoding is unaffected by the direction in which the viewport moves. In theory both, the frame-wise transform and the inter-frame transform, can be combined to one 3D wavelet transform. However, this 3D transform would not offer us the possibility to decode different areas of a frame in different resolutions for the same computational costs. Furthermore, the separation allows us to apply different wavelets and thresholds per transform and respond adaptively to individual circumstances.

Every inter-frame transform of n consecutive frames we call *inter-frame set*. Thereby, n is a power of two value. The number of frames per inter-frame set can be defined per video and may be bigger the less motion is in the video. In contrast to the frame-wise transform, the inter-frame transform is always executed to the last level. For the inter-frame wavelet transform we use the Haar wavelet [Haar 1911]. The Haar wavelet is the only wavelet with no overlapping of the wavelet filters and can therefore be reconstructed by loading only one coefficient per level for the high and low pass filtering. Reconstruction of one specific pixel by a Haar wavelet transform with n levels only requires $log_2(n)$ additions of the correct wavelet coefficients multiplied by the high pass filter position (either -1 or 1). As a result, for the inverse inter-frame transform we can iterate over the uploaded wavelets rather than over all pixels of the target section. This characteristic is unique to the Haar wavelet and allows a rapid inter-frame reconstruction. The speed of the inter-frame reconstruction is important since the inverse inter-frame transform runs on $log_2(n)$ frames every time one frame is decomposed. By iterating over the uploaded pixels rather than a target section we implicitly synthesise only the part of the image that is in our defined FOV.

C.3.3 Thresholding

In order to achieve the necessary storage savings, we have to determine which wavelet coefficients are least essential for the reconstruction. We will refer to this step as *thresholding*. The chosen threshold value is decisive for the intensity of the compression.

Thereby, the threshold always represents a trade-off between quality of the reconstructed image and size of the video file. The threshold of 0 represents lossless compression while a threshold of 1 produces the smallest file. Reconstructing an image with too little frequency information may result in a blurry representation with less details. We derive a threshold T from a user-defined constant α and the level l of the transform:

$$T(x, y) = \alpha \left(\frac{l_{max} - l}{l_{max}} \right)^2 + H, \tag{C.3}$$

where H specifies a mapping factor which depends on the mapping technique (cf. Sec. C.3.6). In the encoding, the frames are thresholded twice: once after the frame-wise 2D FWT, and again following the inter-frame transform. We use two separate threshold operations as both wavelet transforms aim for a different encoding: The frame-wise transform encodes the different frequency information in the respective spatial resolutions. The inter-frame transform encodes temporal frequency information of every wavelet coefficient. Thresholding the values only once is possible but, in our experience, can lead to unwanted interactions and a worse compression rate. Both thresholding operations are independent and have their own threshold value. While the first thresholding is applied on every frame independently, the thresholding of the inter-frames considers all n frames of the inter-frame set. In this latter thresholding operation, the different levels are defined by the relative frame number from t rather than the pixel position inside the frame.

High frequency information was found to be less important for the perceptual quality of an image than low frequency information [Unser and Blu 2003]. We scale the threshold by the frequency level of each coefficient in a quadratic function (cf. Eq. C.3). Accordingly, more coefficients may be zeroed out at high frequencies. This thresholding weighting follows common procedures of other codecs like JPEG2000 [Taubman and Marcellin 2012, Marcellin et al. 2000].

Quantization: Similar to other codecs, we represent the color values of the pixels in the video file by one byte per color component. In the quantization, the 32 bit float color components of the wavelet transform are mapped to the byte representation of the compressed output. We use the extreme values of the wavelet coefficients for normalization in order to achieve the highest possible spatial resolution in this discretization. Therefore, one discrete color value c_d is defined by

$$c_d = \frac{c_n - c_{min}}{c_{max} - c_{min}} * 255 \tag{C.4}$$

with c_n as the floating-point representation of the n-th pixel and c_{min} and c_{max} as the minimum and maximum values of all coefficients, respectively. The normalization

is performed with a separate minimum and maximum for the approximation area (last layer of the transform) and the wavelet layer and for each inter-frame. The normalization is inverted during reconstruction. The minimum and maximum values are stored with the metadata in the file.

C.3.4 File Format

The structure of our video file is illustrated in Figure C.3. We designed the layout to allow for a fast and viewport-dependent streaming of the data. Starting with a file header, general information on the video is offered. This data includes the number of frames, size of the frames and number of levels of the wavelet transform. Following the header, metadata information on every single frame is provided. This frame-wise metadata includes information about where the frame starts and ends in the file or the overall number of wavelets. The frame metadata also provides information on the individual levels of this frame. The header as well as the entire metadata are preloaded when the video is started and are kept in the working memory.

The position (x, y) of one particular wavelet coefficient within the frame is given by an index that is saved along with the wavelet value. However, due to the compression, the position of individual coefficients within the file is unknown. While it is possible to find the data for (x, y) with binary search on the video data, this inconsistent access to the storage drive adds an unwanted delay to the loading process. Instead we divide the transformed frames into a logical grid of small blocks (default size 32×32-pixel). This block allocation is only relevant for the compression but does not affect the wavelet transforms which operate on the entire images. Note, that this is different from blocking in DCT. In the video file we store one pointer for each block, located in the *BlockEnd* section (see Figure C.3). This pointer indicates where the last wavelet coefficient inside the respective block can be found in the video file. In the file the coefficients are stored block after block, which allows a whole series of blocks to be loaded by two of these block-end pointers. During decoding, the block pointers of a frame are preloaded before the frame is processed. By alternately storing the block-end and wavelet data packages of the frames in the video file we can avoid compute-intensive rearrangements of the file during encoding. Inside one block of wavelet coefficients or block-end information the data is ordered level-wise starting with the lowest frequency layer.

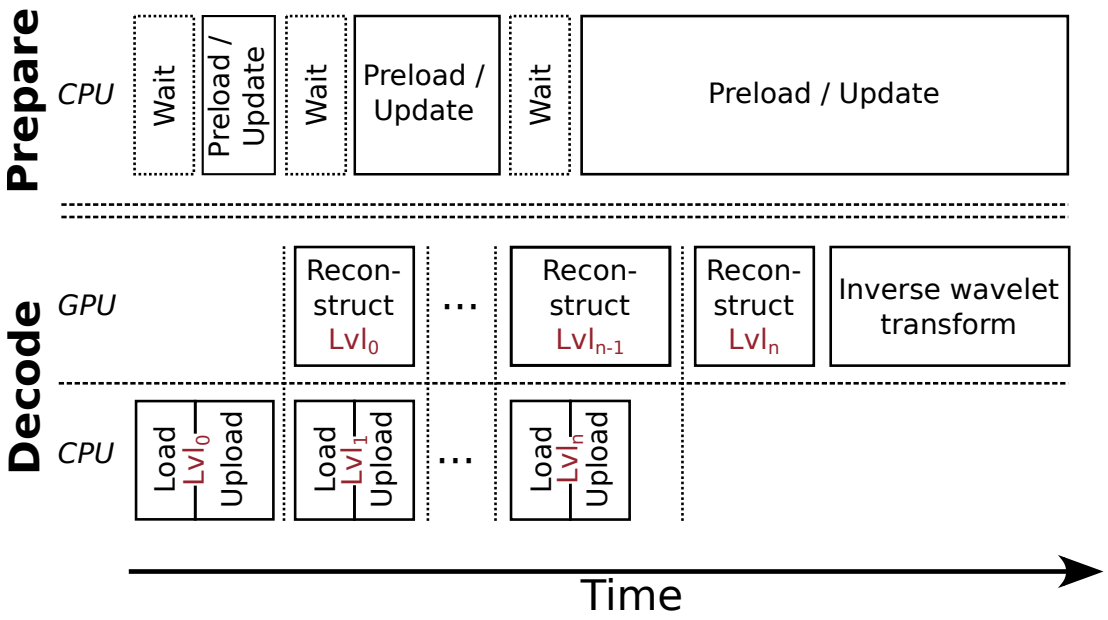

Figure C.4: Parallel processes for the reconstruction of one frame. The prepare thread buffers the next frames. In the *wait* sections the drive is occupied by the decoding process.

C.3.5 Parallel Wavelet Processing

Since VR users move their head, the part of a 360° video that is rendered can change continuously. These viewpoint changes complicate the buffering of subsequent frames with a viewport-dependent video stream. However, buffering is necessary to avoid load peaks and latencies that disrupt the virtual experience and can induce cybersickness [Stauffert et al. 2020]. Therefore, we load the viewport data of the next frames asynchronously while the current frame is decoded. Until the time of rendering, all frames are updated continuously in case that the look direction changes. The number of frames that are preloaded corresponds to the inter-frame size. Preparing more frames is not always useful, as the view direction may change strongly over longer periods of time. During our experiments no substantial latency was measured that arose from the buffering.

Due to the inter-frame compression, one frame is reconstructed by the wavelet coefficients stored in multiple inter-frames of the respective inter-frame set. We rebuild the wavelet representation $W_\psi s$ of one frame on the GPU while we already upload the wavelet data for the next inter-frame in parallel (see Figure C.4). This rebuild already includes the inverse inter-frame transform, as described in Section C.3.2. The reconstruction of the original frame by the 2D iFWT is performed once all inter-frames are processed and the inverse inter-frame transform is completed.

C.3.6 Frame Mapping

360° videos are representations of a recorded 3D sphere, brought to a rectangular frame by a projection. The position of the information in the projected frame is defined by its mapping. Among the most popular mapping techniques are equirectangular mapping and cubemaps. Our codec is defined to be independent of the mapping technique. As the reconstructed areas are given in a low resolution presentation of the frame (cf. Sec. C.3.1), the areas needed for the frame mapping can be set in direct relation to the final reconstruction. We render the final re-projection of the FOV to the spherical presentation in an own shader which is run after the inverse wavelet transform is completed. This shader can react on multiple mapping types and also considers the stereo images. For the experiments we use the equirectangular projection. We tackle the redundancies in the pixel information near the poles by gradually increasing the mapping factor H towards the poles. For an equirectangular projected frame with the dimension S we define $H(y) = 1 - \sin(y\pi/S_y)$. With the adjusted threshold, we experience equal performance at all viewing angles, including upward views.

C.3.7 Foveated Decoding

The information density of visual representations of the human eye are not equally distributed [Silverstein 2008]. The images created on the retina of the human visual system follow a qualitative decline starting from the eye's fixation point. While people can perceive the full resolution of about one sixtieth of a degree in the fovea around the focus point, the information in the peripheral visual area is significantly lower in resolution [Kolb et al. 2020].

So far we only discussed full resolution reconstructions of the viewport. However, when the eye gaze direction of the observer is available by eye tracking, we can utilise the properties of the wavelet transform to achieve what we call *foveated decoding*. With foveated decoding the resolution gradually decreases with the distance from the fovea. Our method is comparable to classical foveated rendering, except that in the periphery the decoding is accelerated while the rendering load is constant. The results are bandwidth savings and higher playback speeds.

For the foveated decoding, we utilise the level-wise structure of the wavelet transform. As described in Section C.3.1, a wavelet representation is composed of individual levels, each corresponding to a defined frequency interval γ. Instead of loading the same FOV at each level, the full FOV of the viewport is loaded and rebuild only at the lowest frequency layer. After that the data that is loaded at every level corresponds to

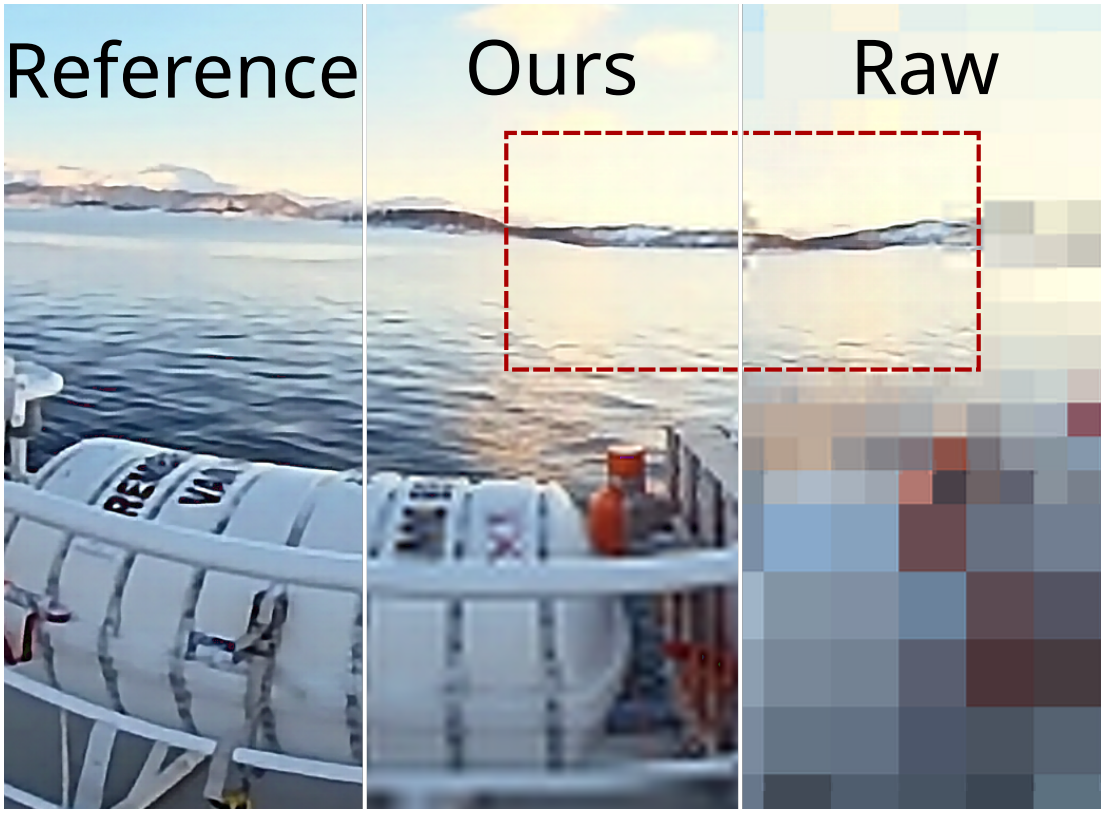

Figure C.5: Our foveated decoding compared with the full-resolution reference frame. On the right, the information density of the foveated decoding is visualised. The dashed line indicates the foveal region.

a increasingly smaller FOV. As a result the foveal area is compact and can be loaded effectively. The sizes of the individual resolution levels of the wavelet transform are defined in regard to the properties of the human eye [Leigh and Zee 2015]. Like human perception, we decrease the quality of the frames at a logarithmic rate [Kolb et al. 2020]. The area with the full video resolution which stimulates the most central foveola is only about two percent wide [Silverstein 2008].

The inverse wavelet transform is executed over the same number of data points as for a full resolution viewport but assumes the coefficients to be zero for the surrounding regions. With foveated decoding, we achieve a peripheral reconstruction in a visually appealing quality with a small number of coefficients (see Figure C.5). With the foveation, up to 80% less data has to be loaded. The reconstructed areas are in rectilinear form and follow recent findings, which indicate advantages over a log-polar presentation [Li et al. 2021].

C.4 Experiments

In objective and subjective evaluations we compare our codec against the common HEVC and AV1 codecs. The benchmarks for display speeds also include a tiled HEVC implementation from related work [Zare et al. 2016].

C.4.1 Dataset

For the evaluation, we aim to analyse a particularly diverse set of videos. Both computer-generated imagery (CGI) content and real world recordings are considered. Furthermore, we investigate moving camera trajectories as well as fixed recording positions.

The CGI reference frames are created with *Unreal Engine 5* with a high level of photorealism. The CGI scenes cover urban and nature scenery. The *City* scene mainly contains geometrical structures and straight lines and the recording takes place in a virtual New York City. The *Mountain* scene depicts a natural landscape with trees and a lake. The content of this scene contains a more heterogeneous shape composition compared to urban imagery. In the CGI scenes the camera trajectory moves along a predefined quarter circle path. The frames are rendered as 360° stereo images with 8192×8192 pixel resolution.

The first set of real-world videos is recorded with a moving camera trajectory and the display of rapid motions. Here, we use our videos from previously published work [Groth et al. 2022, 2021b] which have a higher resolution than typical moving-camera 360° videos due to our custom camera setup. The second category considers videos with a fixed recording position (further denoted by *). These videos were originally recorded by Mühlhausen et al. [2020]. The original real-world videos are recorded with stereoscopic information in 6400×6400-pixel resolution at 30 frames per second (fps).

Reference Data Creation: Pre-recorded videos can cause two problems for the evaluation. For one, the frame rates typically do not match the refresh rates of VR devices. Additionally, the data is already lossy compressed and can include serious compression artifacts. In order to address both problems, we first downscale the video data to 1024×1024-pixel to get rid of high-frequency compression artifacts and then perform temporal interpolation and upscaling of the data with state-of-the-art neural network approaches. The resulting frames are used as reference for our evaluation. The original video data is downscaled with bicubic interpolation by *OpenCV*. The temporal interpolation is run on the downscaled frames with RIFE [Huang et al. 2021]. The information from both eyes is processed individually to avoid artifacts at the edge. We

Table C.1: Display speeds. Tiling refers to Zare et al. [Zare et al. 2016]. The foveated decoding (*FD*) is run with the high resolution version of our codec ($Ours_{HQ}$), all values are averages over multiple runs and given in fps.

Videos	AVG fps ↑					
	HEVC	Tiling	AV1	$Ours_{LQ}$	$Ours_{HQ}$	$Ours_{FD}$
$City_{CGI}$	57.32	88.24	52.7	220.96	194.17	**228.8**
$Mountain_{CGI}$	58.27	87.95	61.57	222.75	202.28	**227.46**
Downhill	60.21	93.21	48.32	193.88	180.70	**206.65**
Horse	62.6	95.94	50.45	197.17	181.18	**207.16**
Climbing	65.36	92.55	52.53	195.63	187.56	**207.3**
Walking	65.32	93.1	52.76	198.24	187.88	**205.89**
Cave*	66.53	111.08	53.93	209.22	207.12	**210.28**
Boat*	67.13	110.15	55.07	205.21	201.06	**208.76**

increase the frame rate from the original 30 fps to 120 fps, which should be in line with the frequency of most modern VR glasses. For the resolution upscaling we use Nvidia VFX to create the final reference frames in 8196×8196-pixel resolution.

C.4.2 Objective Evaluation

We consider a high and low quality version of the wavelet-compressed videos. The inter-frame transform is applied in sets of four frames and compressed with an inter-frame threshold of 0.005. The frame-wise threshold is chosen to be 0.1 and 0.25 for the high and low quality version, respectively. The HEVC and AV1 encodings are performed with FFmpeg. Regarding quality, for HEVC we use a constant rate factor (CRF) of 30 (range 0–51) and for AV1 a CRF of 50 (range 0–65). The videos of both codecs are encoded in the common YUV420 colour space. The OMAF inspired tiling method is realised with HEVC encoded tiles with the fastest tiling scheme of Zare et al. [2016]. However, we extended their tiling scheme for stereoscopic videos to a 6-by-6 grid layout (6-by-3 per eye). Following the former work, the middle row of both eyes is chosen with 90° height and all other rows with 45° height for a better central view performance.

We conduct all of our experiments on a commercially available computer with a NVIDIA RTX 3090 graphics card and an AMD Ryzen 5950X processor. The video data is stored on an on-board SSD. A HTC Vive Pro Eye is chosen as output device. All videos are displayed in our self-programmed video player which uses the Vulkan

API to utilise the GPU. The decoding of HEVC and AV1 video data is performed with the Nvidia NVDECODE API. Thereby, the HEVC and AV1 decoding benefits from the hardware acceleration on the GPU. All videos are created from 1200 reference frames with 8192x8192-pixel resolution. To assure an equal comparison of all experimental conditions, we use head and eye tracking data of participant recordings.

The results regarding **computational time** are shown in Table C.1. Our proposed codec allows for an average increase in performance of 197% compared to HEVC and AV1 and an increase of 91% over the tiling technique. This increase is even more significant when the lower quality version of our codec is used. In the experiment, the foveated decoding ($Ours_{FD}$) is applied on the wavelet-based video with high quality settings. Due to the foveation, the performance increases by 223% over HEVC allowing a better performance than the lower quality wavelet-encoded videos. Please note that we used the hardware accelerated on the GPU for the decoding of HEVC and AV1. The dedicated decoding chips allow for significant increases in decoding speed compared to conventional decoding. Additionally, the compute shaders for the mapping and rendering can be executed in parallel to the decoding through the dedicated chips. A comparable chip for decoding wavelet transforms could also significantly improve the performance of a wavelet-based codec while the compute unit can be used for other tasks.

We compared the results' **quality** of all codecs by the commonly used metrics WS-PSNR, SSIM [Wang et al. 2004], and LPIPS [Zhang et al. 2018]. The given values are averages over all frames and compared with the uncompressed reference frames (see Table C.2). In terms of image quality our method performs equally to the other codecs, HEVC/H.265 and AV1, when high quality settings are chosen. As can be expected, the image quality is on a lower level when the low quality parameters are chosen for the wavelet-based encoding.

The **compression rates** of the wavelet files in both quality configurations as well as the comparison techniques can be seen in Table C.3. With our wavelet-based approach, we are able to compress the raw information to over one hundredth in size for most videos. Despite this significant reduction in file size, our codec does not yet achieve the compression efficiency of HEVC and AV1 in its current state. We would like to emphasize that the focus of this work is on decoding speed and that our wavelet codec is not optimized to produce the smallest possible files. The tiled HEVC videos by the technique of Zare et al. [2016] are on average twice as large as our wavelet-compressed video files due to the significant compression losses of the tiling process.

Table C.2: Objective Quality Comparisons. The results from the computational metrics, WS-PSNR, SSIM (higher is better), and LPIPS (lower is better), on all codecs compared with the reference frames.

Scene	Metrics	HEVC	AV1	$Ours_{LQ}$	$Ours_{HQ}$
$City_{CGI}$	WS-PSNR ↑	**17.89**	17.24	16.13	17.13
	SSIM ↑	.905	.89	.817	**.916**
	LPIPS ↓	.124	.167	.268	**.114**
$Mountain_{CGI}$	WS-PSNR ↑	**18.95**	18.06	16.44	17.43
	SSIM ↑	.927	.925	.851	**.933**
	LPIPS ↓	.158	.157	.32	**.145**
Downhill	WS-PSNR ↑	17.86	14.64	16.51	**17.99**
	SSIM ↑	.954	.95	.921	**.961**
	LPIPS ↓	.082	.103	.161	**.081**
Horse	WS-PSNR ↑	**18.32**	15.35	16.82	18.02
	SSIM ↑	.968	.966	.939	**.97**
	LPIPS ↓	**.054**	.07	.118	.057
Climbing	WS-PSNR ↑	**18.97**	18.93	17.54	18.63
	SSIM ↑	.973	.971	.952	**.976**
	LPIPS ↓	**.051**	.066	.115	.056
Walking	WS-PSNR ↑	18.02	18.13	16.83	**18.22**
	SSIM ↑	**.97**	.969	.928	.967
	LPIPS ↓	**.05**	.064	.117	**.05**
Cave*	WS-PSNR ↑	20.89	**20.99**	18.88	20.15
	SSIM ↑	.971	**.972**	.96	**.972**
	LPIPS ↓	**.039**	.04	.08	.046
Boat*	WS-PSNR ↑	19.44	**19.86**	17.54	18.84
	SSIM ↑	.984	**.986**	.954	.978
	LPIPS ↓	.03	**.027**	.078	.036

Table C.3: Compression ratios of the video files in relation to the uncompressed data. The compression ratios are with respect to the full 360° FOV.

	$City_{CGI}$	$Mountain_{CGI}$	Downhill	Horse	Climbing	Walking	Cave*	Boat*
HEVC	651:1	973:1	431:1	810:1	1059:1	1039:1	5120:1	7408:1
AV1	933:1	1482:1	725:1	1524:1	1994:1	1994:1	9888:1	12659:1
Tiling	102:1	102:1	80:1	90:1	83:1	90:1	204:1	340:1
$Ours_{LQ}$	176:1	250:1	147:1	187:1	250:1	185:1	714:1	312:1
$Ours_{HQ}$	52:1	75:1	77:1	100:1	128:1	100:1	416:1	117:1

C.4.3 Perceptual Evaluation

In a next step, we compare the codecs in terms of observer's preferences to reveal subtle perceptual differences that were not found by the objective metrics. In this perceptual experiment, we are particularly interested in analyzing to what extent the higher frame rates we can achieve with our wavelet codec also contribute to a better VR experience. As before, HEVC and AV1 serve as comparison techniques. The wavelet videos use the high quality version that showed to be most comparable to the other codecs in quality while still maintaining high display speeds. Due to the better data utilization of a wavelet-based codec, a higher resolution of the video or a higher frame rate can be provided. Thus, there are two conditions to be considered: the videos are in the same resolution but have different frame rates (*Speed* condition), or the videos are in different resolutions but provide the same frame rate (*Quality* condition). In the *Speed* condition all videos are at a resolution of 8192×8192 pixels. While the comparison techniques provide common 30fps in this condition, the wavelet videos provides double the frame rate (60fps) based on the results of the display speed measurements. In the *Quality* condition the frame rate of all videos is set to 60fps but the HEVC and AV1 videos, here, have a resolution of 4096×4096 pixels. The length of at videos is 10 seconds. For the perceptual evaluation we only use the CGI scenes. The upscaled real-world videos have a corrupted stereo view because the upscaling algorithms are not designed for stereo footage. This display error should be irrelevant for the image-based metrics, but makes perceptual experiments impossible.

As the abstract feeling of comfort in a VR experience cannot be represented by a linear scale, we utilize the paired comparisons technique [Kendall and Babington-Smith 1940]. Given the same video with two different encodings played immediately after each other, the participants were instructed to choose the video they would prefer for

Table C.4: Perceptual results. The top half refers to the uncorrected data of all participants ($n = 23$). In the bottom half the results are corrected for consistency ($n_{Speed} = 21, n_{Quality} = 15$).

Analysis	Condition	Scene	% Preferences			ζ	u	χ^2	p
			Ours vs. AV1	Ours vs. HEVC	AV1 vs. HEVC				
Raw Data	**Speed**	City	**95.65%**	**82.61%**	4.35%	0.96	0.68	48.13	<.0001
		Mountain	**100.00%**	**56.52%**	8.70%	0.96	0.55	39.09	<.0001
		AVG	**97.83%**	**69.57%**	6.52%	0.96	0.60	84	<.0001
	Quality	City	**56.52%**	**56.52%**	21.74%	0.87	**0.08**	8.13	**.0434**
		Mountain	**86.96%**	43.48%	26.09%	0.74	0.23	18.22	.0004
		AVG	**71.74%**	**50.00%**	23.91%	0.80	**0.13**	20.55	.0001
Corrected for Consistency	**Speed**	City	**100.00%**	**80.95%**	4.76%	1	0.72	46.24	<.0001
		Mountain	**100.00%**	**57.14%**	4.76%	1	0.59	38.62	<.0001
		AVG	**100.00%**	**69.05%**	4.76%	1	0.65	82.48	<.0001
	Quality	City	**80.00%**	**60.00%**	20.00%	1	**0.20**	11.4	.0097
		Mountain	**86.67%**	46.67%	6.67%	1	0.39	19.4	.0002
		AVG	**83.33%**	**53.33%**	13.33%	1	**0.31**	29.60	<.0001

a presentation in VR. The question was intentionally kept open and participants were free to base their decision on the visual quality, the temporal smoothness of the video, or a lower incidence of cybersickness.

Given three codecs, there are three possible comparisons per video: Ours vs. AV1, Ours vs. HEVC, and AV1 vs. HEVC. This results in a total of 12 decisions per person given two scenes for both conditions. In the experiment, the order of the paired stimuli comparisons for each participant was counterbalanced to avoid side effects.

A total of 23 participants took part in the experiment (10 females, age range = 22 – 59, avg. age = 34.83, SD = 13.04) resulting in 184 votes *per codec*. The videos were shown in a HTC Vive Pro Eye HMD. For every comparison, the participants were ask to maintain a fixed head position to compare the same part of the 360° video.

Analysis and Results

On a first analysis, the voting of the participants ($n = 23$) leads us to the displayed results in the top half of Table C.4 ("Raw Data", column "% Preferences"). The results show the analysis for each scene as well as an analysis per condition combining both

scenes (labelled as "AVG"). In order to assess the results of the paired comparisons, we follow the methodology from Setyawan and Lagendijk [2004]. We first study the consistency of choices within one participant as well as the agreement in choices among all participants. The coherence of the answers of the participants is indicated by the **coefficient of consistency** $\zeta \in [0, 1]$, with $\zeta = 1$ implying perfect consistency. Low consistency values of single participants can indicate that these individuals had difficulties to differentiate between the stimuli and thus, we can expect their judgement abilities to be worse than the average. As the number of methods $m = 3$, $\zeta < 1$ can only arise with the occurrence of one kind of circular triad such that $C_1 \rightarrow C_2 \rightarrow C_3$ but $C_3 \rightarrow C_1$. In Table C.4 we present the coefficient of consistency as an average over all participants. The consistency of choices of single participants does not necessarily mean that identical choices were made between all participants. The diversity of preferences for the number of participants n is described by the **coefficient of agreement** u. Complete agreement is achieved with $u = 1$, meaning that all participants favoured the same method for all decisions. High disagreement indicated by low u values, on the other hand, suggests that participants had difficulties to make a joint choice. Disagreement may suggests either that the stimuli were perceived as not distinguishable or a general split of opinions about the stimuli. Here, the minimum u, and accordingly complete disagreement, is given by $u_{min} = -\frac{1}{(n-1)} \approx -0.045$. By the definition of Kendall and Babington-Smith [Kendall and Babington-Smith 1940], the coefficient of agreement u is derived by

$$u = \frac{2\tau}{\binom{n}{2}\binom{m}{2}}, \text{where} \tau = \sum_{i=1}^{m}\sum_{j=1}^{m}\binom{a_{ij}}{2} \tag{C.5}$$

with a_{ij} is the number of times method i is chosen over method j, whereby $i \neq j$. As before, the number of participants is denoted by n and the number of methods by m.

We determine the statistical significance of u by testing against the null hypothesis that all votes were chosen randomly. For a significant u we can conclude the alternative hypothesis that the agreement is above the value one would expect from random choices. To determine the significance under the null hypothesis we perform a chi-squared test (χ^2). As proposed by former research [Siegel and Castellan 1988], with our n we can derive χ^2 in simple form.

$$\chi^2 = \binom{m}{2}[1 + u(n-1)] \tag{C.6}$$

Since we compare three codecs, the χ^2 distribution is evaluated with $\binom{m}{2} = 3$ DoF. Therefore, the statistical significance at level p derives by χ^2_3.

The analysis of the raw data for consistency and agreement is shown in Table C.4.

While the participants show a high consistency, the coefficient of consistency ζ for the *Quality* condition is considerably lower than for the *Speed* condition. The difference in agreement is even more pronounced and the participants show clear disagreement in the ratings of *Quality* (numbers highlighted in red). This mismatch is most apparent in the city scene where results are not statistically significant anymore for the targeted threshold of $p < .01$.

A more in-depth analysis of the consistency of individual ratings shows the cause for this outcome. In the *Quality* condition, a total of eight participants were not able to properly distinguish between stimuli, leading them to produce triads. Also, two participants had difficulties emitting judgement in the *Speed* condition. All the rest of our participants show perfect judgement capability ($\zeta = 1$, no triads), indicating that the inconsistencies arised from the individual participant's ability to judge and not from a problem with a consensus between participants. Therefore, we correct the analysis for a consistent result and remove the votes for the entire condition of all those participants who did not show perfect consistency. By rerunning the analysis with the votings corrected for perfect consistency ($n_{Speed} = 21, n_{Quality} = 15$) we can obtain the values corresponding to participants with perfect discerning abilities. These results are show in the bottom half of Table C.4 ("Corrected for Consistency"). While the *Quality* condition still seems to be more controversial, the agreement of the participants for this condition increases substantially (numbers highlighted in green). With the correction, all differences in preference are statistically significant for $p < 0.01$.

In general, the results of the perceptual study indicated that the videos using both our codec and HEVC were favored over AV1 for all conditions. The conclusions of the comparison between our codec and HEVC are dependent on the analyzed condition. While the participants clearly favoured our codec in the *Speed* condition, the perceived quality overall is on par with the H.265 compressed video. A more in-depth analysis reveals the choice of scene as a decisive factor for the participants' preference. Straight lines and geometric structures are key attributes of urban scenery. In these representations with man-made content, quality differences and artifacts seem to be more conspicuous. The *Mountain* scene, on the other hand, is a natural scene and, as such, its content is less structured and highly semantically homogeneous due to the recurring textures. In this scene, the participants showed difficulties to distinguish the quality of the different compression techniques. These observations are consistent with the findings of previous research [Rubinstein et al. 2010, Castillo et al. 2011]. In the *Speed* condition, the scene attributes seem to became even more salient for the observers' choice, drifting their preferences towards our method (preference $> 80\%$).

C.5 Discussion and Limitations

In the following, we discuss further key points of consideration and address limitations of the current implementation

Experimental Results. We compared our wavelet-based codec against two common video codecs and previous work. For the evaluation, we considered a low and high quality version of the wavelet-encoded videos, because in a practical application either quality or speed may be prioritised. The results show that the codec can be optimised for such requirements by changing the encoding parameters. However, even at the highest quality, we achieve significantly higher decoding speeds than the other methods. The foveated decoding technique leverages the properties of the human visual system, resulting in peripheral resolution differences that are unnoticeable to VR users compared to fully resolved viewports [Leigh and Zee 2015]. At the same time, the foveated decoding allows for the highest decoding speeds. In a perceptual experiment, we studied the importance of the frame rate and video resolution for the perceived quality of the VR experience. The results suggest that for 360° videos in VR, the frame rate is of significantly greater importance to users than the image quality. This emphasis on frame rate for the user experience is consistent with former research [Banitalebi-Dehkordi et al. 2015].

Single Operation Point. Our codec is designed for a very specific use-case. The core motivation is to have 360° videos with a resolution and display speed that does not induce cybersickness and is pleasant to watch. In our investigations, we explored in detail the single operation point that best covers this scenario. For a fair comparison we chose the parameters of our codec so that the quality is on the same level. With this baseline, we then measured the speed of the methods. A broader range of quality-rate scenarios can be explored in the future to utilize wavelet-based coding for a verity of applications.

Streaming. So far, we have primarily addressed videos that are stored on a local drive. Online streaming is another common way to retrieve video data. With online streaming, the amount of data that is transmitted is much more relevant due to bandwidth limitations. For these limitations, a wavelet-based codec benefits from the direct viewport-dependent streaming from file. This property allows to reduce the transfer rates by up to ten times compared to the total size of the video. In its current form, our encoding is not yet optimized for real time execution, which we plan to address as a natural next step. Especially for standalone HMDs, over-air transmission latencies and data transfer rates are often a bottle-neck of the system. In this scenario it would

be beneficial to move the decoding of the wavelet videos to be performed by the HMD hardware. The encoded version of the video is then streamed and decoded on the HMD, thus, reducing the needed bandwidth and transfer time. Given the current hardware development, decoding times can be expected to be inferior to a dedicated consumer GPU. Nevertheless, with further improvements of standalone VR HMDs, an on-device decoding will become highly interesting for cloud-operated video streaming.

File Size. Our wavelet format does not use any container format but is stored in simple binary form. Neither is a color transformation performed, for example to the YUV space. Such techniques are applied by other codecs to reduce their file sizes to the minimum while preserving the best possible quality. In this paper the major focus was on display speed. In future work such techniques may be introduced to further reduce the file sizes of wavelet-based video coding.

Reference Data. Our objective with the reference data scaling of the real-world videos was to generate uncompressed high-resolution, high frame-rate video data. We used a combination of downscaling followed by AI-based upscaling to remove compression artifacts from the original videos. This removal is not perfect and it can be assumed that the compression rate of a wavelet-based codec is significantly higher for raw footage. Such a use of a wavelet-based codec can only be achieved when the encoding is directly performed by the capturing device with the native color information.

Professional Filming. The videos from our experiment are considered as casual recordings. Nevertheless, 360° videos are not only used by amateurs, but also by professional filmmakers. For professional filming, it can be necessary to display different areas of a frame in different qualities, such as the background or the masks of an actor, which stands out as artificial in high resolutions. With conventional methods, this procedure requires post-processing or recapturing of the video. With a wavelet-based codec a pre-adjustment is not necessary and the video can be stored in full resolution. Individual quality levels may be chosen at decoding time for defined parts of the video, comparable to our foveated decoding approach (cf. Sec. C.3.7).

Eyetracking. In VR, eye tracking is nowadays mostly used for computer-generated content, where foveation allows for significant increases in rendering speed. The foveated decoding of our codec opens up the opportunity for an broader use of eye tracking in VR where it can be used to increase the playback speed of 360° videos through unobtrusive quality gradation in the peripheral area.

Wide FOV. When the FOV gets unusually wide, this would affect the performance of our approach since the decoding is viewport dependent. The headset with the widest FOV currently on the market is the *StarVR One* with an overall horizontal FOV of

210° and a vertical FOV of 130° [HMD]. Exploring this scenario, we found that with wavelet coding we are still able to achieve > 100 fps with the high quality configuration in all scenes. With foveated decoding applied, the frame rate is significantly higher. In comparison, the tiling approach with this wide FOV no longer yields any performance benefit and actually performs worse than the native full frame HEVC decoding (≈ 50 fps).

C.6 Conclusion

In this paper we proposed wavelet-based video coding for fast and high-resolution playback of 360° videos. We showed that our wavelet-based compression approach allows for selective loading and decoding of arbitrary video regions, which in the case of 360° videos is key for a fast decoding. While in our experiment our codec reached display speeds at least two times higher than the other methods tested, the quality remained at a comparable level. The importance of high frame rates for a good VR experience is supported by the results of our perceptual experiment. In addition, with our codec we have introduced foveated decoding, allowing for an unobtrusive quality decrease in the outer regions of the view. Foveated decoding can be applied on run-time and further increases the decoding times. In conclusion, wavelet-based video approaches solve the problems that are raised by DCT codecs when a fast or viewport-dependent playback of 360° videos is required. Especially for VR environments, wavelet-based codecs show to be a valuable extension, offering the opportunity to display 360° videos in a quality and speed comparable to renderings of virtual worlds.

D Visual Techniques to Reduce Cybersickness in Virtual Reality

Cybersickness is a unpleasant phenomenon caused by the visually induced impression of ego-motion while in fact being seated. To reduce its negative impact in VR experiences, we analyze the effectiveness of two techniques – peripheral blurring and field of view reduction – through an experiment in an interactive race game environment displayed with a commercial head-mounted display with integrated eye tracker. To measure the level of discomfort experienced by our participants, we utilize self-report and physiological measurements. Our results indicate that, among both techniques, reducing the displayed field of view up to 10 degrees is most efficient to mitigate cybersickness.

Colin Groth and Jan-Philipp Tauscher and Nikkel Heesen and Susana Castillo and Marcus Magnor
Institute for Computer Graphics, TU Braunschweig, Germany

Extended abstract published in *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) 2021,* Pages 486–487. Presented at *IEEE VR 2021.*
DOI: 10.1109/VRW52623.2021.00125

Figure D.1: This paper explores the impact of two different visual techniques – i.e., peripheral blurring (PB) and field of view reduction (FOVR) – to mitigate cybersickness in user-explorable game environments. Towards this goal, we conduct an experiment recording both self-reported (SSQ) and physiological (heart rate and electrodermal activity) data from our participants.

D.1 Introduction

Virtual reality (VR) has opened up the possibility to experience virtual environments with an unprecedented degree of visual realism and immersion. This often comes at a price as virtual experiences may induce cybersickness (CS). CS describes symptoms that are quite similar to those of motion sickness induced by an information mismatch between the visual and vestibular system of the human body as compared to physical motion [Dużmańska et al. 2018]. Besides being inconvenient in entertainment systems, simulators, or high-risk fields like telemedicine, this phenomenon also causes ethical concerns in exposing users willingly to these symptoms. Most importantly, it may limit the adoption of VR technology. Aside from hardware-related factors that can mitigate CS to a certain extent (e.g., high frame-rate renderings, high quality tracking, and reduced-latency systems), multiple techniques have been presented towards weakening the effects of CS by only manipulating the visual information. Typically, these visual techniques aim to limit the amount of optical flow and, thus, reduce the visual movement perceived by the user. Dynamically altering the user's field of view (FOV) based on the virtual movements of the scene is one common visual approach to mitigate CS. Also, gaze-contingent approaches were proposed using real-time eye tracking [Adhanom et al. 2020]. Other methods reduce the image quality of non-foveal regions to mitigate CS by applying blur [Lin et al. 2020b] or sparse rendering [Patney et al. 2016]. Unfortunately, most of the existing visual techniques are intrusive or inefficient.

This work investigates how opaque occluding and blurring users' peripheral FOV affect CS in VR. The techniques are designed to be unobtrusive while at the same time mitigating CS in a reasonable manner. To explore their effectiveness, we present an experiment using commodity eye-tracking hardware that displays a virtual race game scenario.

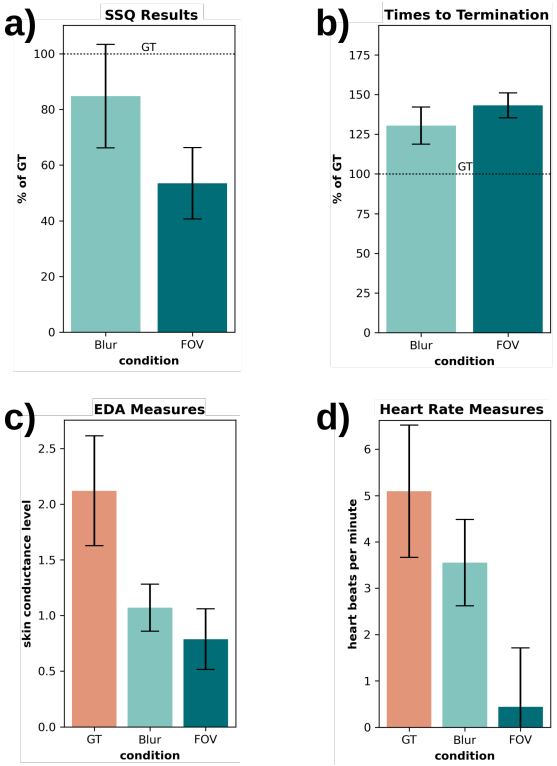

Figure D.2: Statistics averaged for all conditions. Error bars represent the standard error of the mean. **(a)** SSQ relative score differences. **(b)** Average relative termination times. Only participants that aborted at least one of the sessions are considered here (N=6). **(c,d)** Averaged differences of the physiological measurements – **(c)** electrodermal activity (EDA), **(d)** HR.

D.2 Methods and Experiment

We investigate the efficiency of two techniques, i.e., peripheral blurring (PB) and field of view reduction (FOVR). These techniques aim for unobtrusiveness by using eye tracking beside the scene information to gaze-contingently manipulate the peripheral areas of the view. FOVR fully grays out the region outside of the circular foveal area, while PB gradually blurs the peripheral region, with the blur level increasing with the distance to the viewing point. The diameter of the restrictors is based on the linear accelerations and angular movements of the camera's point of view in the virtual scene but is always kept above the minimum of 10° [Weier et al. 2016].

We apply these techniques on a first-person VR racing simulator (cf. Fig. D.1). The scene is designed to evoke a high level of CS by relying on well-known theo-

ries [Reason and Brand 1975, Riccio and Stoffregen 1991] and current research on its causes, e.g., ground sway, sharp curves and uneven roads [Golding et al. 1997, Tauscher et al. 2020]. To measure the level of CS we chose the simulator sickness questionnaire (SSQ) [Kennedy et al. 1993], filled out once before and after immersion. Additionally, we used physiological measures – electrodermal activity (EDA) and heart rate (HR) – associated with CS [Tauscher et al. 2020].

The experiment (approved by the corresponding ethics committee) followed a within-subject design with three sessions per participant, featuring the two restrictors as well as a unrestricted ground truth as comparison, in random order. At least 48 hours lay between the sessions to avoid any carry over effects. A total of 19 participants completed all three sessions (8 females, age range 21–51, mean 27.21, SD 8.02). Their task was to drive fast on a circular road for 10 minutes with the opportunity to leave the experiment in case of severe sickness symptoms. Before the main task, one minute was spent in the virtual environment without any motion to obtain physiological baseline data. During the trial participants were seated. They wore a HTC Vive Pro Eye and a motion controller was used to drive the car.

D.3 Results and Discussion

For the analysis of the experimental results we conducted two-sided dependent t-tests (Bonferroni-corrected) between all combinations of the conditions and a one-way ANOVA. The results indicate the effectiveness of both PB and FOVR on the proposed scenario.

Fig. D.2(a) illustrates the results of the SSQs for both conditions averaged among participants. Both methods were able to reduce the overall sickness level compared to the unrestricted scene. Occluding the FOV showed to be most efficient based on the SSQs ($t = 3.63$, $p = 0.0019$).

As expected, the results confirm that, in general, the participants' experienced level of sickness was directly proportional to their capacity of endurance and, thus, the time they spent in the experiment. The average time of the experiment until dropping out in at least one of the sessions are shown in Fig. D.2(b). Both visual techniques significantly increased the time people were willing to spend in the virtual environment (PB: $t = -2.6$, $p = 0.048$; FOVR: $t = -5.49$, $p = 0.0027$), with FOVR being the most efficient among both. Please note, that the data for one participant will always depend on their own characteristic resilience level, thus the need of gathering all data on a ground truth condition (GT) – scenario without any mitigation technique – to establish the individual baselines per participant. Thus, all previous values were calculated relative

to these baselines.

In contrast, the nature of the physiological measures allows for establishing individual baselines per session and thus, both EDA and HR scores are obtained by the average difference between the baseline and main task data per participant for each session. These values can be seen in Fig. D.2(c,d).Consistently with the results of the SSQ, both techniques achieve a substantial reduction of the measures compared with the averaged scores in the GT condition. The most significant change follows, again, when applying the FOV occlusion to the scene (EDA: $t = 2.36$, $p = 0.03$; HR: $t = 2.32$, $p = 0.033$).

Summarizing the results, both visual techniques efficiently reduce CS in VR scenes where the user is driving the experience, whereas the FOV restriction technique should be preferred when both come into question.

D.4 Conclusion

We investigated the effectiveness of two visual techniques to reduce cybersickness in VR, peripheral blurring and field of view reduction. Our results demonstrate that both techniques can be efficient for this reduction, and, among both, the dynamic reduction of the user's FOV seems to better contribute to a more pleasant experience.

Even though the level of experienced CS decreased in the experiment, the presented scene was a rather extreme scenario. Therefore, our future work will explore the applicability of the presented techniques to everyday VR scenes. Also, we like to investigate the effect of posture (free walking vs. standing still vs. seated) on the results.

E Mitigation of Cybersickness in Immersive 360° Videos

We investigate the mitigation of cybersickness (CS) in 360° videos, a phenomenon caused by the visually induced impression of ego-motion while being physically at rest. We evaluate the effectiveness of scene modulations to reduce motion in the peripheral visual field by deliberately blurring or opaque occluding eccentric view areas of up to ten degrees. Our results indicate that both methods effectively reduce CS in pre-recorded 360° video with the dynamic opaque occlusion method yielding best results.

Colin Groth and Jan-Philipp Tauscher and Nikkel Heesen and Steve Grogorick and Susana Castillo and Marcus Magnor
Institute for Computer Graphics, TU Braunschweig, Germany

Workshop paper published in *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) 2021,* Pages 169–177. Presented at *IEEE VR 2021.*
DOI: 10.1109/VRW52623.2021.00039

Mitigation of Cybersickness in Immersive 360° Videos

Figure E.1: We investigate the effectiveness of two visual techniques to mitigate cybersickness in virtual reality with pre-captured 360° videos. Here, we show exemplary frames of the videos used in our experiment.

E.1 Introduction

Virtual reality (VR) has been around for a very long time and nowadays reaches a broader audience than ever before [Rheingold 1991]. The technology is constantly improving and the industry is releasing an increasing number of games, videos and even movies for VR devices. This raises the bar for expectations and the acceptance of the general public. 360° videos in VR are a novel viewing experience that allows the spectator to be deeply immersed in the content. The viewers can turn their heads in arbitrary directions while a linear story unfolds around them. Similar to traditional linear video formats like TV shows or movies, the story takes place from the position of the camera but the viewers are in control of their own gaze. While a VR world gives the user the freedom of 6 DOF navigation, videos without depth information push them back into the passive consumer seat as they cannot walk around or modify objects. Still videos remain to be experienced in VR and therefore provide a *virtual reality experience* [Tse et al. 2017].

But any discomfort the user may experience while watching video in a VR device could potentially impact the spread of this technology. A major problem common to all types of immersive virtual experiences is cybersickness (CS) [Rebenitsch and Owen 2016, Keshavarz et al. 2014]. This term describes any physical discomfort caused by visually perceived motion that does not correspond to what is actually experienced [Irwin 1881, Lawther and Griffin 1988]. CS describes an extensive collection of symptoms which include, in a low state, oculomotor conditions (e.g., blurry vision), headaches, and dry eyes. In severe cases, the user can even experience disorientation and nausea [Kennedy et al. 2010, Stanney et al. 1997]. One of the most accepted theories, the sensory conflict theory [Reason and Brand 1975], establishes a mismatch in motion perception between the visual and vestibular system as its main cause [Reason and Brand 1975]. This conflict can be triggered in both directions: the person moves while the scene in their

field of view remains static, or when the impression of ego-motion is induced while in fact the individual remains stable [Walker et al. 2010]. Physical negative effects are not the only reason why reducing CS in VR experiences is critical. The occurrence of CS may raise ethical concerns by knowingly exposing participants to discomforting situations and can impact the validity of results from simulation-based studies (e.g. driving simulations [Fisher et al. 2011]). Thus, there is a strong need to prevent CS.

Nowadays, the most common way to prevent CS is to directly avoid simulations that may cause it [Dobie 2019] but this limits the VR experiences that can be presented e.g. no camera motion would be possible in 360° videos. Often, motion is central to videos but the user cannot influence the camera's trajectory. Therefore it is the focus of our current study to mitigate CS in this scenario.

Given that the motion of objects is primarily perceived through peripheral vision of the human eye [Webb and Griffin 2003], manipulating the peripheral area of a scene may reduce CS. In theory, as the foveal area is responsible for sharp central vision, this manipulation should remain unobtrusive. Eye-tracking technology is available for most modern VR hardware to follow the eye gaze and determine the foveal regions of the screen [HTC Vive Pro]. Previous studies already employed manipulation of the users' view but most of them are not gaze-contingent and therefore not subtle due to their static nature, or are not directly focused on the reduction of CS [Lin et al. 2020b, Nie et al. 2017, 2019, Patney et al. 2016, Weier et al. 2016].

In this work, we investigate two visual techniques to gaze-contingently constrain the peripheral field of view (FOV) by applying *blur* and *opaque occlusion* to real-world 360° videos in VR. The FOV is dynamically modified to reduce the peripheral optical flow to which the human eye is most sensitive to. The level of peripheral reduction is determined either by translucent blur [Lin et al. 2020b] or complete occlusion to establish a perceptual threshold. We compare both techniques and their effectiveness on CS. Mitigating CS in real-world videos is of major interest but only few studies have been conducted to prevent sickness in VR such videos. Therefore, our experiment uses real-world videos to investigate the effectiveness of these methods to be used in practice. The experimental results confirm the effectiveness and unobtrusiveness of the proposed methods.

E.2 Previous Work

E.2.1 Theories on Cybersickness Occurrence

There are several theories on the cause of CS and it is still controversial as to which theory is most likely to apply [Dobie 2019].

The most common reason for the cause of CS among the theories is a conflict between different modalities (vision vs. vestibular system) that are expected to be congruous as the foundation of the symptoms [Reason and Brand 1975, Reason 1978, Treisman 1977]. It is called the sensory conflict theory [Reason and Brand 1975]. Reason and Brand describe sensory rearrangement as a core principle of this theory. Sensory rearrangement arises when the perception of one modality in a given situation does not match the perception of functionally related modalities. According to them, the vestibular system has to be involved in any sensory conflict causing CS. Golding et al. [2006] address this thought and even state that any person with an intact vestibular system will experience CS at some degree when the simulation is inducing enough.

Their theory defines two categories for situations that are likely to cause CS: visual-inertial rearrangements (i.e. conflicts between modalities) and canal-otolith rearrangements (i.e. conflicts within the vestibular system) [Reason and Brand 1975]. Furthermore, these categories can be divided into three subcategories each differing in what system sends contradict information (Type 1: both; Type 2 and 3: just one). From this follows, that CS is a visual-inertial rearrangement of type 2 with the visual system sending contradict information and the vestibular system remaining passive.

It was empirically proven that people can adapt to CS over multiple sessions[Reason 1978]. Unfortunately, this adaption was not sufficiently explained by the sensory conflict theory. Therefore, Reason revised his work and published the neural mismatch theory [Reason 1978]. This theory states that CS results from a mismatch between the perceived and expected perceptions in a given situation and not between the perceptions themselves. The name is derived from the mismatch in the brain between the neural activations from the perception and the stored neural connections from learned experiences. With the neural mismatch theory, the adaptation to CS can be sufficiently explained, as the mismatch signal is stored with repeated occurrence.

For Treisman [1977] CS results from evolutionary reasons. In her toxin detector theory she argues that CS arises from problems with motor coordination caused by conflicts between the spatial orientation systems of the body. In her point of view CS is adaptive for the human in an evolutionary sense and the symptoms arising from CS are a warning sign of neurotoxin poisoning which is tried to be removed.

Riccio and Stoffregen [1991] argued that sensory conflicts not only occur in situations that induce sickness but also in non-sickness-inducing situations. Therefore, the sensory conflict cannot be the only reason for CS. Furthermore, because the sensory conflict theory depends on expectations that cannot be measured, it cannot be falsified. They suggest that CS is rather based on prolonged postural instability causing the symptoms to occur. Nowadays, their postural instability theory is also one of the most respected theories for CS occurrence [Riccio and Stoffregen 1991].

E.2.2 Cybersickness and 360° Videos

Real-world 360° videos in VR are of interest due to their sophisticated form of viewing a scene. This form of presentation is considered separately from computer-generated content as recorded content allows for pre-computation of properties like angular velocity and linear acceleration of the camera motion and use them without incurring any latency penalty. In contrast, computer-generated virtual worlds have no pre-defined camera trajectories and the camera's position in the world can only be considered at run-time. Still, the number of works addressing CS in 360° videos is small, especially when it comes to mitigate its occurrence.

The fact that 360° videos can actually provoke CS was demonstrated by former works [Elwardy et al. 2020, Kim et al. 2018a]. Elwardy et al. [2020] evaluated CS in 360° videos with an head-mounted display (HMD). Particular attention was paid to the influence of the level of experience with immersive media on CS. The results indicate that especially participants with a low level of VR experience suffered from CS in the 360° videos.

The high risk of CS of viewers exposed to 360° videos was also recognized by Kim et al. [2018a]. They proposed a neural network solution that predicts a sickness score for videos that are experienced in VR. Thus, they do not try to mitigate CS rather than warn about videos that are most likely to make users sick. Their experimental results reveal two things: 1) prediction correlates for the human perception exist and 2) 360° videos are very likely to cause CS in VR.

The video format (monoscopic, stereoscopic) and audio format (stereo, spatialized) was found to not influence the users' feeling of presence and CS for 360° videos [Narciso et al. 2019]. In contrast, gender is an significant variable that should be considered for VR scenes that display 360° videos [Narciso et al. 2019].

Bala et al. [2018] started investigations on CS mitigation for real-world 360° videos. They used an independent background grid and a fixed FOV restriction as well as a combination of both to reduce CS in their experiment. Their results did not show any

significant difference in CS which was according to their own statement due to the small number of participants [Bala et al. 2018]. In a later work with 360° videos they included more participants and only focused on the combined method (independent background grid and FOV restriction) and were able to show a decrease of CS [Bala et al. 2020].

E.2.3 Mitigation Techniques for Virtual Scenes

Given the ubiquity of CS, a lot of works study the mitigation of CS in user-controllable virtual environments like games that usually give the user 6 degree of freedom (DoF) to explore the scene.

General concepts to reduce CS in immersive experiences include high frame-rate renderings, high quality tracking and reduced latency systems [LaViola Jr 2000, DiZio and Lackner 1997, Sherman 2002]. Since all these approaches require special hardware and setup, visual techniques that do not require this hardware are used increasingly.

Visual methods directly modify the user's perception of motion by reducing the optical flow of the rendering. This is typically done by reducing the users' FOV, either opaque [Fernandes and Feiner 2016, Bos et al. 2010, Lin et al. 2002, Seay et al. 2001, Norouzi et al. 2018, Lim et al. 2020] or semi-transparent [Lin et al. 2020b, Budhiraja et al. 2017, Hillaire et al. 2008, Patney et al. 2016, OculusFFR]. Also, methods that modify the FOV were found to be equally efficient for all genders and do not influence the spatial navigation performance [Al Zayer et al. 2019].

Existing methods commonly employ fixed FOV restrictors that do not follow the eye movement and are centered on screen [Bos et al. 2010, Lin et al. 2002, Seay et al. 2001, OculusFFR]. As these fixed restrictors are clearly visible, they may impact the quality of the VR experience [Adhanom et al. 2020].

Few previous studies use gaze-contingent restrictors to only manipulate the peripheral regions. A first approach is demonstrated by Adhanom et al. with an opaque FOV modulation and the use of eye tracking [2020]. In their experiment the gaze-contingent FOV restrictor was compared against a fixed FOV restriction. The results state that although both method achieved comparable CS scores, the fixed restriction is not only obvious but influences the users' eye gaze behaviour. Conversely, the gaze-contingent modulation was unobtrusive and allows for a natural gaze behaviour of the user [Adhanom et al. 2020].

The potential of semi-transparent FOV modulations to mitigate CS is investigated by Hillaire et al. [2008], who investigated the performance of gamers in response to visual blur effects. They show no performance degradation arise from blurring the periphery. The participants even reported an increase in realism with the active blur [Hillaire et al.

2008]. The recent work of Venkatakrishnan et al. [2020b] used a fixed semi-transparent central window to reduce CS in a explorable VR environment.

Other visual methods to reduce CS involve non-salient blurring of objects in the virtual environment [Nie et al. 2017, 2019], snapping of moving frames [Farmani and Teather 2018, Weißker et al. 2018] and skipping or obscuring translational frames with virtual teleportation [Moghadam et al. 2018, Weißker et al. 2018]. While these techniques promise to potentially minimise CS, they represent a compromise and are often connected to a negative user experience or feeling of presence [Lin et al. 2020b, Moghadam et al. 2018, Weißker et al. 2018].

In this paper, we are the first to directly compare the effects of gaze-contingent peripheral blurring against gaze-contingent peripheral occlusion in the context of seated viewing of pre-recorded 360° videos involving complex camera movements.

E.3 Experimental Framework

We investigate the mitigation of CS in a VR experiment with different visual techniques. The techniques are designed to unobtrusively influence the performance of our participants as little as possible. For this, real-time eye tracking of the HMD and pre-recorded gyroscopic data of the camera is used to only manipulate the peripheral areas of the view. The effectiveness of the methods is validated using psychophysical and physiological measurements.

E.3.1 Stimuli Generation

Restrictor Design

The goal of our visual techniques is to efficiently mitigate CS and be as subtle as possible to the user. In our experiment, we gaze-contingently post-process the presented video in real-time by restricting its periphery during motion by either blurring or opaquely occluding the eccentric area before presenting it on the HMD. For both techniques the foveal region remains unchanged for a natural viewing experience. The two techniques are in the following referred to as *peripheral blurring* and *opaque occlusion*.

The restriction with peripheral blur is realised by a post-process spiral blur filtering technique inspired by the Unreal Engine. The applied filtering technique achieves a uniform blur with a maximum of 64 samples per pixel i.e. 8 distance steps x 8 radial samples. The blur intensity increases radially to the outside [Lin et al. 2020b]. Thus, the farther away a point is from the foveal region the more it is blurred (cf. Fig. E.2).

Figure E.2: Video frame of the unmodified ground truth condition (GT), with applied peripheral blurring (Blur) and opaque occlusion (OCC) of the user's FOV.

The opaque occlusion is implemented in a similar way but instead of blurring the peripheral region, it occludes it completely. Only the foveal region is visible the whole time. The intensity of the occluding gray does not change throughout the scene and was chosen to match the average color intensity of the videos.

The design of both restrictors is shown in Figure E.2.

The diameter of the visual modification techniques changes with the camera motion in the scene but is always kept above a minimum size of 10° to prevent occlusion of the whole scene and be less obtrusive. According to previous research, users are not able to reliably differentiate between full resolution and foveated rendering when the high-resolution area is larger than 10° [Weier et al. 2016].

During the experimental session, we change the size of the circular restrictors based on the linear acceleration a and angular velocity ω of the camera in the scene. This movement data was recorded by the camera's built-in gyroscope along with the video. We link the restrictors' opening diameter to camera acceleration, because the human vestibular system also perceives motion through acceleration in three dimensions [Cullen 2012]. For rotational movements, the angular velocity is taken into account as it is processed similar to linear accelerations in the vestibular system [Meiry 1965]. The restrictor is absent, i.e. fully open, when the camera is not moving following the prevalent theories on CS occurrence [Reason and Brand 1975, Riccio and Stoffregen 1991, Reason 1978]. The restrictor size r at time point t is calculated as in Equation E.1. Here, r_d

represents the full range of the restrictor: $r_d = r_{max} - r_{min}$.

$$r_t = r_{min} + r_d \left(1 - \left(\frac{|a_t|}{2a_{max}} + \frac{|\omega_t|}{2\omega_{max}}\right)\right) \tag{E.1}$$

The maximum linear acceleration a_{max} and angular velocity ω_{max} is fixed to the maximum values of the scene. The restrictor sizes are filtered frame-wise with an exponential moving average filter, with 10% influence of the new size and 90% of the former size, to reduce possible flickering noise.

Virtual Environment

The video content is produced to evoke a high level of CS by relying on well-known theories [Reason and Brand 1975, Riccio and Stoffregen 1991, Reason 1978] and current research on the causes of CS [Lin et al. 2020b, Adhanom et al. 2020, Bala et al. 2020].

We present multiple 360° videos in an HMD. Our pre-recorded 360° videos showed scenes from parkour running and sport climbing with free falls up to 5 meter. We recorded the videos in ego-perspective in collaboration with professional athletes (see Figure E.1). Also, in contrast to controlled, artificially designed scenes, real-videos may often expose a higher visual complexity. We used real-world videos to account for this complexity to investigate the effectiveness of the visual techniques when applied in practice.

Cinematic views like videos are known to cause CS and are therefore highlighted in VR interaction guidelines of leading game engines [UnityVRDesign, Unreal Engine Virtual Reality]. This sickness induction is supported by the fast movements and abrupt stops of the athletes according to the sensory conflict theory [Reason and Brand 1975] and falls and swings supporting the postural instability theory [Riccio and Stoffregen 1991]. Furthermore, some abrupt movements cannot always be anticipated and also increase CS according to the neural mismatch theory. This unpredictability occurs especially when the videos are watched for the first time and explains why we employ a counterbalanced experimental design to prevent adaptation to CS. In our experiment, participants had no control over the video except that they could change their viewing direction by head movement.

E.3.2 Measurements

The level of cybersickness one person is suffering from in a given experience can be measured by both subjective and objective measurements. Each of those methods are used in our experiment.

Subjective Measurement

A frequently used method to measure CS in virtual environments is the simulator sickness questionnaire (SSQ) [Kennedy et al. 1993]. It defines 16 sickness symptoms and derives from the more general pensacola motion sickness questionnaire (MSQ) [Kellogg et al. 1964]. The SSQ is known as an efficient and effective tool to self-assess the level of CS. In addition to the total sickness score, three clusters can be differentiated: nausea, oculomotorics and disorientation. We use the SSQ for subjective measurements of sickness values since simulator sickness evokes the same symptoms as CS experienced in VR [Mazloumi Gavgani et al. 2018]. Furthermore, it was used by multiple previous studies for 360°videos [Narciso et al. 2019, Kim et al. 2018a, Lin et al. 2017, Elwardy et al. 2020].

The total sickness score as well as the corresponding subscores of the SSQ are calculated according to the original procedure of Kennedy et al. [1993]. Participants filled the SSQ before and after the session. The final score is the difference of both questionnaires to measure relative changes.

Objective Measurement

Initially, the SSQ was designed with classical simulators in mind and could therefore raise the question of its validity for 360° videos. As a direct contrast to the self-reported SSQ, we also collect physiological data to objectively assess the level of experienced sickness.

Physiological measurements have been previously studied to identify CS in virtual environments [Tauscher et al. 2020, Gianaros et al. 2003, Dennison et al. 2016, Doweck et al. 1997, Martin et al. 2018, Kim et al. 2005]. Multiple physiological symptoms appeared to be correlated with CS and its associated symptoms.

One indicator is the eye blink rate of a person as shown by Dennison et al. [2016]. They state that the blink rate increases when participants start to feel sick. In this paper, we validate the use of blink rates for CS prediction for 360° videos. We further hypothesize eye gaze behaviour to be also expressed by angular saccade velocities. Gaze behaviour should not be impacted between sessions when the applied visual techniques are truly unobtrusive [Salvucci and Goldberg 2000].

Furthermore, we measured the heart rate (HR) and electrodermal activity (EDA) of our participants as there physiological correlates were previously shown to be strong indicators of CS [Tauscher et al. 2020, Dennison et al. 2016, Kim et al. 2005].

E.3.3 General Methods

In our experiment, we explored real-world video scenes with 3 DoF. We used a within-subjects design with three sessions per participant. Two of the sessions involved the visual techniques – peripheral blur and opaque occlusion – and the third was performed as control condition without visual post-processing. The sessions were counterbalanced and conducted with a two-day-break per participant.

Task

Before the main task, people stayed in the static virtual scene for one minute to obtain physiological baseline data to later control for individual differences and let participants get used to the VR environment.

The duration of the main task was 10 minutes for all sessions. The participants could voluntarily stop and leave the experiment in case of severe sickness symptoms. Participants watched several 360° videos of parkour running and sport climbing with free falling scenes as described in Section E.3.1. The participants were instructed to remain still while watching the videos but were allowed to move their head to freely explore the scene. No input of the user other than the head rotation was required.

Apparatus

We used a commodity HTC Vive Pro Eye HMD with a FOV of 110° and a frame rate of 90 Hz. This HMD has a built-in eye tracking system that records gaze data with 120 Hz at an accuracy of 0.5°– 1.1° [HTC Vive Pro].

The peripheral-physiological data was collected using a MindMedia Nexus-10 MK2 device with sensors placed on the left hand of the participants (cf. Fig. E.3).

The 360° videos were recorded using an Insta360 Pro camera [Insta360 Pro] with $6k$ resolution at 30 FPS. Gyroscopic data was recorded every 3ms. The 360° camera was mounted on a snowboard helmet to record the scenes since we required fast and flexible movements with hands-free control. In order to control the weight and increase wearing comfort, we added extra padding and stabilizers. The final apparatus can be seen in Figure E.4.

Participants

A total of 23 participants were recruited for the experiment. 4 participants had to leave the experiment after the first session because of severe sickness symptoms and 19 subjects completed all three sessions (11 females, age range 20–32, mean age 25.68, SD

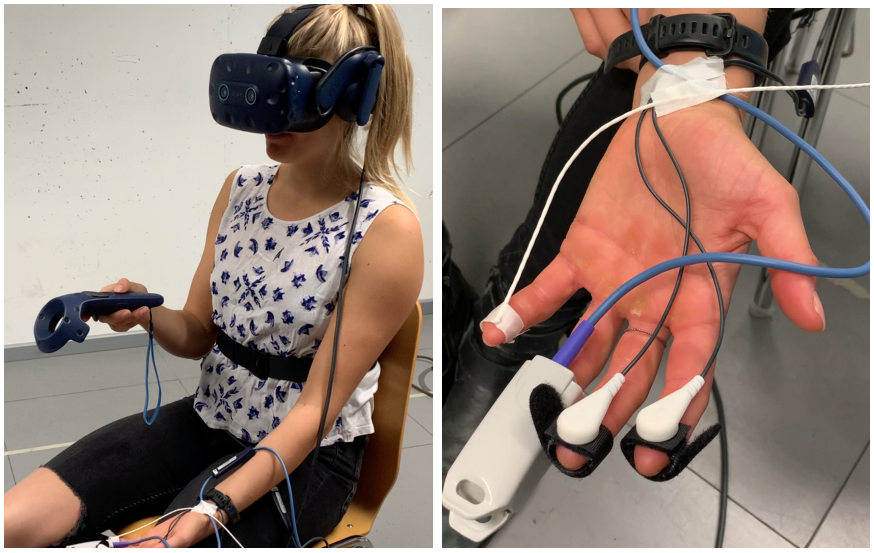

Figure E.3: Participants always sat down during the experiment (left image). The physiological data was measured with a finger clip on the left hand (right image).

4.65). All participants were compensated with a payment of 25 euro. The experiment was approved by the corresponding ethics committee.

Procedure

Each of the three sessions of the experiment was conducted with a two-day-break and at least 48 hours between sessions to avoid any carry over effects, following the design of established studies [Lin et al. 2020b, Ebrahimi et al. 2014, 2015]. Prior to the experiment, participants filled an informed consent form and a demographic questionnaire. The SSQ was filled before and after each session to capture the relative change in well-being. Before the main task was performed (cf. Section E.3.3) the eye tracker was calibrated. During the entire experiment participants sat on a chair and were not allowed to stand up to not break the immersion, minimize the risk of accidents and not impair the physiological measurements.

E.4 Results

For the analysis of the experimental results we used factorial mixed repeated-measures ANOVAs with condition as within-subject and gender as between-subjects factor followed by pair-wise two-sided dependent t-tests for repeated measures with Bonferroni-correction.

Figure E.4: Helmet carrying the 360° camera to capture the recordings for the experiment.

According to previous research, gender has a strong influence on the susceptibility for CS [Narciso et al. 2019]. Therefore, we additionally analyze men and women separately to investigate the influence of gender to mitigate CS.

Figure E.5 illustrates the results of the SSQs and the average times for participants to end a session. The SSQ was analyzed for its total score and the three subscores of nausea, disorientation and oculomotor [Kennedy et al. 1993]. Based on the subjective self-assessments of the participants the results show the same trend for almost all analyses: the unmodified ground truth (GT) condition was perceived as the most sickness inducing while both modification techniques (blur and opaque occlusion) reduced the sickness score. For most observations, the average SSQ level for the session with dynamically occluded FOV (dark blue) was reported the least sickening (cf. Fig. E.5).

The factorial mixed ANOVA shows a significant main effect on the SSQ total score (Fig. E.5a) for condition ($F(2, 34) = 4.261$, $p = 0.022$, $\eta^2 = 0.2$). No significant effect was found for gender, indicating that the methods work equally for both genders. Pair-wise dependent t-tests show a significant difference for females for the SSQ total score when peripheral blurring is applied ($T = 2.84$, $p = 0.0174$) as compared to the control condition (GT). For males there is no such difference. Also for the nausea subscale of the SSQ (Fig. E.5b) a significant main effect for condition is present ($F(2, 34) = 4.121$, $p = 0.025$, $\eta^2 = 0.2$). From the pair-wise t-tests for the females a significant effect is present for both blur ($F = 2.43$, $p = 0.0352$) and opaque occlusion ($F = 2.97$, $p = 0.0141$) when compared to the unmodified view. There is no significant

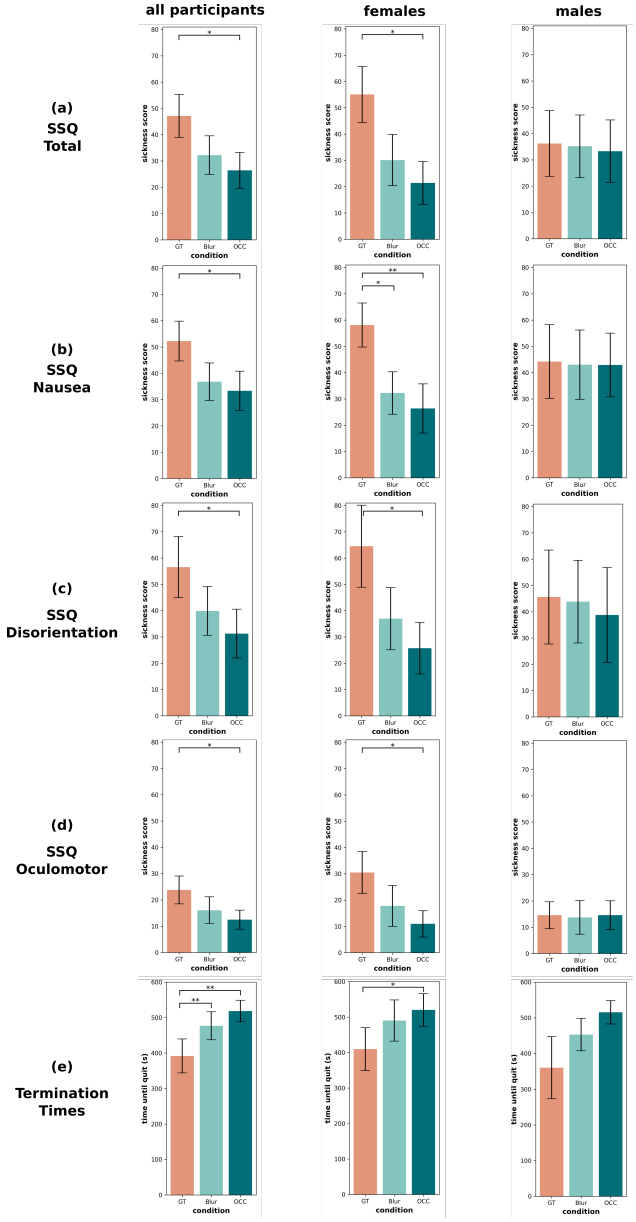

Figure E.5: Averaged SSQ scores and termination times; error bars represent the standard error of the mean (SEM). **(a)** SSQ results for total score ($N_{all} = 19$, $N_f = 11$, $N_m = 8$). **(b-d)** SSQ results for each of the subscales. **(e)** Termination times for each session only considering participants that quit at least once prematurely. ($N_{all} = 11$, $N_f = 7$, $N_m = 4$). Significant results are denoted by '**' (**p≤0.016**, Bonferroni-corrected for multiple comparisons) and '*' (**p≤0.05**)

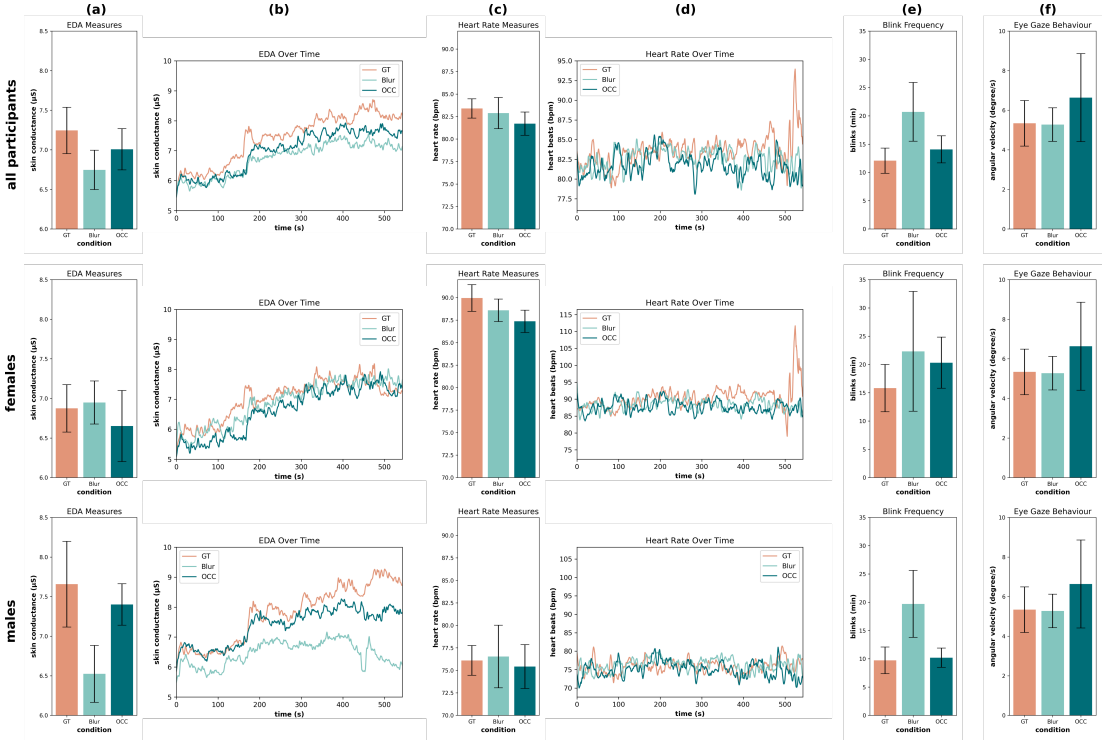

Figure E.6: Physiological measurements ($N_{all} = 19$, $N_f = 11$, $N_m = 8$), error bars represent the SEM. **(a,c)** Baseline corrected EDA and HR averages. **(b,d)** The corresponding EDA and HR over time. Note, that the baseline-corrected values are raised by the constant of average heart rate of all sessions to make them more human readable. **(e,f)** Eye blink rates and eye gaze behaviour.

effect for males. For the disorientation SSQ subscore (Fig. E.5c) of all participants a significant effect is present for condition ($F(2, 34) = 3.282$, $p = 0.05$, $\eta^2 = 0.16$) but not for gender. The pair-wise t-tests show a difference between GT and opaque occlusion for females ($T = 2.27$, $p = 0.0466$). For males this effect is absent. The oculomotor effects (Fig. E.5d) only demonstrate a minor effect for condition when all participants are considered ($F(2, 34) = 2.924$, $p = 0.067$, $\eta^2 = 0.15$). However, for the women the pair-wise t-tests shows a significance for GT compared to a opaque occluded view ($T = 2.37, p = 0.0391$). For the duration participants spent in the virtual scene before they chose to quit the experiment (Fig. E.5e), we can see significant main effects for condition ($F(2, 18) = 11.04$, $p = 0.00074$, $\eta^2 = 0.55$) and gender ($F(1, 9) = 10.04$, $p < 0.011$, $\eta^2 = 0.53$). The pair-wise t-tests confirm this outcome and also reveal a significance for both mitigation techniques (Blur: $T = -3.05$, $p = 0.0123$; OCC: $T = -3.83$, $p = 0.0033$) for all participants. For only the female participants the ef-

fect is present for opaque occlusion against the unmodified GT condition ($T = -2.75$, $p = 0.0332$). Although the average session time also increased for men, the t-test did not reveal any significance, which may be due to the low number of male participants considered here ($N = 4$).

Overall 11 participants decided to terminate one or more of the sessions early because of severe sickness symptoms (57.9%). When the terminations are separated by condition, the participants felt the strongest urge to quit during the unmodified condition (GT = 47.3%, Blur = 36.8%, OCC = 31.5%), giving a first hint that the control condition (GT) was least pleasant to the participants. As expected, most experiment terminations occurred in the first session (S1 = 52.6%, S2 = 36.8%, S3 = 26.3%). Therefore, it is possible that familiarity with the scene increased after the first session and impacts the termination behaviour. The randomized order of the FOV modifications was used to prevent this effect.

Figure E.6 shows the analysis results of the physiological measures. The average EDA and HR measurements over the time-course of the experiment are shown in Figure E.6b/d. Additionally, for the eye tracking data, the blink rate per condition and the gaze behaviour by angular saccade velocities is plotted.

EDA and HR were baseline corrected per participant and session (Section E.3.3). Although averaged physiological results do not indicate a clear trend for every condition, more pronounced changes are visible in the temporal evolution of the signal and indicate higher measures for GT. Note, that the results show a general increase of EDA for all conditions. This drift is not unusual for VR as the devices emit heat.

The blink frequency was highest when peripheral blur of the FOV was applied. For GT, the blink rates were the lowest. For the eye gaze behaviour we found no significant change.

E.5 Discussion

Our experiment investigates techniques that reduce the user's optical flow for 360° videos in VR. Modifying the FOV indicates a reduction of the CS level and, at the same time, increases the time participants are willing to spend in the VR scene, suggesting a more pleasant experience overall. Opaque occlusion of the peripheral regions with a dynamic radius was most effective to improve the SSQ score and the time spent in the VR environment. Blurring the peripheral area was also effective especially for the time until participants ended the experiment. The results suggest that the extent to which optical flow is reduced is related to the efficiency of the methods. Since the manipulation of the

FOV with an opaque occlusion reduces optical flow the most, this technique indicated the highest efficiency in our experiment.

We analysed the single subscores of the SSQ for a more in-depth understanding of the impact of the visual techniques. All subscales show the same trend as the total SSQ results with the highest sickness score for the GT condition and the lowest score for the opaque occlusion. In direct comparison, the oculomotor subscale exhibited the least effects. This low level of oculomotor effects may be explained by the fact that impairments like eye strain and blurred vision usually occur after prolonged VR sessions due to exposition to the display hardware. Here, the sessions had a maximum duration of 10 minutes. The subscales for disorientation and nausea are comparable, both with high values. Disorientation is a problem known to occur after VR sessions. In the case of nausea, on the other hand, some claim that it does not occur as severely in VR experiences [Kim et al. 2018b]. The results of our experiment clearly state that nausea should be considered if 360° videos are experienced in VR.

The results of the physiological measurements show higher values for the GT condition for most of the time, especially for skin conductance (EDA) that varies with the activation of sweat glands of the hands. EDA can be considered as one of the most common observation channels of sympathetic nervous system activity, and manifests itself as a change in electrical properties of the skin. Other studies showed that the nervous system activity is related to CS [Gavgani et al. 2017, Tauscher et al. 2020]. EDA as well as HR measures show more pronounced differences between conditions towards the end of session, indicating an increase of CS over the time-course of the experiment. Interestingly, the strong upward trend of EDA of male participants towards the end of the session for the GT condition suggests that severe sickness is experienced late in the session (Fig. E.6b bottom).

For the results of the eye tracking data, no influence of the FOV modifications on the participants' saccadic length (eye gaze behaviour) is found (Fig. E.6f), which confirms the subtleness of the investigated techniques [Adhanom et al. 2020]. This indicates that the participants did not feel restricted by the peripheral FOV modifications and looked around freely to the same extent in all scenes. The blink frequency was highest for the peripheral blur condition (Fig. E.6e) and lowest for GT in contrast to previous work [Dennison et al. 2016].

As previously discussed, the validity of the SSQ for 360° videos could be questioned as it was neither designed for such scenes nor validated for them. A direct comparison of the physiological data and our SSQ measures reinforces the validity of the SSQ to measure CS for 360° videos in VR.

Results from previous research indicates that CS disproportionately affects women [Narciso et al. 2019]. In our experiment, the level of sickness participants experienced was likewise higher for women, at least for the unmodified GT scene. When the FOV modification comes into place, the level of CS is on the same level for both genders. Therefore, the investigated visual techniques seem to be more effective for female participants. This is the case for the total SSQ score as well as for all subscores. This is indeed an interesting finding as gender differences in the perception of VR have been pointed out in related research [Stanney et al. 2003, Munafo et al. 2017, Grassini and Laumann 2020, Narciso et al. 2019]. Concerns have been raised about possible inequitable barriers when consuming immersive media. Reducing CS to an equal level could again democratize VR technology for all genders. The average SSQ score for female participants decreased by half (Blur: 45.4%, OCC: 61.1%) while, simultaneously, the time they spent in the scene increased (Blur: 20%, OCC: 27%). While men stayed longer in the scene when the visual modulations are applied (Blur: 27%, OCC: 43%), their SSQ scores did not significantly change across conditions. This may suggest that some male participants went up to their limit of tolerable sickness that simply occurred later with a restricted FOV. Furthermore, we can observe that women had a higher heart rate than men overall, with more than 10 bpm difference on average.

The cumulative video duration for the experimental task was ten minutes in total, which is probably shorter than a normal VR session. We chose this rather short time because our 360° videos represent a very challenging scenario with a high probability to cause CS. This strong scenario gives us a better control over the experimental variables. To ensure comparability this increases the likelihood for people to become sick at some point during the exposure time. The capability of our videos to cause CS during the chosen exposure time is further demonstrated by the number of participants leaving the experiment early (57.9%).

E.6 Conclusion

In this work we investigate the mitigation of cybersickness in 360° videos displayed in VR. By modifying the peripheral regions of the FOV with either blurring or opaque occlusion, we show that cybersickness is efficiently reduced, allowing VR users to have a more pleasant experience. The visual modulations were designed to be unobtrusive to the users by following the eye movements and only restricting non-central parts of the view.

For passively watched 360° videos with strong movements, opaque occlusion of the

peripheral FOV is most recommended when both methods come into question.

Furthermore, our results indicate that the investigated techniques work especially well for women with a significant reduction of all sickness scores in our experiment. This is an important factor as VR technology raised the concern of gender bias and our investigated techniques may lead to a more general acceptance and stream-lined experience across all genders. Besides the total SSQ score, all subscales were also analyzed. We found that although all subscores were reduced with a modified FOV, 360° videos primarily caused nausea and disorientation. When videos that are shot along a moving camera trajectory are shown in a VR environment these factors should be considered.

The investigated methods were shown to be suitable for video scenarios with strong movements. In future work, we plan to focus on immersive videos with less strong movements and a longer exposure time. Furthermore, physiological measurements confirmed the SSQ as valid measurement for cybersickness in 360° videos. In the future we will therefore extend our work with more freedom for the user.

F Cybersickness Reduction via Gaze-Contingent Image Deformation

Virtual reality has ushered in a revolutionary era of immersive content perception. However, a persistent challenge in dynamic environments is the occurrence of cybersickness arising from a conflict between visual and vestibular cues. Prior techniques have demonstrated that limiting illusory self-motion, so-called vection, by blurring the peripheral part of images, introducing tunnel vision, or altering the camera path can effectively reduce the problem. Unfortunately, these methods often alter the user's experience with visible changes to the content. In this paper, we propose a new technique for reducing vection and combating cybersickness by subtly lowering the screen-space speed of objects in the user's peripheral vision. The method is motivated by our hypothesis that small modifications to the objects' velocity in the periphery and geometrical distortions in the peripheral vision can remain unnoticeable yet lead to reduced vection. This paper describes the experiments supporting this hypothesis and derives its limits. Furthermore, we present a method that exploits these findings by introducing subtle, screen-space geometrical distortions to animation frames to counteract the motion contributing to vection. We implement the method as a real-time post-processing step that can be integrated into existing rendering frameworks. The final validation of the technique and comparison to an alternative approach confirms its effectiveness in reducing cybersickness.

Colin Groth [1] and Marcus Magnor [1] and Steve Grogorick [1] and Martin Eisemann [1] and Piotr Didyk [2]

[1] Institute for Computer Graphics, TU Braunschweig, Germany

[2] The Perception, Display, and Fabrication Group, Università della Svizzera italiana, Switzerland

Journal article accepted to *ACM Transactions on Graphics (ACM TOG) 2024,* Volume 43, Number 4. To be presented at *ACM SIGGRAPH 2024*.
DOI: 10.1145/3658138

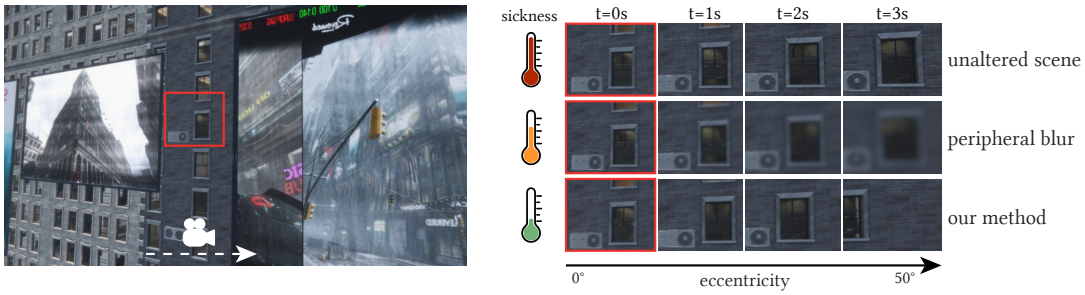

Figure F.1: In this paper, we present a new approach to mitigate cybersickness by applying subtle geometrical distortions to the animation frames. While the modifications significantly reduce cybersickness, the details in the periphery can be preserved better than with the commonly used peripheral blurring.

F.1 Introduction

Virtual reality (VR) rapidly evolves and captivates users with its highly immersing experiences in virtual worlds. Its potential is vast and varied, from the enhancement of entertainment and education to innovative applications in design and therapy. However, a significant barrier to its universal adoption lies in the onset of cybersickness, a negative symptom similar to general motion sickness. Cybersickness arises from a sensory mismatch between the vestibular system's signals and the visually perceived motion of the virtual experience [Reason and Brand 1975]. A crucial factor for cybersickness is vection, i.e., the sensation of self-motion induced by visual stimuli even when the body is physically motionless [D'Amour et al. 2021]. People often experience such an illusion when sitting on a stationary train and watching another train moving. The visual movement observed through the window leads to a feeling of self-motion. In such cases, human perception attributes a higher significance to the information-dense visual stimuli conflicting with the vestibular signals [Bankieris et al. 2017, ter Horst et al. 2015], often resulting in a negative sensation of sickness or discomfort.

In VR, virtual camera movements that are not accompanied by physical motion are a significant source of vection, leading to cybersickness. One of the solutions to reduce cybersickness is to limit the visual vection cues. Common approaches modify the visual content by occluding or blurring the peripheral vision or modifying the camera path [Adhanom et al. 2020, Groth et al. 2021b, Hu et al. 2019]. These approaches aim to reduce visible movements, especially in the periphery, which is considered a more motion-sensitive part of the visual field [Exner 1886, Finlay 1982]. The methods, however, lead to changes in the experience or loss of visual features in the periphery. Other

techniques focus on directly controlling the vestibular system. Their active stimulation can reduce the sensory mismatch through alignment with the visual motion [Groth et al. 2022, Sra et al. 2019]. However, such an enhancement requires specialized equipment that is not yet universally accessible.

This paper proposes a novel technique to mitigate cybersickness that is both subtle and effective. We start by studying linear and rotational camera motion to later address arbitrary camera paths. For linear motion we aim to explicitly reduce the magnitude of perceived motion by slowing down objects in the periphery. To maintain the user's perception of the camera velocity, we demonstrate how the reduction of velocity magnitude in the periphery can be compensated in the foveal region without reintroducing cybersickness. The cybersickness caused by visual rotations is even more severe compared to linear camera movements [Kim et al. 2021, Groth et al. 2022]. To address this type of movement, we propose to create an illusion of objects moving along a linear trajectory instead of rotating around a particular axis. We show that simple geometric distortions that modify the screen-space size of the objects are a powerful tool to achieve this goal. By a series of perceptual experiments, we study the effectiveness and visibility of the proposed manipulations (Section F.3). Based on the results, we present a perceptual model that describes the manipulations that lead to the maximum reduction of vection that can be performed without objectionable changes to the content (Section F.4). The model provides scene and eccentricity-dependent parameters for the final method. While our initial experiments make use of complete control over 3D scenes, such control is not always feasible in complex scenes. Therefore, we propose a method for applying our manipulations using a simple image-based warping method, which modifies the geometrical information of the scene by introducing subtle image deformations (Section F.5). Since the deformations accumulate over time, we exploit saccadic suppression and eye blinks to restore the frame content to the original rendering. The efficiency of the method enables real-time execution and easy integration to any rendering engine. We validate our method in an experiment comparing with subtle blurring of peripheral content (Section F.6). While the comparison was not perfectly analogous, the results demonstrate the effectiveness of our approach in reducing cybersickness while the visual fidelity of the scene is preserved, providing a promising direction for future research.

F.2 Related Work

In this section, we describe how vection leads to cybersickness and explore former research on mitigation of cybersickness in virtual reality. Techniques that reduce cybersickness can generally be separated into two categories: visual content manipulations and vestibular perception enhancement. In the following, we will focus on visual methods that do not require further equipment.

Vection and its Role for Cybersickness Unlike traditional displays such as monitors or TVs, VR displays deliver stereoscopic images that isolate the user from the real world, deepening their immersion in the virtual environment. When determining what is real, the human brain tends to prioritize visual cues [Bankieris et al. 2017, Murovec et al. 2021]. VR delivers strong visual signals that shift the users' feeling of presence to the virtual world. However, the disconnection between visual movements in the virtual environment and actual body motion is a key contributor to cybersickness, a form of motion sickness during VR exposure [Reason and Brand 1975]. When VR users navigate using controllers and continuous movements, they experience motion in the virtual world that does not coincide with their real-world physical state. This discrepancy in perception is evoked by vection – the sensation of self-motion created by visual cues [D'Amour et al. 2021]. In real life, a momentary illusion of self-motion can be evoked, for instance, while sitting in a stationary car and observing another car moving, exemplifying the dominant of our perception in contradictory scenarios. Vection primarily stems from the peripheral vision which is highly sensitive to temporal changes [Thompson et al. 2007, Guo et al. 2021]. Reducing motion cues in the peripheral vision can effectively diminish cybersickness, due to the reduction of vection [Luu et al. 2021]. However, even when suppressing the subconscious effect of vection, humans are still able to experience the movements of the virtual camera, due to the conscious perception of the apparent motion in the fovea. This distinction between apparent motion and vection is a key factor in understanding and properly addressing cybersickness.

Visual Techniques for Cybersickness Reduction Various techniques have been explored that modify the visual stimuli for cybersickness mitigation. Opaque occlusions in the visual field, either in the center region [Bos et al. 2010, Lin et al. 2002, Seay et al. 2001] or based on eye tracking [Adhanom et al. 2020, Groth et al. 2021b,a], are notable for reducing vection and cybersickness. A more common and less intrusive technique includes blurring of the periphery [Groth et al. 2021b, Patney et al. 2016, Hillaire

et al. 2008], fixed outer regions [Lin et al. 2020b] or object-dependent areas [Nie et al. 2017, 2019] using gaussian filters. Also learning approaches were proposed for motion reduction [Kaplanyan et al. 2019]. These approaches, acting as a low-pass filter, reduce contrast and information perception, thereby diminishing cybersickness. However, the information reduction in the periphery can be an issue in applications that require fast reactions like VR shooter games. With our method, we preserve the visual details of the scene over the entire field of view (FOV) by subtly reducing vection with content-aware distortions. Recent methods include integrating reverse optical flow visualizations [Park et al. 2022, Kim and Kim 2022]. The idea is based on the larger pooling of the ganglion cells in the peripheral vision [Anderson et al. 1991]. Park et al. 's experiment with reverse optical flow arrows reduced cybersickness but significantly affected participants' experience due to its application in both peripheral and foveal regions [2022]. We calibrate our modulations to stay under the threshold of detectability to keep the virtual experience immersive and enjoyable. The relationship between geometry appearance, motion perception, and cybersickness remains underexplored. However, temporal geometrical modifications can change the basic visual motion perception of objects and are highly interesting for reducing vection. In a simple prototype of Lou et al. [2022], the geometry of a building is squeezed towards the viewport's edge during forward movements. While their project was not generally applied or experimentally validated the authors' approach motivated the use of geometrical deformations for vection reduction. Aside from these techniques, the overall quality of the virtual experience as well as the content design play a crucial role. In general, high frame-rate renderings, high quality tracking and reduced latency systems reduce the occurrence of cybersickness symptoms [LaViola Jr 2000, DiZio and Lackner 1997, Sherman 2002].

F.3 Perception of Fundamental Movements

Vection is influenced by visual movements in the FOV. The peripheral area has a much stronger influence on this perception of self-motion than the fovea [Exner 1886, Finlay 1982]. At the same time, modifications in the peripheral region are generally tolerated more by users [Patney et al. 2016]. The following investigations are based on the hypothesis that manipulations can be made to a virtual scene that go unnoticed but reduce vection. As a consequence of this hypothesis, we further want to show that vection — a subconscious effect of perceived self-motion — is decoupled from apparent motion — the conscious estimation of visual movement. This section, describes the psychophysical experiments that we conducted to confirm this hypothesis and derive the detection

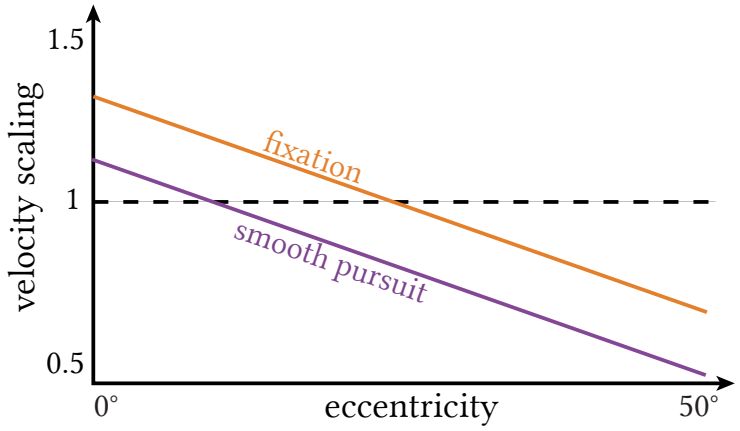

Figure F.2: Visualization of the velocity scaling of an object based on its eccentricity and the type of eye movement.

thresholds that allow for maximizing the reduction of vection. Our investigations are twofold and examine linear movements and rotational movements in isolation. The results of these psychophysical experiments provide the foundation for the perceptual model derived in the next section. All experiments are conducted in VR with a HTC Vive Pro Eye headset.

F.3.1 Vection Compensation for Linear Movements

Our first investigations are focused on the reduction of vection for linear camera movements. In two psychophysical experiments with a simple scene, we directly manipulate the velocity of objects in the VR environment. All modifications are designed to preserve the visually perceived ego-motion in the virtual world. We make two assumptions for our investigations: (1) the human brain derives the apparent motion of a virtual scene as an average of all background motion in the FOV. (2) the velocity of each background object in the visual field can be scaled and the scaling factor can be described by a linear function of eccentricity.

The assumptions are motivated by basic properties of the human visual system (HVS). In the HVS, the effect of information integration of local patterns allows to reach a global consensus even with ambiguities information of local motion (assumption 1). This phenomenon of natural viewing is necessary for humans to characterize ambiguous motion information in local regions that would be raised due to restrictions similar to the aperture problem [Thompson et al. 2011]. The increase in effect strength is motivated with the pooling of peripheral photoreceptor information by the bipolar cells

Figure F.3: Visual scene of the experiment with linear camera movement. The objects in the left image are scaled to move slower in the periphery. The red dot visualizes the view point. The right image shows the unaltered output.

and cortical magnification (assumption 2) [Thompson et al. 2011].

Figure F.2 visualizes how the object motion scaling is described by a linear function of eccentricity. The further away an object from the point of view, the more its velocity is reduced. In contrast, objects in the fovea are accelerated. The function has two unknown properties, the slope and the offset, which are calibrated in two separate experiments.

In the first experiment we search for the slope of the linear function that describes the velocity scaling of objects at the threshold of detectability. This threshold describes the intensity of decreased peripheral speed given a constant foveal velocity. However, since we simplified the scene motion to be a weighted average of the objects at different eccentricities, the overall perceived speed of the camera can be altered by the scaling. Such a perceived change in speed would reshape our understanding of the scene with potential implications on the user experience and performance. Therefore, we conduct a second experiment that calibrates the offset of the linear function for the calibrated function slopes (cf. Figure F.2).

Experiment 1: Slope Calibration

The maximum difference in velocity between the fovea and the outermost point in the periphery is defined by the magnitude of the slope of the scaling function. Here, we use the description of object velocity instead of camera velocity based on the formally made assumption that the linear camera motion derives by averaging the velocity of all background objects in the visual field.

Table F.1: Perceptual thresholds for velocity scaling intensity at two tested camera velocities.

	Velocity 1.0		**Velocity 1.4**	
Task	**Mean**	**± StDev**	**Mean**	**± StDev**
Up	0.3344	± 0.1685	0.28	± 0.1582
Down	0.4035	± 0.2657	0.3631	± 0.2721

Experimental Design: The experiment is conducted with a simple scene that displays an infinite straight wall with multiple lines of windows (see Figure F.3). The camera faces the wall while moving to the right. The horizontal spacing between the windows is randomly varied to avoid that the results are influenced by regular patterns. A random color is assigned to each window to increase contrast and recognizability. A small red ball in the forward direction defines the gaze point and participants are ask to always look at this ball. In the experiment the gaze of participants was tracked and the scene was blended out when the participants' gaze deviates more than 5° from the ball's direction. The experiment is designed as an Up/Down task with the slope of the linear curve being manipulated. The linear curve defines how the speed of a window is adjusted based on eccentricity. Thereby, the objects are increased in speed in the foveal area at the same amount as they are slowed down in the periphery. For clarification, a slope value of 0.3 describes a scene where the windows are 30% increased in speed at the focal point and 30% slowed down in speed at the outermost point in the periphery. For the up trials the slope start value is 0 which gets continuously increased by 0.05 per second. The down trials start with a 0.95 slope value which is decreased by 0.05 per second. For every trial participants are ask to press a button when they *notice a difference in speed or distance of the windows between the inner and outer area of the FOV*. With the button press the result is logged and the next trial begins. Each trial starts with a random idle period between 2 and 5 seconds to avoid that participants recognize regular patterns. We investigated the slope for two different speeds, 1 m/s and 1.4 m/s, which are counterbalanced. The higher speed, thereby, corresponds to half of the maximum speed for smooth movement perception of the eye [Daly 2001], allowing us to stay within this limit even with a maximum scaling of 200%. In our within-subject experiment we had 5 repetitions for each, the Up and Down tasks, and both speeds, yielding a total of 20 trials per participant. A total of 11 participants took part in the experiment (2 females, age = 25.5 ± 4.31).

Results: Table F.1 shows the results of the experiment. In general, the results of the psychophysical experiment support the initial assumptions. Even in the worst case scenario, the speed of objects is reduced by 28% at an eccentricity of 50°. Stronger modifications go unnoticed for slower camera movements. To ensure subtlety, we interpret the results in a conservative manner and apply the lower values of the Up task in the following.

Experiment 2: Speed Adjustment

The offset property of the scaling function derives the default scaling factor at zero eccentricity. Therefore, this parameter weights the influence of objects in the inner and outer half of the FOV on the overall perceived scene motion. Consider the two curves in Figure F.2 that describe the same perceptual camera velocity: since people rely more on the foveal content during smooth pursuit eye movements, the objects' speed in the periphery can be reduced more.

Experimental Design: The scene of the second experiment remains unchanged, with the camera moving along a wall with multiple lines of windows. While we consider a constant slope based on the results of Experiment 1, the overall speed of the camera, i.e. the function offset, can be adjusted by the participants. We show two versions of the scene to the participants, which can be switch as often as required. The first version provides the unaltered baseline with slope 0 as a reference. The second version is constructed similarly, but participants are able to change the speed of the camera. We ask participants to match the speed of the adjustable scene to the baseline. They are allowed to take as much time and scene switches as needed. When the scenes are toggled, a gray screen of 1.3 seconds is displayed to avoid direct comparison from the change and rather rely on the participants' memory. Similar to the first experiment we display a focus point and occlude the scene when participants deviate from it. However, additionally to the fixed focus point we also investigate the perception for moving gaze. In this moving gaze scenario the focus point is moved at the speed of the windows and displaced to the right when it gets close to the edge of the screen. With this scenario, we induce smooth pursuit eye movements, while the static focal point investigates eye fixations. These two types of fundamental eye movements are highly relevant for real-world scenarios, but can significantly diverge in the perception of movement speeds. As in the former experiment, we investigate both scenarios for the baseline camera velocities of 1 m/s (slope: 0.334) and 1.4 m/s (slope: 0.28). To verify the validity of the results, we further introduce a control condition where the adjustable version

Table F.2: Summary of the results for the second experiment assessing participants' perception of movement speeds during fixations and smooth pursuit eye movements. The table presents the matched speeds for two different camera velocities (1 m/s and 1.4 m/s) across two conditions: with slope (as per the findings of Experiment 1) and control (uniformly moving windows). Note: when the values get lower, the perceived scene speed is higher than the actual camera velocity and, consequently, scene objects can be decelerated more.

Scenario	**Condition**	**Speed: Mean ± Std**
Fixation	With Slope	1.0: 0.9906 ± 0.0905
		1.4: 1.3750 ± 0.1132
	Control	1.0: 0.9922 ± 0.0663
		1.4: 1.4078 ± 0.1133
Smooth Pursuit	With Slope	1.0: 0.8047 ± 0.1018
		1.4: 1.2141 ± 0.1501
	Control	1.0: 1.0266 ± 0.0848
		1.4: 1.4219 ± 0.1075

of the scene is without manipulation and shows uniformly moving windows just like the baseline. The purpose of this control condition is to assess participants' general ability to match the speed of objects during fixations and smooth pursuit to a baseline. We conducted the within-subject experiment with three repetitions for both velocities, conditions and scenarios. The order of the trials within each scenario was randomly distributed and participants did not know about the different conditions. A total of 17 participants took part in the experiment (8 females, Age = 24.2 ± 3.68).

Results: Table F.2 illustrates the perceived velocities for a given baseline. The results for the two scenarios differ considerably. For fixations our assumption that the influence of the inner and the outer part of the vision contribute evenly to the perceived camera velocity holds true, letting the measured values match the baseline almost perfectly. For smooth pursuit eye movements the influence of the foveal region for the overall perceived camera velocity is significantly higher. The decelerated periphery contributes around three times less to the scene speed perception highlighting a significant shift compared with the fixation scenario (see Figure F.2). The results of the control condition confirm the validity of participants' speed assessment ability. Accordingly, the deviating results of different eye movement scenarios are not based on a change in accuracy of speed estimations but rather on the shift in the importance of different image areas for motion

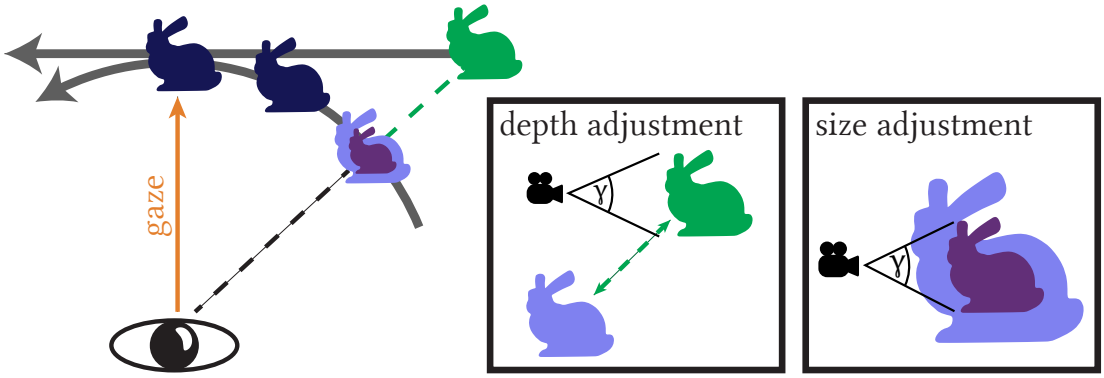

Figure F.4: Visualization of the depth adjustment and scaling method for vection reduction of rotating objects. Blue bunnies represent the unmanipulated state, while green and violet are the adjusted state.

perception during active eye movement. In conclusion, for both types of eye movement, the vection in the motion-sensitive periphery can be greatly reduced without changing the apparent motion of the virtual camera, which we later exploit to effectively mitigate cybersickness.

F.3.2 Compensation of Visual Rotation

Next, we take a closer look at rotational camera movements inside virtual environments. In a psychophysical experiment with a simple scene, we transform objects in the periphery to create the illusion of the objects moving along a linear trajectory. The manipulations are designed to preserve the conscious scene understanding while reducing cybersickness. The study is motivated by a key observation of former research that rotational movements in virtual environments induce considerably more cybersickness than linear camera movements [Groth et al. 2022, Kim et al. 2021, Groth et al. 2021b]. We make an assumption based on preliminary findings: the subconscious interpretation of rotational scene movements can be tricked to perceive linear motion, while the active scene understanding still results with the actual angular movement. The assumption, therefore, suggests a decoupling of the illusory self-movements (vection) and the apparent motion of the scene. If this assumption holds true, the applied scene adjustments will change the perceived vection and have a positive effect on cybersickness even when they are below the threshold of detectability.

Experiment 3: Rotation Reduction

Humans derive visual rotations of a static scene by the change of object sizes over time as well as from depth clues e.g. disparity [Leigh and Zee 2015]. This experiment investigates how much both factors contribute to the visual perception of angular movement.

Scaling methods: Figure F.4 illustrates how the scene objects are transformed in our investigations to create the illusion of movement on a linear path. For this visual linear motion, we scale the size of the 3D objects by a function of eccentricity. While the position of the modified objects in the 3D space remains on the circular path, their scaled size in relation to the other objects provide the visual cues of a position further away (see Figure F.2).

The distance d of the object on the virtual linear path, orthogonal to the cameras' forward vector f can be derived by $d = \frac{r}{\cos\phi}$ with r representing the radius and ϕ the angle between the forward vector and the vector from the camera to the object. Consequently, the adjusted size $s = \frac{w}{d}$ for an object is calculated using distance d and the objects width w. However, visual size is not the only cue to provide information about the 3D position. In our experiment, we introduce another condition in which the actual position of the scene objects is modified (depth condition). Therefore, in this condition, both the object's size on the screen and its disparity are altered. For the depth condition, the depth of the objects is modulated by displacing the position of the objects in accordance with d in the direction of the vector from the camera to the object. The depth condition is meant as a baseline and provides further insights about the importance of different cues for the perception of visual movements.

Experimental design: The scene in the experiment displays a circular wall with lines of windows and the camera rotating in the center (see Figure F.5). The angular velocity of the camera is modulated by a sine function. We investigate two different approaches, firstly the scaling of the objects in size and secondly the adjustment of the objects' positions in depth. In the experiment we want to find the intensity of the methods which describe the threshold of detectability. Intensity reflects the amount to which the methods manipulate the perceptual trajectory of the objects in the visual field at maximum angular velocity. At an intensity of 1 the methods manipulate the perceived object trajectory into a straight line and for 0 they are on the circular path. We use a 1 up/1 down procedure as psychophysical estimation method to find the conservative detection threshold (CDT) [Zenner et al. 2021]. In this experimental method the intensity of the manipulation technique is increased (+0.1) when participants do not

Figure F.5: Visual scene of the experiment with rotating camera. In the upper image the objects are scaled in size to create the visual impression of a linear movement trajectory. The lower image is the unaltered output.

notice any distortions in a trial and decreased (-0.1) in the opposite case. Furthermore, we apply a staircase design with interleaved ascending (start with intensity = 0) and descending (start with intensity = 1) sequences. The sequences are terminated after five reversals and the average of the last four reversals yield the sequence threshold estimate [Zenner et al. 2021]. In the experiment, the participants are ask to indicate by a button press when noticing any changes of the scene compared to the unaltered reference. The reference movement (normal circular path) is given in the first trial of each block which is properly communicated with the participants. We decided to show the reference condition only at the beginning of a block because we are interested in the participants' comparison with conscious expectations rather than a side-by-side comparison that does not compare to the final application. Each sequence of acceleration and deceleration of the camera to 0 is considered as one trial without pauses in-between. Also, each method is tested for yaw and pitch rotations. These 4 blocks (2 methods * 2 directions) are counterbalanced by a 4x4 Williams design Latin square. In the experiment, participants are ask to keep their head straight. A total of 15 participants took part in the experiment (7 females, Age = 28.8 ± 2.66).

Table F.3: Results of the calibration experiment for rotational movement compensation. The values indicate the CDT of the visual manipulation methods applied to the scene for different rotations. A value of 1 would correspond to a perceptually linear movement of the objects (full manipulation), while at 0 the initial circular path is displayed.

Condition	**Yaw**	**Pitch**
Scale	0.624 ± 0.2052	0.674 ± 0.2314
Depth	0.716 ± 0.2126	0.782 ± 0.184

Results: Table F.3 shows the CDTs for both methods and rotation directions. The results of the third experiment are in line with our assumption that subtle vection reduction is feasible. Substantial modulations can be made to the visual trajectory of the scene objects before participants report any difference. Overall, we were able to compensate for 62% of the rotational movement in the horizontal FOV of the VR glasses. Pitch rotations, that are due to hardware limitations displayed with smaller vertical angles can be compensated more. In the control condition, where the 3D objects are displaced in depth, compensations are around 10% higher compared to the modification in size. Overall, the results suggest that the majority of the movement information is derived by the contents' temporal change in size rather than from the disparity which motivates an important part of our method for cybersickness reduction to rely on transformations in geometrical size. As a limitation, the standard deviation is relatively high, which could reflect uncertainty among the participants. On the other hand, the thresholds are conservative and manipulations can be expected be less detected in practice when users do not pay attention to possible deformations. Also, more complex scene could mask the deformations to a certain degree.

F.4 Perceptual Model

In this section, we derive a perceptual model that estimates a scaling factor for each objects' optical flow between frames. The model is designed to maximize the reduction of vection while maintaining the observers' perception of the scene. For linear and angular camera movements, we provide two separate components F_{lin} and F_{rot} which both are functions of eccentricity. For the definition of F_{lin}, we assume a linear function that defines the scaling factor of the objects' motion based on eccentricity. To keep the perceptual scene speed intact, we compensate the reduced speed in the periphery with faster movements around the gaze point (cf. Section F.3). For angular movements the

function F_{rot} computes a scaling factor that modulates the objects' geometrical size to be smaller in the periphery. In the image domain, the modulations of rotational movements with F_{rot} give the impression of the scene content moving on a linear trajectory that is anchored at the gaze point. Our model is based on the observations of the psychophysical experiments of Section F.3. The supplementary material provides an additional list of all parameters defined here.

Linear Movement Component First, we define the linear movement component F_{lin} of our perceptual model. The results of the psychophysical experiments for linear movement compensation confirmed our initial assumption that objects can be scaled by a linear function of eccentricity (see Section F.3.1). In these experiments, the function was calibrated to the threshold that allows for the strongest modulations that stay undetected. Based on these findings, the linear component F_{lin} is defined as a linear function with slope a and offset b. The function derives a scaling factor of the object motion per input point with the current camera speed v_{cam}, the gaze velocity v_{eye} and a set of geometry points $p \in P$ as an input.

$$F_{lin}(P, v_{cam}, v_{eye}) = 1 - (a(v_{cam}) \cdot \min(\theta(p) \cdot n_1, 1) + b(v_{eye})) \tag{F.1}$$

Here, the eccentricity function θ is normalized by $n_1 = \frac{1}{\pi} \cdot \frac{9}{5}$ to be within $[0, 1]$. In the calibration experiments we placed the gaze point in the middle of the screen and, therefore, normalize with half the FOV (50°). In the experiments, we found that for higher movement speeds the scaling slope has to be slightly less intense (see Section F.3.1). We assume a linear dependency of the velocity for the change of the slope function and interpolate the slope for velocities that are between the calibrated ones. For fast movements above 1.4 m/s the slope is extrapolated following this trend making it more flat. This means that an increase in scene speed decreases the intensity of the modifications. Velocities below 1 m/s are handled conservatively and a constant slope of $s_{max} = 0.33$ (33% velocity reduction at the highest eccentricity) is assumed to avoid that the modifications become overly intensive and are recognized by the user. Specifically, we derive the following slope function with the former calibration values:

$$a(v_{cam}) = \min(s_{max}, m \cdot v_{cam} + s_{max} - m) \tag{F.2}$$

The value m models the decrease of the slope with increasing speed, $m = (s_f - s_{max})/(v_2 - V_1) = -0.125$, defined by the calibrated slopes $s_{max} = 0.33$ and $s_f = 0.28$ for the speeds $v_1 = 1$ and $v_2 = 1.4$.

The function offset parameter b of F_{lin} defines the perceptual difference in velocity between the actual linear speed of the scene camera and the perceived movement speed.

The formulation of b is based on the experimental results for velocity adjustment for different eye movements, described in Section F.3.1. We found that the perceived velocity of the modified scene highly varies with the type of eye movement the user performs. For object following, the HVS relies intensively on the foveal region. For fixations, on the other hand, different parts of the visual field contribute more equally to the velocity perception. We define offset b by a fixation parameter b_{fix} and an additional term that adjusts the offset of the scaling function for smooth pursuit eye movements.

$$b(v_{eye}) = b_{fix} + o_{sp} \cdot \min(v_{eye} \cdot n_2, 1) \tag{F.3}$$

With $b_{fix} = -0.25 \cdot a$ the default offset for full fixations allows for slowing down the object velocity in the outer half of the FOV proportionally to the speed up in the inner circle. The constant $o_{sp} = 0.2$ is derived from the former calibration and defines the offset adjustment that is necessary to address the increased importance of the fovea for smooth pursuit eye movements. The formula considers a normalization of the gaze velocity v_{eye} with $n_2 = \frac{\delta t}{\pi} \cdot \frac{9}{3}$ for gaze movements below 30°/s, where smooth pursuit eye movements fully take place, based on former findings [Leigh and Zee 2015]. We interpolate gaze velocities below 30°/s, since the studies of Leigh et al. show that no hard threshold can be defined between fixations and smooth pursuit [2015]. Note that the exact velocity where a movement is categorized as smooth pursuit is debatable and other works also suggest lower values [Komogortsev and Karpov 2013]. By relying on the gaze velocity, we can implicitly model both fundamental types of eye movements in F_{lin}.

Rotational Movement Component Next, we define the rotational component of the perceptual model by the function F_{rot} using the discoveries of the psychophysical experiments in Section F.3.2. The function output is a scaling factor that moderates the objects' geometrical size to visually align to a linear movement path. For this illusion, the size decreases stronger with higher eccentricity in the direction of motion. We define:

$$F_{rot}(P, p_{eye}) = (1 - \cos(\min(|d_{eye} - d_p|, |d_p|) \cdot \frac{\text{FOV} \cdot \pi}{180})) \cdot \alpha \tag{F.4}$$

The scalars d_{eye} and d_p are the respective distances of the projected object point and gaze point onto the axis of movement. The scaling modifications by F_{rot} are anchored on the gaze point as well as the forward vector of the head to increase temporal stability, represented by $\min(|d_{eye} - d_p|, |d_p|)$. The free parameter α determines the strength of the effect and was calibrated in the psychophysical experiments to $\alpha = 0.624$. A value of 1

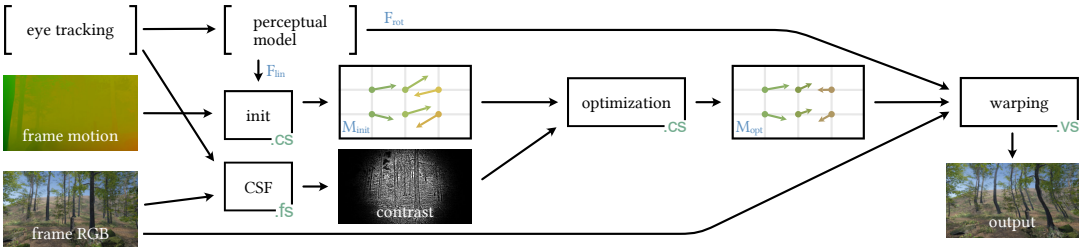

Figure F.6: With a series of shader operations (shader identifier in box corners), we apply screen-space geometrical modifications that reduce vection.

would correspond to a visually perfect linear trajectory. The scalars $d_{eye} := \langle p_{eye}, \hat{\mathbf{u}}_{cam} \rangle$ and $d_p := \langle p, \hat{\mathbf{u}}_{cam} \rangle$ are the distances of the gaze point p_{eye} and the geometrical point p to the origin after projection onto the axis of movement $\mathbf{u}_{cam}$. We assume the origin to be at the forward vector position in the image plane.

F.5 Application

In this section, we describe the implementation of our method for cybersickness mitigation based on our perceptual model. In the psychophysical experiments of Section F.3, we directly transformed the objects' 3D geometry in the scene. However, such a full control is not feasible in complex scenes. For general applicability, our implementation therefore operates as an effective and lightweight post-process on the rendering output of a VR application. Unlike before, the scaling modifications are applied using image warping in the 2D screen-space. Figure F.6 provides a general overview of the shader operations that are performed on the RGB image to generate the distorted output. Since our framework operates in the image domain of the rendering, it can be applied to arbitrary scenes without further adjustment. The implementation runs in Unity 2022.3 using renderer features and GPU-based shader operations in combination with OpenXR for VR support. The motion of the pixels between frames is derived by motion flow.

Warping In our implementation, we apply geometrical distortions to reduce vection. Without knowledge of the 3D scene, the modifications are applied to the rendered RGB frame. Since the topology of objects is unknown in screen-space, it cannot be adjusted directly. Therefore, the implementation relies on the optical flow to adjust the content movement. Rotational movements are compensated by foveated displacements that transform the content towards the axis of movement. This operation results in

a reduction of object sizes by distortions comparable to a concave lens laying over the image. For an efficient implementation that leverages the standard capabilities of graphics hardware, deformations are realized with a sparse vertex grid and UV warping in the vertex shader. The final image is rendered in the fragment shader by simple UV sampling.

The overall displacement D of the vertices $p \in P$ in the deformation grid are derived by combining the displacement vectors for rotations D_{rot} and the vectors of the linear movement modulation M_{opt}. The vector field of $M_{opt} = M_{raw}+F_{lin}$ combines the scenes' motion flow M_{raw} with the linear movement scaling F_{lin} of the perceptual model (see paragraph 'Optimization').

$$D(P) = D_{rot} + M_{opt} \cdot \max(1 - \frac{\omega_{cam}}{\omega_{max}}, 0) \tag{F.5}$$

The camera's angular velocity ω_{cam} is normalized with a maximum speed constant ω_{max} which can be adjusted per scene. In our experiments, we calibrated ω_{max} to 40°/s. The predominance of rotational movements makes the compensation of linear movements negligible during strong rotational movements [Groth et al. 2022, Kim et al. 2021]. Accordingly, the deformations by the linear component, described by M_{opt}, are scaled by the angular velocity of the camera ω_{cam}. The displacement vector field D_{rot} is derived by applying the scaling factor F_{rot} of the perceptual model to a vector orthogonal to the camera movement direction $\mathbf{u}_{cam}$.

$$D_{rot}(P, \omega_{cam}, \mathbf{u}_{cam}) = F_{rot} \cdot (\mathbf{u}_{cam} \cdot d_p - p) \cdot \frac{\omega_{cam}}{\omega_{max}} \cdot 2 \tag{F.6}$$

The length of the displacement D_{rot} depends on the eccentricity of the vertex point p. Here, $\mathbf{u}_{cam} \cdot d_v - p_v$ gives us the vector that points from p to the axis of movement.

Optimization The movement of the scene content between frames is derived by motion flow. To slow-down the velocity of peripheral content, the movement can be transformed to the inverse of the motion vectors. This counter-movement is modulated by the scaling parameter of the linear component of the perceptual model F_{lin} to derive the initial vector representation M_{init}. However, image overlaps can arise from opposing motion vectors. To avoid overlaps and find a globally optimal solution, we optimize M_{init} to find the final vector field M_{opt} which is used for the image warping. The goal is to determine a final state, ensuring each vertex closely aligns with its desired target position including considerations of the importance of different image areas. The optimization is required to maintain the non-penetration constraint.

Like for the warping, we opt for an efficient sparse grid (32x32p) for the optimization rather than pixel-level resolution. The optimization uses a semi-implicit Euler approach (step constant = 0.02). The forces acting on the vertices are the accumulated sum of the displacement forces to each of the four neighboring vertices. The rest length between vertices is the sum of the grid cell width (32p) and the temporally accumulated deformation of former frames. This dynamic definition of the spring rest length allows for more temporal stability than a static definition. In the optimization, we want to preserve content with high visibility, e.g. high contrast edges, while content under the threshold of sensibility can be deformed arbitrarily. The stiffness parameter of each spring between vertices in the grid is defined by the luminance contrast of the surrounding content given by the eccentricity-based contrast sensitivity function (CSF) [Tursun et al. 2019]. After the contrast sensitivity is masked by a transducer model [Zeng et al. 2000], the final visibility value is derived by summing up the individual frequency layers. Higher visibility corresponds to greater stiffness which preserves the content. The damping constant was chosen with $d = 0.9$ based on experimental validation. For each frame, we optimize the grid for 50 iterations on the GPU, taking approximately 3 ms.

For each vertex computation, the positions are evaluated in a local coordinate system, centered at the vertex's default grid position. This local approach simplifies displacement calculations, independent of global positioning and actual grid dimensions. Although the vertices do not have a global understanding, the global effects still have an influence as they spread over the multiple iterations of one time-step. Lastly, to preserve image content, the line of vertices along the image border remains fixed in position.

Recovery of the Original Rendering The geometrical distortions accumulate over time since the modulations operate as a temporal effect. During blinks and saccades, the phenomenon of change blindness occurs, which masks changes made to the visual content. In our implementation, we leverage this natural effect to recover to the original rendering. Human eye blinks occur around every 3 seconds and allow for an instantaneous reset of the entire image [Nakano et al. 2013]. The recovery with saccadic suppression, on the other hand, is more challenging, because a full reset would alter the eye's target position and disrupt the inherent expectation of edge consistency. Therefore, we only reset content c that based on the gaze position g is located against the direction of movement $\mathbf{m}$. It applies $\langle (c-g), \mathbf{m} \rangle < 0$. To avoid a sharp edge between the recovered and deformed content, the reset intensity is scaled with a gradient. Formally,

the reset intensity I is quantified by:

$$I(b, p_v, p_{eye}) = \|\langle \hat{\mathbf{m}}, p_v \rangle - \langle \hat{\mathbf{m}}, p_{eye} \rangle\| \tag{F.7}$$

Here, p_v represents the position of the respective vertex in UV coordinates, and p_{eye} denotes the gaze point of the eye.

Sun et al. found that scene displacements during saccades are undetectable for movements up to 12.6°/s even in the target region [Sun et al. 2018]. During saccades, we use this threshold to subtly reduce content deformations located in the direction of movement. In the case of insufficient resets, the distortions are kept below a maximum offset of 20° from the original content. The threshold is chosen based on our empirical analysis and ensures that the warping effect is both effective and visually pleasing.

For the detection of eye blinks, we rely on the capabilities of the SRanipal library and the HTC Vive Pro Eye HMD. The detection of saccades is based on the gaze velocity using the algorithm of Imaoka et al. [2020].

F.6 Validation

We validate the effectiveness of our implementation in a comprehensive experiment using a realistic VR scenario with multiple scenes. We run a within-subjects experiment with three sessions per participant to explore the effect of the proposed deformation technique. While the virtual environment was the same for all sessions, the visual post-processing was altered, comparing our method against blurring of the peripheral vision and an unaltered control condition. We counterbalanced the order in which the three conditions were shown to the participants. Furthermore, a recovery time of at least 48 hours was maintained between sessions to avoid carry-over effects.

F.6.1 Experiment Design

The experiment is approved by the university's ethics committee.

Stimuli

Conditions We investigate three different conditions: Our technique that does content-specific adjustments to the scene content while preserving visual fidelity, the commonly used method of peripheral blurring and a control condition as baseline. Peripheral burring is one of the most common techniques to subtly prevent occurrences of cybersickness. Based on former work, we implemented a foveated blurring where the quality

Figure F.7: In the experiment, we used two scenes, both displayed twice per session with a smooth transition.

degradation from the foveal point is scaled by the camera's linear and angular velocity [Groth et al. 2021b,a]. At maximum speed, the foveal region is kept the smallest with a diameter of 10° [Groth et al. 2021b]. The strength of the blur increases linearly with eccentricity. Following the results of Lin et al. [Lin et al. 2020a], the linear increase is set with a maximum kernel size of 13 which was found to be at a 50% detection probability. Comparably, the underlying principles of our warping are likewise calibrated to this commonly used detection threshold (see Section F.3) and adjusted based on the methodology described in the implementation section. We perform the manipulations in image space for a more dynamic response to scene content and broad applicability of the technique to a variety of applications.

Virtual Environment In the experiment we present a photorealistic virtual environment with two different scenes, urban and nature. While the urban scene mainly contains geometrical structures and straight lines, the nature scene explores the perception of a more heterogeneous shape composition. The camera path in the experiment is predefined and cannot be altered by the participants. While the path completes its circular trajectory, both environments are shown for the same amount of time, resulting in four scene changes per session in the same order. We carefully design the camera path to feature a balanced mixture of different camera movements. For linear movements, the camera constantly accelerates and decelerates in a sinusoidal manner to investigate movements that are most relevant for cybersickness [O'Hanlon et al. 1974, Groth et al. 2022]. Furthermore, we include sections with unpredictable movements with a linear and angular component and full 360° rotations.

Apparatus

For the experiment we used a HTC Vive Pro Eye head-mounted display (HMD) with a visible FOV of 100° and a frame rate of 90 Hz. The resolution of that HMD is 2880 x 1600 pixel. The rendering is performed on a commodity computer with a NVIDIA RTX 4090 graphics card.

Participants

A total of 25 participants completed all three sessions (12 females, 1 non-binary, Age range = 19 - 35, Avg age = 24.9, SD = 4.29). The participation was compensated with 40€. Due to the within-subjects design, every participant experienced all of the conditions. The order of the conditions are counterbalanced. For the analysis, we separate the participants in two disjoint groups based on the occurrence of sickness symptoms in the baseline condition. In line with former research, a simulator sickness questionnaire (SSQ) total score of 20 is chosen as clustering factor [Groth et al. 2022]. The group with individuals that got negatively affected consists of 17 participants (11 females).

Measurement

We used the SSQ for participant feedback on cybersickness [Kennedy et al. 1993]. Following common procedure, we let participants fill in the SSQ twice, before and after each session of the experiment, to counteract different daily conditions. The total sickness score as well as the corresponding subscores of the SSQ are calculated according to the original procedure of Kennedy et al. [Kennedy et al. 1993]. During the experiment we also asked the participants to press different buttons every time their feeling in comfort got worse or better. This discomfort includes all symptoms of the SSQ and can be in intervals as coarse or fine as participants choose. The responses allow for the calculation of the participants' individual level of discomfort over time [Groth et al. 2022]. After each session of the experiment we hold a semi-structured interview with the participants. The questions were:

1. Have you noticed anything unusual, and if so, what?
2. *Optionally:* How often did the effect occur?
3. *Optionally:* How disturbing did you find the effect?
4. How was your overall feeling during this session?

5. How immersed were you in the VR environment?

We chose an interview over subjective rankings of subtleness and immersion since the nature of the two manipulations is very different and comparing them on one numerical scale is prone to misinterpretation. The interview gave us the chance to dig deeper into the effects that were actually perceived by the participants and gain substantial insights about the impact of the investigated methods.

Procedure

The experiment for every participant was conducted in three sessions with one condition in each session (control, blur, or warp). The procedure of each session followed the same structure. In the beginning, every participant gave written consent and was informed about the experimental procedure and the possibility that negative symptoms may arise. Demographics have already been provided for participant registration. In each session, participants first filled the pre-experimental SSQ to capture their initial state, adjusted the VR glasses and performed a calibration routine for the eye tracker. In the experiments, participants were seated and ask to keep their head straight to increase comparability between sessions. The in-game camera was moved along the predefined path and participants could constantly indicate their well-being over the respective buttons. We ask participants to indicate when they experience severe negative feelings and the session was ended immediately in that case. Otherwise, the total time of the VR experience was 13 minutes. After every session of the experiment, participants first filled in the second SSQ, before we then conducted a semi-structured interview.

F.6.2 Analysis and Results

For the analysis of the experimental results, we performed pairwise two-sided dependent t-tests for repeated measures comparing both manipulation techniques to the control condition and to each other. Qualitative results of the interviews were determined by thematic analysis [Braun and Clarke 2006]. The further results focus only on the group of participants for whom the virtual simulation evoked significant cybersickness, since mitigation methods are less interesting for participants that do not get sick in the first place. As expected, the statistical results of the group of participants that were unaffected by the simulation show no significant change in the SSQ scores or times participants spent in the VR environment. Data from two participants had to be discarded from the time-based discomfort analysis due to improper task execution.

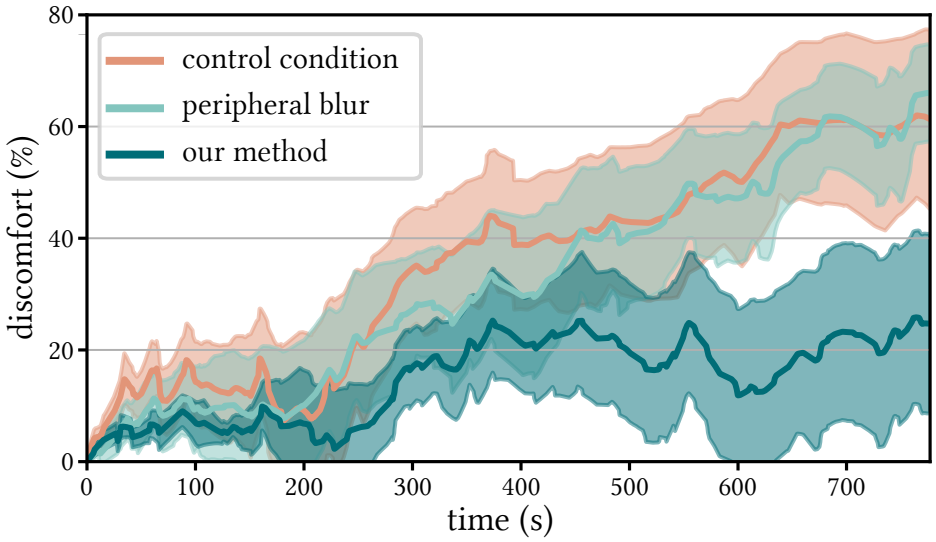

Figure F.8: Progression of the relative discomfort of all participants over the sessions, with shaded areas for the standard error of the mean (SEM).

Results for Effectiveness Figure F.9 shows the SSQs scores for total sickness and the three subscales nausea, disorientation, and oculomotor effects. The results confirm our technique's effectiveness in mitigating cybersickness. Our method significantly reduced the sickness scores (-31.8% over control) across all SSQ subscales (total: $T = 4.87$, $p = 0.0002$); nausea: $T = 3.81$, $p = 0.0015$; disorientation: $T = 4.37$, $p = 0.0005$; oculomotor: $T = 2.68$, $p = 0.0166$). While the peripheral blur technique also reduces the average cybersickness (-20.2%), it does not achieve a significant effect in our experiment. Simultaneously, the time participants are willing to spend in the VR environment (600.3s in control) was significantly increased when post-processing the output (Blur: +54.8s, Warp: +92.9s). However, only our warp condition had a significant influence on the increase of time in the VR environment ($T = -2.97$, $p = 0.009$). By introducing content distortions, the drop-out rate was reduced by 50% over the control session (control: 41.7%, blur: 29.2%, warp: 20.8%). From the qualitative results, a majority of 84% of participants stated that the session with our technique was the most enjoyable without being aware of the differences.

The positive impact of content-aware image warping on the comfort of users is further supported by the real-time data of the discomfort analysis (see Figure F.8). When applying our modulations, the general level of relative discomfort over time was significantly lower. While the relative level of discomfort is around 25% at the end, the temporal progression of the warp session only shows a rising trend between 240 and

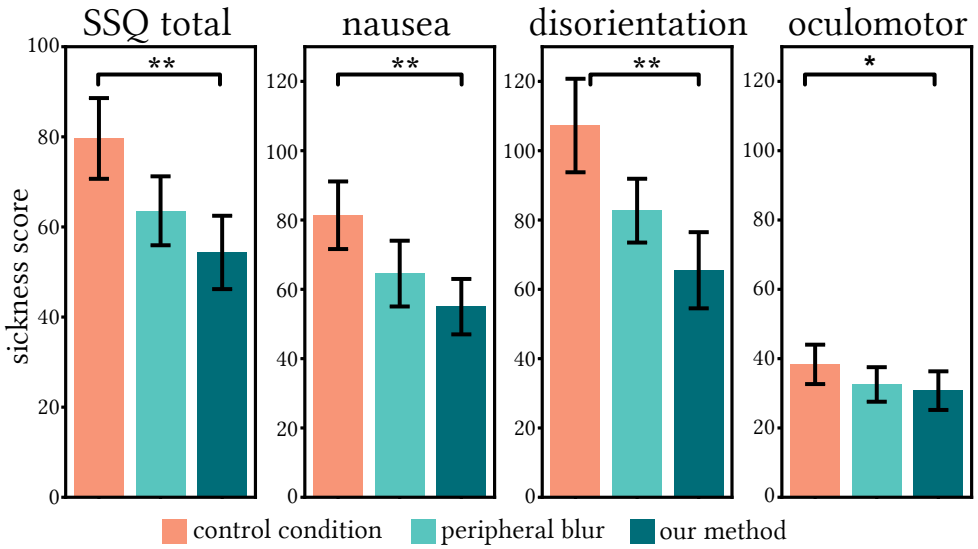

Figure F.9: SSQ results for all conditions. Error bars represent the SEM. Significance is denoted by '**' ($\mathbf{p \leq 0.01}$) and '*' ($\mathbf{p \leq 0.05}$).

300 seconds. The much larger remaining intervals, on the other hand, tend to oscillate around their constant mean. Participants were even able to recover from some of their discomfort in certain parts of the simulation, e.g., between 450 to 520 seconds. Such a recovery was not evident in the other conditions.

With the pre-defined camera trajectory and continuous discomfort metrics, we can further assess the relative efficacy of individual motion types (linear and angular). The mean discomfort increment (as percentage per minute) for rotational motions registered at 15.97 in the control scenario, 15.88 with peripheral blur applied, and 11.37 when warping was used. Linear motion was less severe and increased discomfort rates by 0.53% per minute without post-processing and 0.70% per minute with peripheral blur. With our warp implementation the data suggests even a decrease in discomfort during linear motion scenarios with -2.29% per minute.

Results for Subtlety A decisive factor for the widespread use of methods for cybersickness reduction is, in addition to efficiency, the unobtrusiveness of the methods. The perceptual model of our method is based on novel findings from our studies of the HVS (Section F.3). The free parameters of the model are calibrated to the conservative detection threshold in dedicated experiments. Further constraints in the implementation prevent excessive distortion of the scene. In the validation experiment, we investigated the subtleness of the applied image manipulations and the extent of distracting partici-

pants from their virtual experience. From the results of the semi-structured interviews, for 36% of the participants it was impossible to detect any changes made to the scene over the whole 13 minutes of the experiment. Surprisingly, the peripheral blur was less noticed in our experiment (undetected for 64%) than in former research with equivalent parameters. On the other hand, it was noted to be more distracting when perceived by participants.

A detailed investigation of the eye tracking results leads to the assumption that most of the detected artifacts of the warping were not caused by the visibility of the deformations themselves, but rather by insufficient resets during the saccades. Better hardware and saccade detection techniques have the potential to improve the subtleness of the resetting of scene deformations.

Further Findings Overall, the results show the same trend for men and women, with the control condition perceived as the most sickness inducing and the warping condition rated as the most pleasant experience (see Figure F.11). In line with former findings [Groth et al. 2022, Narciso et al. 2019], women experienced increased sickness symptoms, have a higher average sickness score, and dropped out of the experiment earlier. However, with our scene deformation technique applied, cybersickness was mitigated significantly for both, men ($T = 3.01$, $p = 0.0299$) and women ($T = 3.86$, $p = 0.0032$). The result give a positive indication for the use of warping methods for effective reduction of the general gender bias of VR experiences. This results, however, should be considered with care due to the small group size of male participants that actually got sick in the experiment.

Based on our qualitative inquiry, the influences on cybersickness are grouped into two themes: Influences of the scene and the type of movement. Most of the participants reported the forest scene to be worse than the city presentation. This was further reported to be influenced by the closeness of the surrounding objects, e.g. trees, which were further away in the urban area. In line with former research, rotational movements caused more severe cybersickness [Kim et al. 2021, Groth et al. 2022]. Participants reported that pitch rotations had an especially strong influence. One participant related the severity of pitch rotations to the high buildings of the city environment [P17]. Movements to the side were further claimed to be critical for sickness symptoms, while forward movements were not seen as a problem. The majority of participant experienced a high level of presence in the VR experience (15 out of 25 participants; 60%). Another 8 individuals reported the presence level to be at a decent amount, describing the experience as “medium immersing” or “somewhat real”. Participants’ preferences among

the session comparing immersion was less clear. Most of the participants reported that all sessions were equally immersing (10 participants). Followed by a preference for the control session and warping with 7 and 6 votes, respectively. The blurring was less accepted for an immersive experience (2 votes), which likely be attributed to the attenuation of details in the periphery.

F.7 Limitations and Future Work

The experiment reveals the peripheral blur to be more subtle than in prior studies using comparable intensities. Therefore, our method's predominance in effectiveness, though indicative, might differ when stronger blurring is applied. Notably, more pronounced blurring would further diminish peripheral visual details, while our method preserves the visual fidelity of the scene.

The current implementation of our method uses the motion flow of the rendering pipeline to retrieve content motion between frames. For content not relying on a 3D scene, like 360° videos, the computation would require a different input, e.g. vision-based optical flow. Hardware-accelerated optical flow algorithms promise real-time computation in milliseconds [Fast Optical Flow].

Another limitation of our method is the prediction of complex camera movements in short time intervals. Unpredictable changes in direction can disrupt the motion energy accumulation, leading to artifacts in the rendered scene. The full scene resets, triggered by the user's blinking, recover the original rendering every three seconds on average.

The distortions of the warped frames may also become visible in scenarios with near-camera presentations and scenes that are extensively regular, e.g. very regular low-poly scenes. Especially in the forest scene, multiple objects of the virtual environment got close to the camera and some participants reported recognizing artifacts of the distortions in these scenarios. Strong regularity, on the other hand, has the potential to disrupt the subtlety of the method because the optimization often results in a curvature effect of straight lines. In a scene consisting mostly of high contrast lines and regular patterns, the manipulations can become visible. However, while the presented city scene already contains considerable regularity, our method remained subtle in this scenario.

Against our expectations, the peripheral blur method was more subtle and less effective than in comparable implementations that use the same gradient [Lin et al. 2020a, Groth et al. 2021a]. Compared with the experiments of Lin et al. [2020a], our participants could not control the camera movements. Also, the scenes in our experiment were more complex, which may had an influence on the detectability of

the blurring. In contrast, our warping method was adjusted to a 50% detection rate which is confirmed by the experimental results. Using a more aggressive blurring of the peripheral content can result in a more effective mitigation of cybersickness. However, increasing the blur would result in even stronger suppression of peripheral details. The key motivation of our work is to allow effective cybersickness mitigation without the suppression of any visual details.

Our implementation does not yet separate between background and dynamic foreground objects. Dynamic objects do not induce vection, and therefore no cybersickness [Seno et al. 2009]. For scenes with extensively moving objects, a separation between dynamic and static objects would avoid unnecessary scene deformations. All visual movements that originate to the users' body movements, e.g. turning the head, should not be compensated because the corresponding vestibular signals are triggered accordingly. Our implementation considers this fact and only counteracts passive camera movements.

Potential future work that exploits the HVS are modulations of disparity to counteract forward and backward movements. These modulations involve adjusting the virtual environment's disparity to negate motion effects, potentially reducing sensory mismatches.

F.8 Conclusion

This research introduces a new technique for mitigating cybersickness in virtual reality environments without compromising the visual details of the scene. The perceptual model of our technique is based on two key findings: Firstly, reducing the visual motion of objects in the motion-sensitive peripheral field does not alter the camera velocity perception as long as the movement speed in the fovea is adjusted accordingly. Secondly, angular camera movements can be adjusted to visually move on a linear trajectory by scaling the image content with eccentricity. By means of calibration experiments, perceptual thresholds were found that keep the effects unnoticeable while maximizing the reduction of vection. We presented an effective implementation of our method that runs in real-time as a post-process on the rendered image. Experimental validation confirms the effectiveness of our approach to mitigate cybersickness, while the modifications often remained completely undetected.

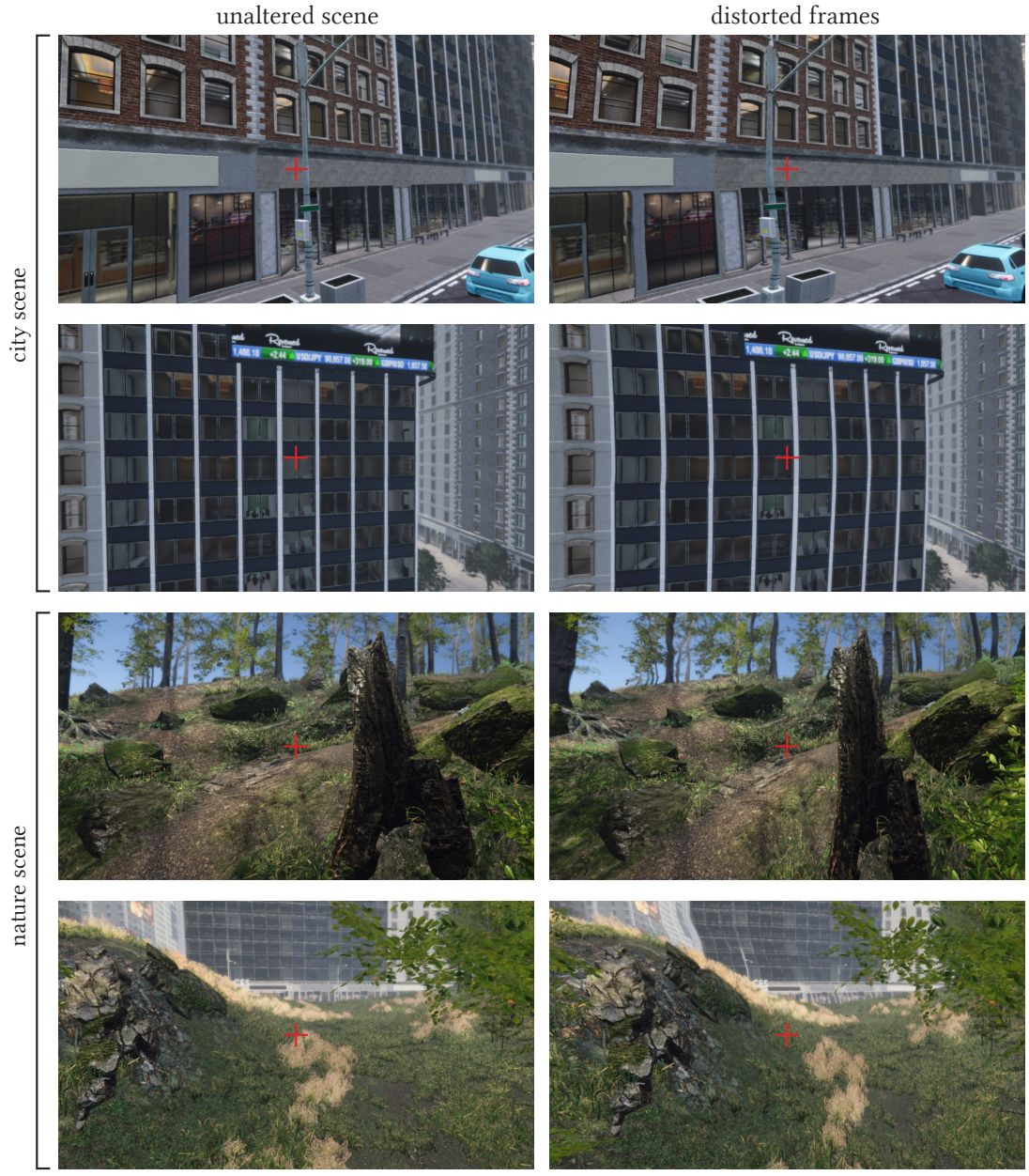

Figure F.10: Exemplary frames of the validation experiment. In the left column we present the unmanipulated frames of the control condition. After applying our method, the frames are distorted to reduce the visual motion (right column). The gaze point is marked with the red cross. In the VR glasses, the effect is gaze-contingent and mostly unnoticeable to users.

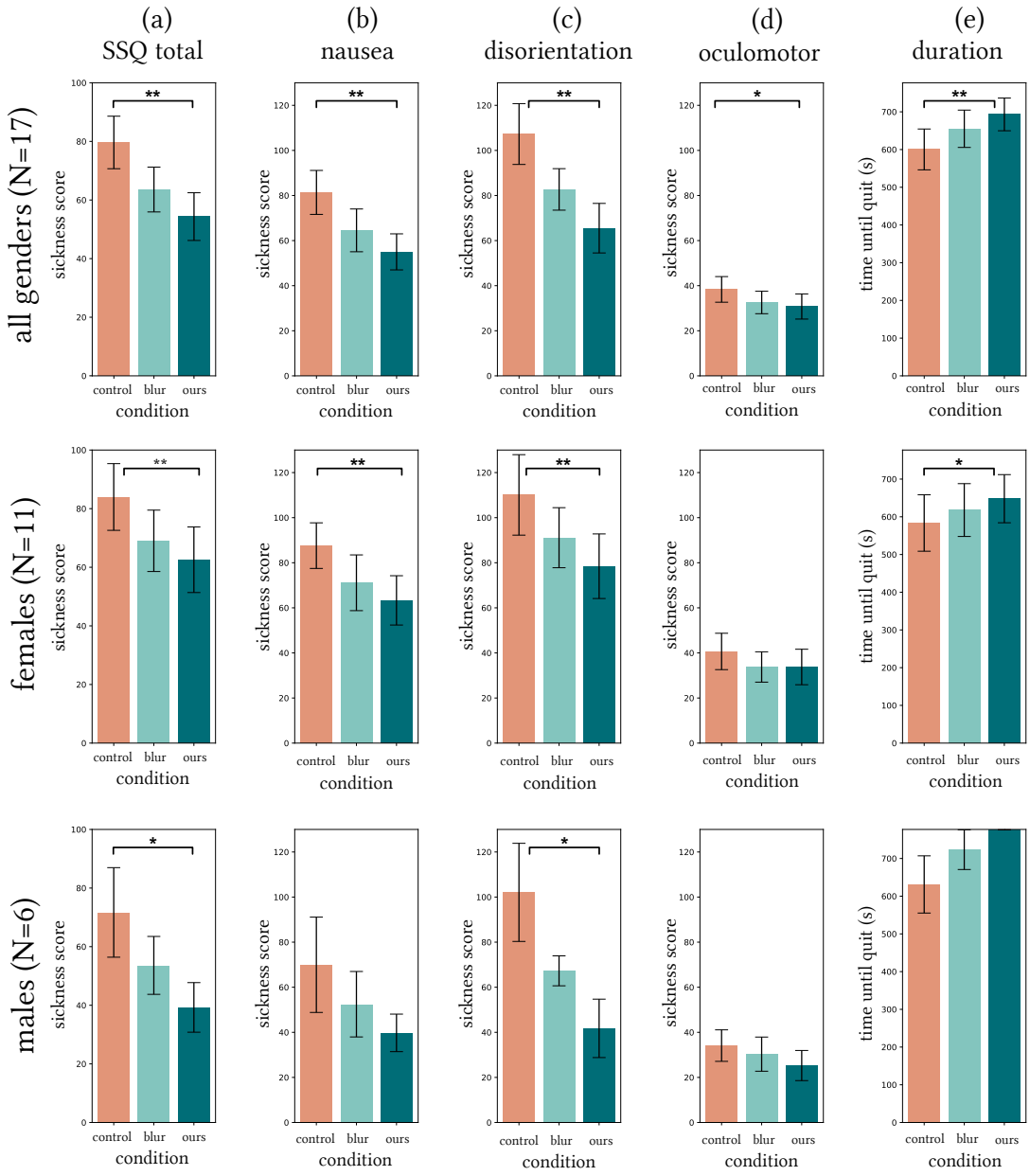

Figure F.11: Results of the SSQ and durations for the control condition, the peripheral blur and our warping technique. This analysis only includes participants that were negatively affected (SSQ total score > 20). Error bars represent the SEM. **(a)** SSQ results for the total score. **(b-d)** results for each of the SSQ subscales. **(e)** duration people were willing to spend in the virtual environment. Significant results are denoted by '**' ($\mathbf{p \leq 0.01}$) and '*' ($\mathbf{p \leq 0.05}$).

4 Conclusion and Future Perspectives

The publications that formed the main part of this dissertation presented a variety of techniques and perceptual aspects crucial to maintain engaging and satisfactory user experiences in virtual reality. This chapter summarizes the methods, key insights, and contributions of the presented works and concludes with an outlook toward future research directions and possible applications.

4.1 Conclusion

This dissertation comprises six publications each presenting different techniques to enhance the user experience in certain VR scenarios. While our proposed methods differ in their generalization and applicability, they follow the same general concept: We consider beneficial properties of human perception to develop techniques that are subtle to the observer but effectively enrich the user experience of VR applications. A central aspect of this goal is the mitigation of negative effects, which were found to be the most influential factor of VR experiences [Nichols and Patel 2002, Stanney et al. 2003, Cobb et al. 1999]. Especially, cybersickness and breaks in immersion can lead to rapid degradation in comfort and user experience, often resulting in full rejection of the technology. This thesis presents techniques that can significantly reduce these negative aspects, leaving the user with a more comfortable and enjoyable experience. The individual works require different levels of adaptability of the final applications (see Figure 1.2). Applications that allow considerable adjustments favor techniques that are highly efficient in controlling negative symptoms. In contrast, methods with no or only minimal requirements for the final application are most generalizable. In the following, the individual works will be collectively discussed in the global context of system requirements and final application scenarios.

Hardware Augmentation with Additional Sensors

The first two works presented in this dissertation (A and B) investigate the augmentation of existing VR systems with additional stimulation hardware. The two papers explore two key aspects of VR user experience: mitigating negative physiological ef-

fects that arise as a reaction to the visual-vestibular mismatch evoked by the display of visual-only motion (Paper A), and maintaining immersion through coherent tactile interaction feedback in scenarios where the virtual world diverges from physical reality (Paper B).

For the mitigation of cybersickness, we extended the VR hardware with professional stimulation devices to induce artificial vestibular signals in seated participants. We only considered low, electric currents ($< 2.5\,\mathrm{mA}$) that are safe and comfortable for the users. The stimulation induces a feeling of self-motion by vestibular signals, which we showed can be controlled for arbitrary rotation axes in three-dimensional space. When the stimulation is coupled with the visual motion cues of the VR presentation, significant reductions in cybersickness are achieved, while users feel overall more comfortable.

To convey convincing hand interactions, we exploit the change blindness effect which masks instantaneous visual changes from users that are applied to a virtual scene. We found that actively triggered eye blinks cause a short-time change blindness effect starting consistently 100 ms after the stimulation. In our work, we provoked natural eye blinks by actuating the orbicularis oculi muscle in both eyelids with low electric current stimulation. The subsequent temporary imperceptibility allowed us to perform subtle repositioning of the virtual hand to counteract mismatches between the real-to-virtual comparison and, therefore, achieve a redirection onto the trajectory towards the real-world target.

In conclusion, substantial future potential lies in the integration of stimulation devices into VR systems to improve critical factors of user experience. In both works, the effects of active stimulation are leveraged to create an experience that is not only transported by visual information but is multimodal. In the first work, the vestibular sense was incorporated into the experience, while in the second work, the experience is enhanced with haptic feedback. The inclusion of multiple sensory impressions in the VR experience enables more comprehensive, immersive experiences that bring the user closer to a truly convincing virtual sensation.

Adjusting the Rendering Pipeline

While the introduction of additional stimulation hardware showed significant success in providing better user experiences, these techniques require custom hardware which is not yet broadly available. In contrast, publication C proposes a software-only solution where the rendering pipeline is adjusted. The proposed solution specifically considers immersive video experiences where high frame rates and high quality are most challenging for 360° videos. We presented a novel video codec that compresses animation

frames with wavelet-based transforms. We show that with this way of data coding, performance-efficient foveation can be achieved even for pre-recorded content. The perceptual experiments further confirmed that the higher playback speeds that we can achieve for 360° videos contribute to a more comfortable and enjoyable user experience.

Nowadays, pre-recorded content in VR is typically attributed a secondary role in research compared to live-rendered 3D scenes. At the same time, the realistic visualization of panoramic content is highly challenging and rarely achieves the same visual quality as fully virtual content. Viewport-dependent video coding and adapted pipelines have considerable potential here; an optimized codec can not only achieve the quality and playback times at the display frequency of modern VR glasses but also enables foveation of videos for headsets with eye tracking. As a result, with the right procedures, 360° videos can be displayed with the same immersive experience as rendered 3D worlds in VR.

Post-Process Modulations

The techniques so far needed to be incorporated into the application development and required alterations in the rendering pipeline. However, customization of individual applications for integrating the proposed techniques cannot be guaranteed in all cases. Instead, direct manipulations of the animation frames are broadly applicable on the system level and can directly be incorporated into the software of the VR system. For this group of applications, we presented several approaches (Publications D, E, and F) that operate by post-processing the rendering output. These approaches target the mitigation of cybersickness by reducing the visually-induced motion information. The first techniques (D and E) apply screen space blurring operations to lower the perceptual threshold of temporal change in the periphery. The quality degradation was foveated and was mostly imperceptible for participants. When we compared the technique with an opaque restriction of the FOV and an unaltered scene, peripheral blurring offered reasonable improvements in the user experience. In our final work (Paper F), we followed the same line of thought and focused on the adjustment in the perception of self-motion that is induced by visual motion cues in the periphery. We found that humans perceive the apparent movement speed of a scene as a consensus of the individual local regions. Given a considerable natural tolerance for changes that stay undetected, we demonstrated that peripheral motion can be reduced by 30% when the movement is compensated in the foveal region. We further demonstrated that reductions in vection can further be achieved for rotational camera movements by scaling the scene objects according to their eccentricity. By incorporating our technique into a realistic

VR application we illustrated the enormous potential of fine-tuned scene manipulations to reduce cybersickness with no to minimal compensation in visual fidelity.

In conclusion, even highly restricted operations in image space for highly generalizable methods offer great potential for the improvement of user comfort. Due to the dominance of human vision for the overall perception of self-motion in scenes [Land and Fernald 1992], significant improvements can be achieved by subtle adaptation of self-motion effects, i.e., vection. The presented real-time capable methods enable adaptation at the system level and therefore allow an improvement of the user experience for arbitrary VR applications through systematic embedding into the render pipeline of HMDs.

In essence, this dissertation demonstrates substantial progress in creating more immersive, comfortable, and accessible VR experiences through both hardware augmentations and software innovations. Key considerations for future VR implementations include the following:

- Incorporating multiple sensory inputs (vestibular, tactile) to augment the visual presentation significantly enhances VR experiences in realism and comfort.
- Adjustments in video compression and 3D rendering through perceptual effects can greatly improve the frame rate and visual quality of VR content without additional hardware.
- Techniques developed for post-process modulations in image space are promising for broad application in future HMDs by improving user comfort with minimal compromise in visual fidelity.

4.2 Future Research and Applications

This dissertation presents meaningful advancements in the field of virtual reality by overcoming major limitations in user experience through novel techniques and approaches. The investigations and research results underscore the dynamic nature of VR technology and the benefits of focusing on human perception for the development of VR applications. While the thesis has successfully addressed multiple key challenges in VR, such as the mitigation of cybersickness, enhancement of sensory immersion, and improvement of visual fidelity, it also highlights various directions for future research.

Future Perspectives

One interesting direction for future work is the exploration of advanced locomotion techniques. The common continuous movement behavior is directly transferred from traditional locomotion in desktop video games but evokes strong symptoms of cybersickness in VR. More sophisticated movement methods that are precisely adjusted for VR usage have the potential to drastically lower the occurrence of negative physiological symptoms from the motion. Current investigations of related work in this field focus on the incorporation of in-place physical body movements for virtual steering [Wang et al. 2023, Feasel et al. 2011, Kassler et al. 2010, Iwata et al. 2005, Avila and Bailey 2014, Cakmak and Hager 2014], redirected walking to utilize the physical space [Sun et al. 2018, Langbehn et al. 2018, Nguyen and Kunz 2018, Grechkin et al. 2016, Thomas and Rosenberg 2020], and modification of the general movement type [Weißker et al. 2018, Moghadam et al. 2018].

Redirected walking is a major area of interest in VR research. Current techniques exploit natural perception effects like change blindness during eye blinks [Langbehn et al. 2018] and saccadic suppression [Sun et al. 2018] for undetected alteration of the user's path. In our research, we demonstrated that the efficiency of redirection for rapid hand movements could be significantly enhanced with actively stimulated eye blinks. Including active stimulated change blindness effects in redirected walking techniques would allow one to react instantaneously to time-critical situations, e.g., impending collision, by steering the user away without the need to break the immersion.

While many investigations of this dissertation considered immersive video material as a medium for VR, following common procedures, we used virtual 3D spaces for hand interaction scenarios. However, increasing interactability might especially be intriguing for real-world content. To interact with recorded scenes, the stereo content has to be interpreted to form a three-dimensional scene. Successful implementation of this step would not only allow interaction via touch but also enable users to decide on their movement trajectories in these worlds. While the reconstruction of a recorded scene is technically challenging, recent developments in neural scene reconstruction are promising to provide solutions of sufficient quality and speed [Kerbl et al. 2023, Hahlbohm et al. 2023].

In Paper F, we reduced cybersickness by deceleration of object velocity in the periphery and accompanied acceleration of foveal objects. We found that this balanced manipulation retains the global scene speed perception. For this work, a linear model was used to describe the relationship of movement scaling based on eccentricity. While the model was proven to be a sufficient solution for cybersickness reduction, a linear re-

lationship may not reflect the natural perception behavior, and other models potentially describe the natural behavior even closer. Therefore, future research should investigate if other representations are more reasonable for the use case.

The screen space distortions in Paper F adjust the animation frames to displace objects for reduced vection. Another promising direction for effective cybersickness reduction without scene knowledge involves the introduction of visual counter-vection. Recent research already indicates the neutralization of optical flow by the induction of opposing visual markers as an effective method to reduce cybersickness [Park et al. 2022]. However, the current methods are still highly noticeable and interfering. Future work could build upon these findings to develop more sophisticated methods to subtly reduce cybersickness.

Future techniques can further consider directly manipulating the 3D scene data for cybersickness reduction. With this approach, scene objects can be manipulated in real time by deformation or displacement that is aligned with the user's motion and perceptual expectations. While these techniques would require an adjustment on the application level, they are likely to provide an effective reduction of visual-vestibular conflicts with minimal visual artifacts that can arise from post-processing methods.

Finally, preliminary research indicates that noisy GVS could be a promising avenue for investigations of cybersickness [Weech et al. 2020]. This research involves applying non-regular, fluctuating electrical stimuli to the vestibular system, shifting the attention of the scene perception more toward the virtual information. The aim would be to refine this technique to achieve an optimal balance between effectiveness in cybersickness reduction and user comfort.

Applications

The future of VR, as envisioned in this dissertation, is set to transform future VR applications and system-level development. The vision of this thesis foresees a world where multimodal stimulations are seamlessly integrated into VR systems, providing not only visual but also vestibular and tactile cues, especially in training simulations for military, medical, and emergency response applications. Such an integration will evoke an unrivaled level of realism in immersive environments allowing for effective skill development and the realization of unique opportunities.

For immersive video content, further innovation calls for establishing global standards for new video codecs specifically designed for panoramic VR content. Such standards should address the unique challenges of VR like the need for dynamic video decoding and high frame rates, to ensure seamless and high-quality 360° experiences.

The direct integration of both advanced post-processing solutions and stimulation sensors into VR headsets offers great potential for future VR systems. The inclusion of hardware for galvanic vestibular stimulation (GVS) and blink stimulation as standard features can enable more realistic and immersive experiences by default, enhancing the perception of motion and interaction in virtual environments. Integrating post-processing techniques into the rendering pipeline of HMDs can play a pivotal role in improving user comfort and excitement. These post-processing solutions can be implemented in the software of the VR system and, thus, promise universally enhanced user experiences for arbitrary applications and a potential widespread adoption in the consumer market. Together, the advancements motivate a transformation in VR, where system design aligns visual processing with physical sensory feedback to redefine our engagement with virtual realms.

Overall, this dissertation presents fundamental contributions to the field of virtual reality, laying the groundwork for innovative approaches and technologies that are more immersive, accessible, and user-friendly.

Bibliography

P. Abtahi and S. Follmer. Visuo-haptic illusions for improving the perceived performance of shape displays. In *Conf. Human Factors in Computing Systems*, pages 1–13, 2018.

P. Abtahi, B. Landry, J. Yang, M. Pavone, S. Follmer, and J. A. Landay. Beyond the force: Using quadcopters to appropriate objects and the environment for haptics in virtual reality. In *Conf. Human Factors in Computing Systems*, pages 1–13, 2019.

R. Addams. An account of a peculiar optical phænomenon seen after having looked at a moving body. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 5(29):373–374, 1834.

E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2):284–299, 1985.

E. H. Adelson and J. A. Movshon. Phenomenal coherence of moving visual patterns. *Nature*, 300(5892):523–525, 1982.

I. B. Adhanom, N. N. Griffin, P. MacNeilage, and E. Folmer. The effect of a foveated field-of-view restrictor on vr sickness. In *IEEE Conf. Virtual Reality and 3D User Interfaces*, pages 645–652. IEEE, 2020.

M. Al Zayer, I. B. Adhanom, P. MacNeilage, and E. Folmer. The effect of field-of-view restriction on sex bias in vr sickness and spatial navigation performance. In *Conf. Human Factors in Computing Systems*, pages 1–12, 2019.

G. J. Andersen. Perception of self-motion: psychophysical and computational approaches. *Psychological bulletin*, 99(1):52, 1986.

S. J. Anderson, K. T. Mullen, and R. F. Hess. Human peripheral spatial resolution for achromatic and chromatic stimuli: limits imposed by optical and retinal factors. *The Journal of physiology*, 442(1):47–64, 1991.

K. Aoyama, H. Iizuka, H. Ando, and T. Maeda. Four-pole galvanic vestibular stimulation causes body sway about three axes. *Scientific Reports*, 5:10168, 2015.

BIBLIOGRAPHY

K. Aoyama, M. Mizukami, T. Maeda, and H. Ando. Modeling the enhancement effect of countercurrent on acceleration perception in galvanic vestibular stimulation. In *Proc. Int. Conf. Advances in Computer Entertainment Technology*, 2016.

D. Au, J. Tong, R. Allison, and L. Wilcox. The impact of conflicting ordinal and metric depth information on depth matching. *Journal of Vision*, 22(14):3739–3739, 2022.

L. Avila and M. Bailey. Virtual reality for the masses. *IEEE computer graphics and applications*, 34(05):103–104, 2014.

M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Conf. Human Factors in Computing Systems*, pages 1968–1979, 2016.

P. Bala, D. Dionísio, V. Nisi, and N. Nunes. Visually induced motion sickness in 360° videos: Comparing and combining visual optimization techniques. In *IEEE Int. Symp. Mixed and Augmented Reality Adjunct*, pages 244–249. IEEE, 2018.

P. Bala, I. Oakley, V. Nisi, and N. Nunes. Staying on track: a comparative study on the use of optical flow in 360 video to mitigate vims. In *ACM Int. Conf. Interactive Media Experiences*, pages 82–93, 2020.

R. W. Baloh and K. Kerber. *Baloh and Honrubia's clinical neurophysiology of the vestibular system*, volume 77. Oxford university press, 2011.

S. Ban and K. H. Hyun. Directional force feedback: Mechanical force concentration for immersive experience in virtual reality. *Applied Sciences*, 9(18), 2019.

A. Banitalebi-Dehkordi, M. T. Pourazad, and P. Nasiopoulos. The effect of frame rate on 3d video quality and bitrate. *3D Research*, 6(1):1–13, 2015.

K. R. Bankieris, V. R. Bejjanki, and R. N. Aslin. Sensory cue-combination in the context of newly learned categories. *Scientific Reports*, 7(1):1–10, 2017.

G. R. Barnes. Cognitive processes involved in smooth pursuit eye movements. *Brain and cognition*, 68(3):309–326, 2008.

N. Bartlett, J. Brown, Y. Hsia, C. Mueller, and L. Riggs. *Vision and Visual Perception*. John Wiley & Sons, 1965.

B. Benda, S. Esmaeili, and E. D. Ragan. Determining detection thresholds for fixed positional offsets for virtual hand remapping in virtual reality. In *IEEE Int. Symp. Mixed and Augmented Reality*, pages 269–278, 2020.

A. Berni and Y. Borgianni. Applications of virtual reality in engineering and product design: Why, what, how, when and where. *Electronics*, 9(7):1064, 2020.

N. Birbaumer and R. F. Schmidt. *Hören und Gleichgewicht*, pages 415–438. 2010. doi:10.1007/978-3-540-95938-0_18.

G. Bjontegaard. Calculation of average PSNR differences between RD-curves. *ITU-T SG16/Q6 input document VCEG-M33*, 2001.

Blausen Medical. Medical gallery of blausen medical. *WikiJournal of Medicine*, 2(1), 2014. doi:10.15347/wjm/2014.010.

R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International journal of computer vision*, 1(1):7–55, 1987.

G. Boopathi and S. Arockiasamy. Image compression: Wavelet transform using radial basis function (rbf) neural network. In *India Conf.*, pages 340–344. IEEE, 2012.

G. Borg. *Borg's perceived exertion and pain scales.* Human Kinetics, 1998.

J. E. Bos, S. C. de Vries, M. L. van Emmerik, and E. L. Groen. The effect of internal and external fields of view on visually induced motion sickness. *Applied ergonomics*, 41(4):516–521, 2010.

R. Bouville, V. Gouranton, and B. Arnaldi. Virtual reality rehearsals for acting with visual effects. In *International Conference on Computer Graphics & Interactive Techniques*, pages 1–8, 2016.

O. Braddick. A short-range process in apparent motion. *Vision research*, 14(7):519–527, 1974.

T. Brandt, J. Dichgans, and E. Koenig. Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental brain research*, 16:476–491, 1973.

V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

W.-P. Brinkman, A. R. Hoekstra, and R. van EGMOND. The effect of 3d audio and other audio techniques on virtual reality experience. *Annual Review of Cybertherapy and Telemedicine*, pages 44–48, 2015.

BIBLIOGRAPHY

F. P. Brooks Jr, M. Ouh-Young, J. J. Batter, and P. Jerome Kilpatrick. Project grope – haptic displays for scientific visualization. *ACM SIGGRAPH*, 24(4):177–185, 1990.

V. Bruce, P. Green, and M. Georgeson. *Visual Perception: Physiology, Psychology, & Ecology*. Psychology Press, 2003.

A. R. Brunoni, J. Amadera, B. Berbel, M. S. Volz, B. G. Rizzerio, and F. Fregni. A systematic review on reporting and assessment of adverse effects associated with transcranial direct current stimulation. *Int. Journal of Neuropsychopharmacology*, 14 (8):1133–1145, 2011.

P. Budhiraja, M. R. Miller, A. K. Modi, and D. Forsyth. Rotation blurring: use of artificial blurring to reduce cybersickness in virtual reality first person shooters. *arXiv preprint arXiv:1710.02599*, 2017.

G. C. Burdea and P. Coiffet. *Virtual reality technology*. John Wiley & Sons, 2003.

E. Burns, S. Razzaque, A. T. Panter, M. C. Whitton, M. R. McCallus, and F. P. Brooks Jr. The hand is more easily fooled than the eye: Users are more sensitive to visual interpenetration than to visual-proprioceptive discrepancy. *Presence: Teleoperators and Virtual Environments*, 15(1):1–15, 2006.

R. Byrne, J. Marshall, and F. Mueller. Balance ninja: towards the design of digital vertigo games via galvanic vestibular stimulation. In *Proc. Annual Symp. Computer-Human Interaction in Play*, pages 159–170, 2016.

T. Cakmak and H. Hager. Cyberith virtualizer: a locomotion device for virtual reality. In *ACM SIGGRAPH Emerging Technologies*, pages 1–1. 2014.

K. C. Cassidy, J. Šefčík, Y. Raghav, A. Chang, and J. D. Durrant. Proteinvr: Web-based molecular visualization in virtual reality. *PLoS computational biology*, 16(3): e1007747, 2020.

S. Castillo, T. Judd, and D. Gutierrez. Using eye-tracking to assess different image retargeting methods. In *ACM SIGGRAPH Symp. Applied Perception in Graphics and Visualization*, pages 7–14, 2011.

M. J. Cevette, J. Stepanek, D. Cocco, A. M. Galea, G. N. Pradhan, L. S. Wagner, S. R. Oakley, B. E. Smith, D. A. Zapala, and K. H. Brookler. Oculo-vestibular recoupling using galvanic vestibular stimulation to mitigate simulator sickness. *Aviation, space, and environmental medicine*, 83(6):549–555, 2012.

L.-P. Cheng, E. Ofek, C. Holz, H. Benko, and A. D. Wilson. Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop. In *Conf. Human Factors in Computing Systems*, page 3718–3728, 2017.

B. Choi, Y. Wang, M. Hannuksela, Y. Lim, and A. Murtaza. Information technology–coded representation of immersive media (mpeg-i)–part 2: Omnidirectional media format. *ISO/IEC*, pages 23090–23092, 2018.

C.-H. Chu, W. Bernstein, Y. Zhang, V. R. Krishnamurthy, and J. Ma. Extended reality in design and manufacturing. *Journal of Computing and Information Science in Engineering*, 24(3), 2024.

Citroën. SEETROËN glasses, 2024. URL https://lifestyle.citroen.com/glasses-without-lenses-seetroen-inspired-by-you.html.

S. V. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson. Virtual reality-induced symptoms and effects (vrise). *Presence: Teleoperators & Virtual Environments*, 8(2):169–186, 1999.

A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5):485–560, 1992. doi:https://doi.org/10.1002/cpa.3160450502.

Conference Générale des Poids et Mesures. Resolution 1 of the 17th cgpm. *Bureau International des Poids et Mesures*, 1983.

X. Corbillon, G. Simon, A. Devlic, and J. Chakareski. Viewport-adaptive navigable 360-degree video delivery. In *Int. Conf. Communications*, pages 1–7. IEEE, 2017.

S. Corvino, A. Piazza, T. Spiriev, R. Tafuto, F. Corrivetti, D. Solari, L. M. Cavallo, A. Di Somma, J. Enseñat, M. de Notaris, and G. Iaconetta. The sellar region as seen from transcranial and endonasal perspectives: Exploring bony landmarks through new surface photorealistic 3d models reconstruction for neurosurgical anatomy training. *World Neurosurgery*, 2024. doi:10.1016/j.wneu.2024.02.022.

G. Cruccu and G. Deuschl. The clinical use of brainstem reflexes and hand-muscle reflexes. *Clinical Neurophysiology*, 111(3):371–387, 2000.

K. E. Cullen. The vestibular system: multimodal integration and encoding of self-motion for motor control. *Trends in neurosciences*, 35(3):185–196, 2012.

BIBLIOGRAPHY

S. Daly. Engineering observations from spatiovelocity and spatiotemporal visual models. In *Vision Models and Applications to Image and Video Processing*, pages 179–200. 2001.

L. Dandona and R. Dandona. Revision of visual impairment definitions in the international statistical classification of diseases. *BMC medicine*, 4:1–7, 2006.

F. Danieau, J. Fleureau, A. Cabec, P. Kerbiriou, P. Guillotel, N. Mollet, M. Christie, and A. Lécuyer. Framework for enhancing video viewing experience with haptic effects of motion. In *2012 IEEE Haptics Symposium*, pages 541–546, 2012.

R. P. Darken, W. R. Cockayne, and D. Carmein. The omni-directional treadmill: a locomotion device for virtual worlds. In *Proceedings of the ACM symposium on User interface software and technology*, pages 213–221, 1997.

P. Das, M. Zhu, L. McLaughlin, Z. Bilgrami, and R. L. Milanaik. Augmented reality video games: new possibilities and implications for children and adolescents. *Multimodal Technologies and Interaction*, 1(2):8, 2017.

B. Day, A. Severac Cauquil, L. Bartolomei, M. Pastor, and I. Lyon. Human body-segment tilts induced by galvanic stimulation: a vestibularly driven balance protection mechanism. *The Journal of physiology*, 500(3):661–672, 1997.

M. S. Dennison, A. Z. Wisti, and M. D'Zmura. Use of physiological signals to predict cybersickness. *Displays*, 44:42–52, 2016.

J. Dichgans and T. Brandt. Visual-vestibular interaction: Effects on self-motion perception and postural control. In *Perception*, pages 755–804. 1978.

P. DiZio and J. Lackner. Circumventing side effects of immersive virtual environments. *Advances in human factors/ergonomics*, 21:893–896, 1997.

T. G. Dobie. *Motion sickness: a motion adaptation syndrome*, volume 6. Springer, 2019.

L. Dominjon, A. Lécuyer, J.-M. Burkhardt, P. Richard, and S. Richir. Influence of control/display ratio on the perception of mass of manipulated objects in virtual environments. In *IEEE Conf. Virtual Reality and 3D User Interfaces*, pages 19–25, 2005.

K. Dooley. Storytelling with virtual reality in 360-degrees: a new screen grammar. *Studies in Australasian Cinema*, 11(3):161–171, 2017. doi:10.1080/17503175.2017.1387357.

I. Doweck, C. R. Gordon, A. Shlitner, O. Spitzer, A. Gonen, O. Binah, Y. Melamed, and A. Shupak. Alterations in r–r variability associated with experimental motion sickness. *Journal of the Autonomic Nervous System*, 67(1):31 – 37, 1997. ISSN 0165-1838. doi:10.1016/S0165-1838(97)00090-8.

D. D. Doyle, A. L. Jennings, and J. T. Black. Optical flow background estimation for real-time pan/tilt camera object tracking. *Measurement*, 48:195–207, 2014.

N. Dużmańska, P. Strojny, and A. Strojny. Can simulator sickness be avoided? a review on temporal aspects of simulator sickness. *Front. Psychol.*, 9:2132, 2018.

S. D'Amour, L. R. Harris, S. Berti, and B. Keshavarz. The role of cognitive factors and personality traits in the perception of illusory self-motion (vection). *Attention, Perception, & Psychophysics*, 83:1804–1817, 2021.

E. Ebrahimi, B. Altenhoff, L. Hartman, A. Jones, S. Babu, C. Pagano, and T. Davis. Effects of visual and proprioceptive information in visuo-motor calibration during a closed-loop physical reach task in immersive virtual environments. In *ACM Symp. Applied Perception*, pages 103–110, 2014.

E. Ebrahimi, B. Altenhoff, C. Pagano, and S. Babu. Carryover effects of calibration to visual and proprioceptive information on near field distance judgments in 3d user interaction. In *IEEE Symp. 3D User Interfaces*, pages 97–104. IEEE, 2015.

M. El Beheiry, S. Doutreligne, C. Caporal, C. Ostertag, M. Dahan, and J.-B. Masson. Virtual reality: beyond visualization. *Journal of molecular biology*, 431(7):1315–1321, 2019.

M. Elwardy, H.-J. Zepernick, Y. Hu, T. M. C. Chu, and V. Sundstedt. Evaluation of simulator sickness for 360° videos on an hmd subject to participants' experience with virtual reality. In *IEEE Conf. Virtual Reality and 3D User Interfaces Abstracts and Workshops*, pages 477–484. IEEE, 2020.

M. D. S. Eremita and J. Chitra. Effect of virtual reality therapy with brain gym exercises for sleep-deprived individuals: A randomized clinical trial. *Archives of Medicine and Health Sciences*, 2023.

BIBLIOGRAPHY

S. Esmaeili, B. Benda, and E. D. Ragan. Detection of scaled hand interactions in virtual reality: The effects of motion direction and task complexity. In *IEEE Conf. Virtual Reality and 3D User Interfaces*, pages 453–462, 2020.

C. Evinger, K. A. Manning, and P. A. Sibony. Eyelid movements. mechanisms and normal data. *Investigative Ophthalmology & Visual Science*, 32(2):387–400, 1991.

S. Exner. Ein versuch über die netzhautperipherie als organ zur wahrnehmung von bewegungen. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, 38(1): 217–218, 1886.

M. Faridan, M. Friedel, and R. Suzuki. Ultrabots: Large-area mid-air haptics for vr with robotically actuated ultrasound transducers. In *Adjunct Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 1–3, 2022.

Y. Farmani and R. J. Teather. Viewpoint snapping to reduce cybersickness in virtual reality. In *Proc. Graphics Interface Conf.*, pages 168–175. Canadian Human-Computer Communications Society, 2018.

Fast Optical Flow. Nvidia optical flow sdk. https://developer.nvidia.com/optical-flow-sdk, 2023. Accessed: 2024-01-17.

J. Feasel, M. C. Whitton, L. Kassler, F. P. Brooks, and M. D. Lewek. The integrated virtual environment rehabilitation treadmill system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(3):290–297, 2011.

G. T. Fechner. *Vorschule der Aesthetik*, volume 1. Breitkopf & Härtel, 1876.

A. S. Fernandes and S. K. Feiner. Combating vr sickness through subtle dynamic field-of-view modification. In *IEEE Symp. 3D User Interfaces (3DUI)*, pages 201–210. IEEE, 2016.

D. Finlay. Motion perception in the peripheral visual field. *Perception*, 11(4):457–462, 1982.

D. L. Fisher, M. Rizzo, J. Caird, and J. D. Lee. *Handbook of driving simulation for engineering, medicine, and psychology*. CRC Press, 2011.

R. Fitzpatrick, D. Burke, and S. C. Gandevia. Task-dependent reflex responses and movement illusions evoked by galvanic vestibular stimulation in standing humans. *The Journal of physiology*, 478(2):363–372, 1994.

R. C. Fitzpatrick and B. L. Day. Probing the human vestibular system with galvanic stimulation. *Journal of applied physiology*, 96(6):2301–2316, 2004.

D. J. Fleet and K. Langley. Computational analysis of non-fourier motion. *Vision research*, 34(22):3057–3079, 1994.

A. Frigerio, J. T. Heaton, P. Cavallari, C. Knox, M. H. Hohman, and T. A. Hadlock. Electrical stimulation of eye blink in individuals with acute facial palsy: progress toward a bionic blink. *Plastic and Reconstructive Surgery*, 136(4):515–523, 2015.

C. Fujimoto, Y. Yamamoto, T. Kamogashira, M. Kinoshita, N. Egami, Y. Uemura, F. Togo, T. Yamasoba, and S. Iwasaki. Noisy galvanic vestibular stimulation induces a sustained improvement in body balance in elderly adults. *Scientific reports*, 6(1): 1–8, 2016.

L. Galvani. De viribus electricitatis in motu musculari commentarius. 1791.

S. Garbutt, Y. Han, A. N. Kumar, M. Harwood, C. M. Harris, and R. J. Leigh. Vertical optokinetic nystagmus and saccades in normal human subjects. *Investigative ophthalmology & visual science*, 44(9):3833–3841, 2003.

R. J. García-Hernández, C. Anthes, M. Wiedemann, and D. Kranzlmüller. Perspectives for using virtual reality to extend visual data mining in information visualization. In *2016 IEEE Aerospace Conference*, pages 1–11. IEEE, 2016.

A. M. Gavgani, K. V. Nesbitt, K. L. Blackmore, and E. Nalivaiko. Profiling subjective symptoms and autonomic changes associated with cybersickness. *Autonomic Neuroscience*, 203:41–50, 2017.

P. J. Gianaros, K. S. Quigley, E. R. Muth, M. E. Levine, R. C. Vasko, Jr, and R. M. Stern. Relationship between temporal changes in cardiac parasympathetic activity and motion sickness severity. *Psychophysiology*, 40(1):39–44, 2003.

J. J. Gibson. Adaptation, after-effect and contrast in the perception of curved lines. *J. Exp. Psychol.*, 16(1):1, 1933.

J. J. Gibson. The perception of the visual world. 1950.

J. J. Gibson. The senses considered as perceptual systems. 1966.

S. Girod, S. C. Schvartzman, D. Gaudilliere, K. Salisbury, and R. Silva. Haptic feedback improves surgeons' user experience and fracture reduction in facial trauma simulation. *Journal of Rehabilitation Research & Development*, 53(5), 2016.

BIBLIOGRAPHY

J. Golding, M. Finch, and J. Stott. Frequency effect of 0.35-1.0 hz horizontal translational oscillation on motion sickness and the somatogravic illusion. *Aerosp. Med. Hum. Perform.*, 68(5):396–402, 1997.

J. F. Golding. Motion sickness susceptibility. *Autonomic Neuroscience*, 129(1-2):67–76, 2006.

E. J. Gonzalez, P. Abtahi, and S. Follmer. Reach+ extending the reachability of encountered-type haptics devices through dynamic redirection in vr. In *Proc. Symp. User Interface Software and Technology*, pages 236–248, 2020.

S. Grassini and K. Laumann. Are modern head-mounted displays sexist? a systematic review on gender differences in hmd-mediated virtual reality. *Front. Psychol.*, 11, 2020.

T. Grechkin, J. Thomas, M. Azmandian, M. Bolas, and E. S. Rosenberg. Revisiting detection thresholds for redirected walking: Combining translation and curvature gains. In *ACM Symp. Applied Perception*, pages 113–120, 2016.

L. Greenemeier. Virtual reality for all, finally. *Scientific American*, 2015. URL https://www.scientificamerican.com/article/virtual-reality-for-all-finally-video/.

R. L. Gregory. *Eye and brain: The psychology of seeing*, volume 38. Princeton university press, 2015.

C. Groth, J.-P. Tauscher, N. Heesen, S. Castillo, and M. Magnor. Visual techniques to reduce cybersickness in virtual reality. In *IEEE Conf. Virtual Reality and 3D User Interfaces*, pages 486–487, 2021a.

C. Groth, J.-P. Tauscher, N. Heesen, S. Grogorick, S. Castillo, and M. Magnor. Mitigation of cybersickness in immersive 360°videos. In *IEEE VR Workshop on Immersive Sickness Prevention*, pages 169–177, 2021b.

C. Groth, J.-P. Tauscher, N. Heesen, M. Hattenbach, S. Castillo, and M. Magnor. Omnidirectional galvanic vestibular stimulation in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2234–2244, 2022. doi:10.1109/TVCG.2022.3150506.

M. Groß. *Visual Computing*. Springer, 1994.

X. Guo, S. Nakamura, Y. Fujii, T. Seno, and S. Palmisano. Effects of luminance contrast, averaged luminance and spatial frequency on vection. *Experimental Brain Research*, 239:3507–3525, 2021.

M. A. Gutiérrez Alonso, F. Vexo, and D. Thalmann. Stepping into virtual reality. *(No Title)*, 2008.

A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 71 (1):38–53, 1911.

F. Hahlbohm, M. Kappel, J.-P. Tauscher, M. Eisemann, and M. Magnor. Plenoptic-points: Rasterizing neural feature points for high-quality novel view synthesis. In *Proc. Vision, Modeling and Visualization (VMV)*, pages 53–61. Eurographics, 2023. doi:10.2312/vmv.20231226.

G. Halmagyi, I. Curthoys, P. Cremer, C. Henderson, M. Todd, M. Staples, and D. D'cruz. The human horizontal vestibulo-ocular reflex in response to high-acceleration stimulation before and after unilateral vestibular neurectomy. *Experimental brain research*, 81:479–490, 1990.

G. Hammond, T. Thompson, T. Proffitt, and P. Driscoll. Functional significance of the early component of the human blink reflex. *Behavioral neuroscience*, 110(1):7, 1996.

D. T. Han, M. Suhail, and E. D. Ragan. Evaluating remapped physical reach for hand interactions with passive haptics in virtual reality. *IEEE transactions on visualization and computer graphics*, 24(4):1467–1476, 2018.

Y. Han, J. T. Somers, J. I. Kim, A. N. Kumar, and R. J. Leigh. Ocular responses to head rotations during mirror viewing. *Journal of Neurophysiology*, 86(5):2323–2329, 2001.

M. M. Hannuksela, Y.-K. Wang, and A. Hourunranta. An overview of the omaf standard for 360 video. In *Data Compression Conf.*, pages 418–427, 2019.

HaptX. Haptx gloves g1. URL https://g1.haptx.com/.

J. Hartfill, J. Gabel, L. Kruse, S. Schmidt, K. Riebandt, S. Kühn, and F. Steinicke. Analysis of detection thresholds for hand redirection during mid-air interactions in virtual reality. In *ACM Symp. Virtual Reality Software and Technology*, 2021.

J. Hegdé, T. D. Albright, and G. R. Stoner. Second-order motion conveys depth-order information. *Journal of Vision*, 4(10):1–1, 2004.

BIBLIOGRAPHY

M. Heilig. Sensorama simulator. U.S. Patent 3050870, 1962. URL https://patents.google.com/patent/US3050870A/en.

W. Heisenberg. Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Zeitschrift für Physik*, 43:172–198, 1927. doi:10.1007/BF01397280.

S. Hillaire, A. Lécuyer, R. Cozot, and G. Casiez. Depth-of-field blur effects for first-person navigation in virtual environments. *IEEE computer graphics and applications*, 28(6):47–55, 2008.

K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Conf. Human Factors in Computing Systems*, pages 452–458, 1994.

HMD. Starvr one. https://www.starvr.com/. Accessed: 2022-10-14.

P. Hock, S. Benedikter, J. Gugenheimer, and E. Rukzio. Carvr: Enabling in-car virtual reality entertainment. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 4034–4044, 2017.

A. Hollingworth. Visual memory for natural scenes. *Visual memory*, pages 123–162, 2008.

B. K. Horn. Closed-form solution of absolute orientation using unit quaternions. *Josa a*, 4(4):629–642, 1987.

T. S. Horowitz and J. M. Wolfe. Visual search has no memory. *Nature*, 394(6693): 575–577, 1998.

HTC Vive Pro. Htc vive pro eye technical specifications. https://developer.vive.com/resources/knowledgebase/vive-pro-eye-specs-user-guide/.

P. Hu, Q. Sun, P. Didyk, L.-Y. Wei, and A. E. Kaufman. Reducing simulator sickness with perceptual camera control. *ACM Trans. Graph.*, 38(6), 2019.

J.-Y. Huang. An omnidirectional stroll-based virtual reality interface and its application on overhead crane training. *IEEE Transactions on Multimedia*, 5(1):39–51, 2003.

Z. Huang, T. Zhang, W. Heng, B. Shi, and S. Zhou. Rife: Real-time intermediate flow estimation for video frame interpolation. *arXiv preprint arXiv:2011.06294*, 2021.

W. J. Hutchins. *Machine Translation: Past, Present, Future*. 1986.

Y. Imaoka, A. Flury, and E. D. De Bruin. Assessing saccadic eye movements with head-mounted display virtual reality technology. *Frontiers in Psychiatry*, 11:1–19, 2020.

Insta360 Pro. Insta360 pro. https://www.insta360.com/product/insta360-pro. Accessed: 2021-11-26.

J. Irwin. The pathology of sea sickness. *The Lancet*, 118(3039):907–909, 1881.

R. Islam, K. Desai, and J. Quarles. Cybersickness prediction from integrated hmd's sensors: A multimodal deep fusion approach using eye-tracking and head-tracking data. In *IEEE international symposium on mixed and augmented reality (ISMAR)*, pages 31–40, 2021.

ISO. *ISO/IEC 15444-1:2019*, volume 642. Int. Organization for Standardization, 2019.

H. Iwata. The torus treadmill: Realizing locomotion in ves. *IEEE Computer Graphics and Applications*, 19(6):30–35, 1999.

H. Iwata and T. Fujii. Virtual perambulator: a novel interface device for locomotion in virtual environment. In *Proceedings of the IEEE Virtual Reality Annual International Symposium*, pages 60–65, 1996.

H. Iwata, H. Yano, H. Fukushima, and H. Noma. Circulafloor. *IEEE Computer Graphics and Applications*, 25(1):64–67, 2005.

H. Iwata, H. Yano, and M. Tomiyoshi. String walker. In *ACM SIGGRAPH 2007 emerging technologies*. 2007.

J. Jerald. *The VR Book: Human-Centered Design for Virtual Reality*. ACM Books, 2015.

R. K. Jones and D. N. Lee. Why two eyes are better than one: the two views of binocular vision. *Journal of Experimental Psychology: Human Perception and Performance*, 7 (1):30, 1981.

K. Jung, S. Kim, S. Oh, and S. H. Yoon. Hapmotion: motion-to-tactile framework with wearable haptic devices for immersive vr performance experience. *Virtual Reality*, 28 (1):13, 2024.

A. S. Kaplanyan, A. Sochenov, T. Leimkühler, M. Okunev, T. Goodall, and G. Rufo. Deepfovea: neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Transactions on Graphics*, 38(6):1–13, 2019.

BIBLIOGRAPHY

L. Kassler, J. Feasel, M. D. Lewek, F. P. Brooks Jr, and M. C. Whitton. Matching actual treadmill walking speed and visually perceived walking speed in a projection virtual environment. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 161–161, 2010.

R. Kellogg, R. Kennedy, and A. Graybiel. Motion sickness symptomatology of labyrinthine defective and normal subjects during zero gravity maneuvers. *AMRL-TR. Aerospace Medical Research Laboratories (US)*, page 1, 1964.

M. G. Kendall and B. Babington-Smith. On the method of paired comparisons. *Biometrika*, 31:324–345, 1940. doi:10.1093/biomet/31.3-4.324.

R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The Int. Journal of Aviation Psychology*, 3(3):203–220, 1993. doi:10.1207/s15327108ijap0303_3.

R. S. Kennedy, J. Drexler, and R. C. Kennedy. Research in visually induced motion sickness. *Applied ergonomics*, 41(4):494–503, 2010.

B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023.

B. Keshavarz, H. Hecht, and B. Lawson. Visually induced motion sickness: characteristics, causes, and countermeasures. *Handbook of virtual environments: Design, implementation, and applications*, pages 648–697, 2014.

H. G. Kim, H.-T. Lim, S. Lee, and Y. M. Ro. Vrsa net: Vr sickness assessment considering exceptional motion for 360° vr video. *IEEE Trans. Image Process.*, 28(4): 1646–1660, 2018a.

H. K. Kim, J. Park, Y. Choi, and M. Choe. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics*, 69:66–73, 2018b.

J. Kim, S. Palmisano, W. Luu, and S. Iwasaki. Effects of linear visual-vestibular conflict on presence, perceived scene stability and cybersickness in the oculus go and oculus quest. *Front. Virtual Real.*, 2:42, 2021.

S. M. Kim and G. J. Kim. Sparse peripheral counter-vection flow visualization with reverse optical flow for vr sickness reduction. In *International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 798–799, 2022.

Y. M. Kim, I. Rhiu, and M. H. Yun. A systematic review of a virtual reality system from the perspective of user experience. *International Journal of Human–Computer Interaction*, 36(10):893–910, 2020.

Y. Y. Kim, H. J. Kim, E. N. Kim, H. D. Ko, and H. T. Kim. Characteristic changes in the physiological components of cybersickness. *Psychophysiology*, 42(5):616–625, 2005.

F. A. Kingdom and N. Prins. Adaptive methods. In *Psychophysics-A Practical Introduction*, pages 119–148. 2016.

S. A. Klein. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63(8):1421–1455, 2001.

R. Klinke. *Gleichgewichtssinn, Hören, Sprechen*, pages 291–319. 1993. doi:10.1007/978-3-662-09336-8_12.

L. Kohli, M. C. Whitton, and F. P. Brooks. Redirected touching: The effect of warping space on task performance. In *IEEE Symp. 3D User Interfaces*, pages 105–112, 2012.

H. Kolb, R. F. Nelson, P. K. Ahnelt, I. Ortuño-Lizarán, and N. Cuenca. The architecture of the human fovea. *Webvision: The Organization of the Retina and Visual System*, 2020.

O. V. Komogortsev and A. Karpov. Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior research methods*, 45:203–215, 2013.

G. Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, 22(7):1912–1931, 2015.

E. Kugelberg. Facial reflexes. *Brain*, 75(3):385–396, 1952.

I. Ladeira, G. Marsden, and L. Green. Designing interactive storytelling: A virtual environment for personal experience narratives. In *Proceedings of Human-Computer Interaction (INTERACT)*, pages 430–437, 2011.

T. D. Lamb, S. P. Collin, and E. N. Pugh. Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nature Reviews Neuroscience*, 8(12):960–976, 2007.

BIBLIOGRAPHY

M. F. Land and R. D. Fernald. The evolution of eyes. *Annual review of neuroscience*, 15(1):1–29, 1992.

E. Langbehn, F. Steinicke, M. Lappe, G. F. Welch, and G. Bruder. In the blink of an eye: Leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Trans. Graph.*, 37(4), 2018.

E. Langbehn, F. Steinicke, P. Koo-Poeggel, L. Marshall, and G. Bruder. Stimulating the brain in vr: Effects of transcranial direct-current stimulation on redirected walking. In *ACM Symposium on Applied Perception*, 2019a.

E. Langbehn, F. Steinicke, P. Koo-Poeggel, L. Marshall, and G. Bruder. Stimulating the brain in vr: Effects of transcranial direct-current stimulation on redirected walking. In *ACM Symp. Applied Perception*, 2019b.

J. J. LaViola. A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32 (1):47–56, 2000. doi:10.1145/333329.333344.

J. J. LaViola Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, 2000.

A. Lawther and M. Griffin. A survey of the occurrence of motion sickness amongst passengers at sea. *Aviation, space, and environmental medicine*, 59(5):399–406, 1988.

D. Le Gall and A. Tabatabai. Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques. In *Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 761–764, 1988.

R. J. Leigh and D. S. Zee. *The Neurology of Eye Movements*. Contemporary neurology series. Oxford University Press, 2015.

Y. Leng, C.-C. Chen, Q. Sun, J. Huang, and Y. Zhu. Energy-efficient video processing for virtual reality. In *Int. Symp. Computer Architecture*, pages 91–103, 2019.

D. T. Levin and D. J. Simons. Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, 4:501–506, 1997.

D. Li, R. Du, A. Babu, C. D. Brumar, and A. Varshney. A log-rectilinear transformation for foveated 360-degree video streaming. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2638–2647, 2021.

K. Liao, M. F. Walker, A. Joshi, M. Reschke, M. Strupp, and R. J. Leigh. The human vertical translational vestibulo-ocular reflex: Normal and abnormal responses. *Annals of the New York Academy of Sciences*, 1164(1):68–75, 2009.

K. Lim, J. Lee, K. Won, N. Kala, and T. Lee. A novel method for vr sickness reduction based on dynamic field of view processing. *Virtual Reality*, pages 1–10, 2020.

J.-W. Lin, H. B.-L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proc. IEEE Virtual Reality*, pages 164–171. IEEE, 2002.

Y.-C. Lin, Y.-J. Chang, H.-N. Hu, H.-T. Cheng, C.-W. Huang, and M. Sun. Tell me where to look: Investigating ways for assisting focus in 360 video. In *Conf. Human Factors in Computing Systems*, pages 2535–2545, 2017.

Y.-X. Lin, R. Venkatakrishnan, R. Venkatakrishnan, E. Ebrahimi, W.-C. Lin, and S. V. Babu. How the presence and size of static peripheral blur affects cybersickness in virtual reality. *ACM Transactions on Applied Perception*, 17(4), 2020a. doi:10.1145/3419984.

Y.-X. Lin, R. Venkatakrishnan, R. Venkatakrishnan, E. Ebrahimi, W.-C. Lin, and S. V. Babu. How the presence and size of static peripheral blur affects cybersickness in virtual reality. *ACM Trans. Appl. Percept.*, 17(4), 2020b.

Linden ResearchF. Second life. URL https://secondlife.com/.

Y.-H. Liu, H.-Y. Su, H.-C. Lin, C.-Y. Li, Y.-H. Wang, and C.-H. Huang. Investigation of achieving ultrasonic haptic feedback using piezoelectric micromachined ultrasonic transducer. *Electronics*, 11(14):2131, 2022.

M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, 3(2):JCMC321, 1997.

R. Lou, R. H. Y. So, and D. Bechmann. Geometric deformation for reducing optic flow and cybersickness dose value in vr. In *Eurographics 2022 - Posters*, 2022.

E. Loup-Escande, E. Jamet, M. Ragot, S. Erhel, and N. Michinov. Effects of stereoscopic display on learning and user experience in an educational virtual environment. *International Journal of Human–Computer Interaction*, 33(2):115–122, 2017.

BIBLIOGRAPHY

Z.-L. Lu and G. Sperling. Three-systems theory of human visual motion perception: review and update. *Journal of the Optical Society of America A*, 18(9):2331–2370, 2001.

W. Luu, B. Zangerl, M. Kalloniatis, and J. Kim. Effects of stereopsis on vection, presence and cybersickness in head-mounted display (hmd) virtual reality. *Scientific reports*, 11(1):12373, 2021.

J. Lylykangas, M. Ilves, H. Venesvirta, V. Rantanen, E. Mäkelä, A. Vehkaoja, J. Verho, J. Lekkala, M. Rautiainen, and V. Surakka. Artificial eye blink pacemaker-a first investigation into the blink production using constant-interval electrical stimulation. In *Proc. European Medical and Biological Engineering Conf.*, pages 522–525, 2018.

J. Lylykangas, M. Ilves, H. Venesvirta, V. Rantanen, E. Mäkelä, A. Vehkaoja, J. Verho, J. Lekkala, M. Rautiainen, and V. Surakka. Electrical stimulation of eye blink in individuals with dry eye symptoms caused by chronic unilateral facial palsy. In *Proc. Int. Conf. Medical and Biological Engineering*, pages 7–11, 2020.

T. Maeda, H. Ando, and M. Sugimoto. Virtual acceleration with galvanic vestibular stimulation in a virtual reality environment. In *IEEE Conf. of Virtual Reality*, pages 289–290, 2005.

A. T. Maereg, A. Nagar, D. Reid, and E. L. Secco. Wearable vibrotactile haptic device for stiffness discrimination during virtual interactions. *Frontiers in Robotics and AI*, 4:42, 2017.

K. A. Manning, L. A. Riggs, and J. K. Komenda. Reflex eyeblinks and visual suppression. *Perception & Psychophysics*, 34(3):250–256, 1983.

R. K. Mantiuk, M. Ashraf, and A. Chapiro. stelacsf: a unified model of contrast sensitivity as the function of spatio-temporal frequency, eccentricity, luminance and area. *ACM Transactions on Graphics (TOG)*, 41(4):1–16, 2022.

M. W. Marcellin, M. J. Gormish, A. Bilgin, and M. P. Boliek. An overview of jpeg-2000. In *Proc. Data Compression Conf.*, pages 523–541. IEEE, 2000.

D. Martin, X. Sun, D. Gutierrez, and B. Masia. A study of change blindness in immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 29(5): 2446–2455, 2023.

N. Martin, N. Mathieu, N. Pallamin, M. Ragot, and J.-M. Diverrez. Automatic recognition of Virtual Reality sickness based on physiological signals. In *IBC*, Amsterdam, Netherlands, 2018. URL https://hal.archives-ouvertes.fr/hal-02284832.

S. Martinez-Conde, S. L. Macknik, X. G. Troncoso, and D. H. Hubel. Microsaccades: a neurophysiological analysis. *Trends in neurosciences*, 32(9):463–475, 2009.

J. Mateer. Directing for cinematic virtual reality: how the traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice*, 18(1):14–25, 2017. doi:10.1080/14682753.2017.1305838.

G. Maus, Z. Chen, and R. Denison. Illusory occlusion can trump binocular disparity. *Journal of Vision*, 16(12):837–837, 2016.

A. Mazloumi Gavgani, F. R. Walker, D. M. Hodgson, and E. Nalivaiko. A comparative study of cybersickness during exposure to virtual reality and "classic" motion sickness: are they different? *Journal of Applied Physiology*, 125(6):1670–1680, 2018.

D. McDonnall, K. S. Guillory, and M. D. Gossman. Restoration of blink in facial paralysis patients using fes. In *Int. Conf. Neural Engineering*, pages 76–79, 2009.

M. McGill, A. Ng, and S. Brewster. I am the passenger: how visual motion cues can influence sickness for in-car vr. In *Conf. Human Factors in Computing Systems*, pages 5655–5668, 2017.

R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics*, 18(4):626–633, 2012.

E. Medina, R. Fruland, and S. Weghorst. Virtusphere: Walking in a human size vr "hamster ball". In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 52, pages 2102–2106, 2008.

J. L. Meiry. *The vestibular system and human dynamic space orientation.* PhD thesis, Massachusetts Institute of Technology, 1965.

Microsoft. Microsoft hololens 2. URL https://www.microsoft.com/de-de/hololens.

F. Miles. The neural processing of 3-d visual information: evidence from eye movements. *European journal of neuroscience*, 10(3):811–822, 1998.

BIBLIOGRAPHY

P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. In *IEICE Transactions on Information and Systems*, 1994.

P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–292. Spie, 1995.

S. R. Mitroff, D. J. Simons, and D. T. Levin. Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & psychophysics*, 66:1268–1281, 2004.

K. R. Moghadam, C. Banigan, and E. D. Ragan. Scene transitions and teleportation in virtual reality and the implications for spatial awareness and sickness. *IEEE Transactions on Visualization and Computer Graphics*, 2018.

S. T. Moore, V. Dilda, and H. G. MacDougall. Galvanic vestibular stimulation as an analogue of spatial disorientation after spaceflight. *Aviation, space, and environmental medicine*, 82(5):535–542, 2011.

M. Mühlhausen, M. Kappel, M. Kassubeck, P. M. Bittner, S. Castillo, and M. Magnor. Temporal consistent motion parallax for omnidirectional stereo panorama video. In *ACM Symp. Virtual Reality Software and Technology*, 2020. doi:10.1145/3385956.3418965.

M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP*, 2(331-340):2, 2009.

J. Munafo, M. Diedrick, and T. A. Stoffregen. The virtual reality head-mounted display oculus rift induces motion sickness and is sexist in its effects. *Experimental brain research*, 235(3):889–901, 2017.

R. A. M. Murillo, S. Subramanian, and D. M. Plasencia. Erg-o: Ergonomic optimization of immersive virtual environments. In *Proc. Symp. User Interface Software and Technology*, 2017.

B. Murovec, J. Spaniol, J. L. Campos, and B. Keshavarz. Multisensory effects on illusory self-motion (vection): The role of visual, auditory, and tactile cues. *Multisensory Research*, 34(8):869–890, 2021.

D. Mustafi, A. H. Engel, and K. Palczewski. Structure of cone photoreceptors. *Progress in retinal and eye research*, 28(4):289–302, 2009.

T. Nakano, M. Kato, Y. Morito, S. Itoi, and S. Kitazawa. Blink-related momentary activation of the default mode network while viewing videos. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2):702–706, 2013.

K. Nakayama and G. H. Silverman. The aperture problem—ii. spatial integration of velocity information along contours. *Vision Research*, 28(6):747–753, 1988.

D. G. Narciso, M. Bessa, M. C. Melo, A. Coelho, and J. Vasconcelos-Raposo. Immersive 360° video user experience: impact of different variables in the sense of presence and cybersickness. *Univers. Access Inf. Soc.*, 2019.

NASA Ames Research Center. Virtual environment workstation (view) project, 1985. URL https://human-factors.arc.nasa.gov/index.php.

G. Neely, G. Ljunggren, C. Sylven, and G. Borg. Comparison between the visual analogue scale (vas) and the category ratio scale (cr-10) for the evaluation of leg exertion. *Int. J. Sports Med.*, 13(02):133–136, 1992.

A. Nguyen and A. Kunz. Discrete scene rotation during blinks and its effect on redirected walking algorithms. In *ACM Symp. Virtual Reality Software and Technology*, pages 1–10, 2018.

S. Nichols and H. Patel. Health and safety implications of virtual reality: a review of empirical evidence. *Applied ergonomics*, 33(3):251–271, 2002.

G. Nie, Y. Liu, and Y. Wang. Prevention of visually induced motion sickness based on dynamic real-time content-aware non-salient area blurring. In *IEEE Int. Symp. Mixed and Augmented Reality Adjunct*, pages 75–78, 2017.

G.-Y. Nie, H. B.-L. Duh, Y. Liu, and Y. Wang. Analysis on mitigation of visually induced motion sickness by applying dynamical blurring on a user's retina. *TVCG*, 2019.

Nintendo Co., Ltd. Virtual boy, 1995.

M. A. Nitsche, P. S. Boggio, F. Fregni, and A. Pascual-Leone. Treatment of depression with transcranial direct current stimulation (tdcs): a review. *Experimental neurology*, 219(1):14–19, 2009.

H. Noma and T. Miyasato. Design for locomotion interface in a large scale virtual environment atlas: Atr locomotion interface for active self motion. In *ASME In-*

ternational Mechanical Engineering Congress and Exposition, volume 15861, pages 111–118, 1998.

N. Norouzi, G. Bruder, and G. Welch. Assessing vignetting as a means to reduce vr sickness during amplified head rotations. In *ACM Symp. Applied Perception*, pages 1–8, 2018.

OculusFFR. Oculus quest and go: Fixed foveated rendering. https://developer.oculus.com/documentation/unreal/unreal-ffr/. Accessed: 2021-11-14.

N. Ogawa, T. Narumi, and M. Hirose. Effect of avatar appearance on detection thresholds for remapped hand movements. *IEEE transactions on visualization and computer graphics*, 27(7):3182–3197, 2020.

J. F. O'Hanlon, M. E. McCauley, et al. Motion sickness incidence as a function of the frequency and acceleration of vertical sinusoidal motion. *Aerospace medicine*, 45(4): 366–369, 1974.

M. S. Ollonazarovich. The latest advancements in virtual reality therapy for mental health conditions. *Scientific trends in medicine*, 1(1):2–4, 2024.

M. Orchard and G. Sullivan. Overlapped block motion compensation: an estimation-theoretic approach. *IEEE Trans. Image Process.*, 3(5):693–699, 1994.

J. K. O'regan and A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5):939–973, 2001.

D. Osorio and M. Vorobyev. Colour vision as an adaptation to frugivory in primates. *Proceedings: Biological Sciences*, 263(1370):593–599, 1996.

S. E. Palmer. *Vision science: Photons to phenomenology*. MIT press, 1999.

S. H. Park, B. Han, and G. J. Kim. Mixing in reverse optical flow to mitigate vection and simulation sickness in virtual reality. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, 2022.

A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graph.*, 35(6), Nov. 2016. ISSN 0730-0301. doi:10.1145/2980179.2980246.

W. Powell, B. Stevens, S. Hand, and M. Simmons. Blurring the boundaries: The perception of visual gain in treadmill-mediated virtual environments. In *IEEE VR workshop on perceptual illusions in virtual environments*, 2011.

P. Preechayasomboon and E. Rombokas. Haplets: Finger-worn wireless and low-encumbrance vibrotactile haptic feedback for virtual and augmented reality. *Frontiers in Virtual Reality*, 2, 2021.

D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. Lamantia, J. O. McNamara, and S. M. Williams. *Neuroscience, 3rd Edition.* 2004.

S. Ramat and D. S. Zee. Binocular coordination in fore/aft motion. *Annals of the New York Academy of Sciences*, 1039(1):36–53, 2005.

V. Rantanen, A. Vehkaoja, J. Verho, P. Veselỳ, J. Lylykangas, M. Ilves, E. Mäkelä, M. Rautiainen, V. Surakka, and J. Lekkala. Prosthetic pacing device for unilateral facial paralysis. In *Mediterranean Conf. Medical and Biological Engineering and Computing*, pages 653–658, 2016.

J. T. Reason. Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine*, 71(11):819–829, 1978.

J. T. Reason and J. J. Brand. *Motion sickness.* Academic press, 1975.

L. Rebenitsch and C. Owen. Review on cybersickness in applications and visual displays. *Virtual Reality*, 20(2):101–125, 2016.

R. J. Reed-Jones, J. G. Reed-Jones, L. M. Trick, and L. A. Vallis. Can galvanic vestibular stimulation reduce simulator adaptation syndrome? 2007.

R. A. Rensink. Change detection. *Annual review of psychology*, 53(1):245–277, 2002.

R. A. Rensink, J. K. O'regan, and J. J. Clark. To see or not to see: The need for attention to perceive changes in scenes. *Psychological science*, 8(5):368–373, 1997.

H. Rheingold. *Virtual reality: exploring the brave new technologies.* Simon & Schuster Adult Publishing Group, 1991.

G. E. Riccio and T. A. Stoffregen. An ecological theory of motion sickness and postural instability. *Ecological psychology*, 3(3):195–240, 1991.

B. E. Riecke. Compelling self-motion through virtual environments without actual self-motion: using self-motion illusions ("vection") to improve user experience in vr. In J.-J. Kim, editor, *Virtual reality*, volume 8, pages 149–178. 2011.

BIBLIOGRAPHY

M. Rietzler, F. Geiselhart, J. Gugenheimer, and E. Rukzio. Breaking the tracking: Enabling weight perception using perceivable tracking offsets. In *Conf. Human Factors in Computing Systems*, pages 1–12, 2018.

J. G. Robson. Spatial and temporal contrast-sensitivity functions of the visual system. *J. Opt. Soc.*, 56(8):1141–1142, 1966.

Y. Ruan, S. Jiang, and A. Gericke. Age-related macular degeneration: role of oxidative stress and blood vessels. *International journal of molecular sciences*, 22(3):1296, 2021.

P. Rubin. The inside story of oculus rift and how virtual reality became reality. *Wired*, 2015. ISSN 1059-1028. URL https://www.wired.com/2014/05/oculus-rift-4/.

M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir. A comparative study of image retargeting. *ACM Trans. Gr.*, 29(5), 2010.

G. Rushworth. Observations on blink reflexes. *Journal of neurology, neurosurgery, and psychiatry*, 25(2):93, 1962.

D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proc. symposium on Eye tracking research & applications*, pages 71–78, 2000.

M. Samad, E. Gatti, A. Hermes, H. Benko, and C. Parise. Pseudo-haptic weight: Changing the perceived weight of virtual objects by manipulating control-display ratio. In *Conf. Human Factors in Computing Systems*, pages 1–13, 2019.

W. Säring and D. Von Cramon. The acoustic blink reflex: Stimulus dependence, excitability and localizing value. *Journal of Neurology*, 224:243–252, 1981.

R. Schmid and F. Lardini. On the predominance of anti-compensatory eye movements in vestibular nystagmus. *Biological Cybernetics*, 23(3):135–148, 1976.

M. J. Schuemie, P. van der Straaten, M. Krijn, and C. A. van der Mast. Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, 4(2):183–201, 2001. doi:10.1089/109493101300117884.

R. M. Schumacher. *Handbook of Global User Research.* 2010. doi:https://doi.org/10.1016/B978-0-12-374852-2.00001-X.

S. Schwartz, N. Gomel, A. Loewenstein, and A. Barak. Use of a novel beyeonics one three-dimensional head-mounted digital visualization platform in vitreoretinal surgeries. *European Journal of Ophthalmology*, page 11206721241229115, 2024.

A. F. Seay, D. M. Krum, L. Hodges, and W. Ribarsky. Simulator sickness and presence in a high fov virtual environment. In *Proc. IEEE Virtual Reality*, pages 299–300. IEEE, 2001.

A. R. See, J. A. G. Choco, and K. Chandramohan. Touch, texture and haptic feedback: a review on how we feel the world around us. *Applied Sciences*, 12(9), 2022.

T. Seno, H. Ito, and S. Sunaga. The object and background hypothesis for vection. *Vision research*, 49(24):2973–2982, 2009.

SenseGlove. Senseglove nova 2. URL https://www.senseglove.com/product/nova-2/.

I. Setyawan and R. L. Lagendijk. Human perception of geometric distortions in images. In *Security, Steganography, and Watermarking of Multimedia Contents VI*, volume 5306, pages 256 – 267, 2004. doi:10.1117/12.526726. URL https://doi.org/10.1117/12.526726.

C. Sherman. Motion sickness: review of causes and preventive strategies. *Journal of travel medicine*, 9(5):251–256, 2002.

K. Sheykholeslami and K. Kaga. The otolithic organ as a receptor of vestibular hearing revealed by vestibular-evoked myogenic potentials in patients with inner ear anomalies. *Hearing research*, 165(1-2):62–67, 2002.

D. Shin. Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior*, 78:64–73, 2018. doi:https://doi.org/10.1016/j.chb.2017.09.012.

S. Siegel and J. Castellan, N. J. *Nonparametric statistics for the behavioral sciences*. Mcgraw-Hill Book Company, 2. edition, 1988.

L. D. Silverstein. Foundations of vision. *Color Research and Application*, 21:142–144, 2008.

D. J. Simons and D. T. Levin. Failure to detect changes to people during a real-world interaction. *Psychon. Bull. Rev.*, 5:644–649, 1998a.

D. J. Simons and D. T. Levin. Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, 5:644–649, 1998b.

A. Singla, S. Göring, D. Keller, R. R. Ramachandra Rao, S. Fremerey, and A. Raake. Assessment of the simulator sickness questionnaire for omnidirectional videos. In *IEEE Conf. Virtual Reality and 3D User Interfaces*, pages 198–206, 2021.

BIBLIOGRAPHY

M. Slater and S. Wilbur. *A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments.* Massachusetts Institute of Technology, 1997.

M. Slater, J. McCarthy, and F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human factors*, 40(3):469–477, 1998.

B. J. Snow and R. W. Frith. The relationship of eyelid movement to the blink reflex. *Journal of the Neurological Sciences*, 91(1-2):179–189, 1989.

B. Song, P. Wen, T. Ahfock, and Y. Li. Numeric investigation of brain tumor influence on the current distributions during transcranial direct current stimulation. *IEEE Trans. on biomedical engineering*, 63(1):176–187, 2015.

Z. Song, X. Zhang, X. Xu, J. Dong, W. Li, Y.-K. Jan, and F. Pu. The effects of immersion and visuo-tactile stimulation on motor imagery in stroke patients are related to the sense of ownership. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

J. L. Souman, P. R. Giordano, M. Schwaiger, I. Frissen, T. Thümmel, H. Ulbrich, A. D. Luca, H. H. Bülthoff, and M. O. Ernst. Cyberwalk: Enabling unconstrained omnidirectional walking through virtual environments. *ACM Transactions on Applied Perception (TAP)*, 8(4):1–22, 2011.

G. Sperling. Movement perception in computer-driven visual displays. *Behavior Research Methods & Instrumentation*, 8:144–151, 1976.

J. Spillmann, S. Tuchschmid, and M. Harders. Adaptive space warping to enhance passive haptics in an arthroscopy surgical simulator. *IEEE transactions on visualization and computer graphics*, 19(4):626–633, 2013.

M. Sra. Steering locomotion by vestibular perturbation in room-scale vr. In *IEEE Virtual Reality*, pages 405–406, 2017.

M. Sra, X. Xu, and P. Maes. Galvr: a novel collaboration interface using gvs. In *ACM Symp. Virtual Reality Software and Technology*, pages 1–2, 2017.

M. Sra, A. Jain, and P. Maes. Adding proprioceptive feedback to virtual reality experiences using galvanic vestibular stimulation. In *Conf. Human Factors in Computing Systems*, pages 1–14, 2019.

K. K. Sreedhar, A. Aminlou, M. M. Hannuksela, and M. Gabbouj. Viewport-adaptive encoding and streaming of 360-degree video for virtual reality applications. In *Int. Symp. Multimedia*, pages 583–586, 2016.

K. M. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is not simulator sickness. *Proc. Human Factors and Ergonomics Society Annual Meeting*, 41(2):1138–1142, 1997. doi:10.1177/107118139704100292.

K. M. Stanney, K. S. Hale, I. Nahmens, and R. S. Kennedy. What to expect from immersive virtual environment exposure: Influences of gender, body mass index, and past experience. *Human Factors*, 45(3):504–520, 2003.

J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cybersickness: impact, causes, and measures. a review. *Front. Virtual Real.*, 1:1–10, 2020.

Q. Sun, A. Patney, L.-Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. McGuire, D. Luebke, and A. Kaufman. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Transactions on Graphics*, 37(4):1–13, 2018.

Q. Sun, A. Taherin, Y. Siatitse, and Y. Zhu. Energy-efficient 360-degree video rendering on fpga via algorithm-architecture co-design. In *Int. Symp. Field-Programmable Gate Arrays*, pages 97–103, 2020.

I. E. Sutherland. A head-mounted three dimensional display. *AFIPS Conference Proceedings*, 33(1):757–764, 1968.

I. E. Sutherland et al. The ultimate display. In *Proceedings of the IFIP Congress*, volume 2, pages 506–508, 1965.

D. Swapp, J. Williams, and A. Steed. The implementation of a novel walking interface within an immersive display. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 71–74, 2010.

J. Takatalo, G. Nyman, and L. Laaksonen. Components of human experience in virtual environments. *Computers in Human Behavior*, 24(1):1–15, 2008. doi:https://doi.org/10.1016/j.chb.2006.11.003.

D. Taubman and M. Marcellin. *JPEG2000 image compression fundamentals, standards and practice*, volume 642. Springer Science & Business Media, 2012.

BIBLIOGRAPHY

J.-P. Tauscher, A. Witt, S. Bosse, F. W. Schottky, S. Grogorick, S. Castillo, and M. Magnor. Exploring neural and peripheral physiological correlates of simulator sickness. *Computer Animation and Virtual Worlds*, 2020.

A. C. ter Horst, M. Koppen, L. P. J. Selen, and W. P. Medendorp. Reliability-based weighting of visual and vestibular cues in displacement estimation. *PloS one*, 10(12): 1–15, 2015.

BBC Research. Dirac specification (version 2.2.3). https://web.archive.org/web/20150503015104/http://diracvideo.org/download/specification/dirac-spec-latest.pdf, 2008.

J. Thomas and E. S. Rosenberg. Reactive alignment of virtual and physical environments using redirected walking. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 317–323, 2020. doi:10.1109/VRW50115.2020.00071.

B. Thompson, B. C. Hansen, R. F. Hess, and N. F. Troje. Peripheral vision: Good for biological motion, bad for signal noise segregation? *Journal of Vision*, 7(10):1–12, 2007.

W. Thompson, R. Fleming, S. Creem-Regehr, and J. K. Stefanucci. *Visual Perception from a Computer Graphics Perspective*. CRC Press, 2011.

M. Thüring and S. Mahlke. Usability, aesthetics and emotions in human–technology interaction. *International journal of psychology*, 42(4):253–264, 2007.

P. Tognini, I. Manno, J. Bonaccorsi, M. C. Cenni, A. Sale, and L. Maffei. Environmental enrichment promotes plasticity and visual acuity recovery in adult monocular amblyopic rats. *PloS one*, 7(4), 2012.

M. Treisman. Motion sickness: an evolutionary hypothesis. *Science*, 197(4302):493–495, 1977.

A. Tse, C. Jennett, J. Moore, Z. Watson, J. Rigby, and A. L. Cox. Was i there? impact of platform and headphones on 360 video immersion. In *Conf. Human Factors in Computing Systems*, pages 2967–2974, 2017.

C. Tursun and P. Didyk. Perceptual visibility model for temporal contrast changes in periphery. *ACM Transactions on Graphics*, 42(2):1–16, 2022.

O. T. Tursun, E. Arabadzhiyska-Koleva, M. Wernikowski, R. Mantiuk, H.-P. Seidel, K. Myszkowski, and P. Didyk. Luminance-contrast-aware foveated rendering. *ACM Trans. Gr.*, 38(4):1–14, 2019.

UnityVRDesign. Unity vr best practice. https://learn.unity.com/tutorial/vr-best-practice. Accessed: 2021-11-14.

Unreal Engine Virtual Reality. Unreal engine virtual reality best practices. https://docs.unrealengine.com/en-US/Platforms/VR/DevelopVR/ContentSetup/index.html. Accessed: 2021-11-14.

M. Unser and T. Blu. Mathematical properties of the jpeg2000 wavelet filters. *IEEE Trans. Image Process.*, 12(9):1080–1090, 2003. doi:10.1109/TIP.2003.812329.

M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr. Walking> walking-in-place> flying, in virtual environments. In *Annual Conf. Computer graphics and interactive techniques*, pages 359–364, 1999.

M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503, 2000.

K. S. Utz, V. Dimova, K. Oppenländer, and G. Kerkhoff. Electrified minds: Transcranial direct current stimulation (tdcs) and galvanic vestibular stimulation (gvs) as methods of non-invasive brain stimulation in neuropsychology—a review of current data and future implications. *Neuropsychologia*, 48(10):2789–2810, 2010.

F. VanderWerf, P. Brassinga, D. Reits, M. Aramideh, and B. Ongerboer de Visser. Eyelid movements: behavioral studies of blinking in humans under different stimulus conditions. *Journal of Neurophysiology*, 89(5):2784–2796, 2003.

M. Veit, A. Capobianco, and D. Bechmann. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pages 51–58, 2009. doi:10.1145/1643928.1643942.

Vestibulator. Good vibrations engineering: Gvs. https://goodvibrationsengineering.com/GVS.html. Accessed: 2023-06-08.

E. Viirre, D. Tweed, K. Milner, and T. Vilis. A reexamination of the gain of the vestibuloocular reflex. *Journal of neurophysiology*, 56(2):439–450, 1986.

BIBLIOGRAPHY

J. Vora, S. Nair, A. K. Gramopadhye, A. T. Duchowski, B. J. Melloy, and B. Kanki. Using virtual reality technology for aircraft visual inspection training: presence and comparison studies. *Applied ergonomics*, 33(6):559–570, 2002.

A. D. Walker, E. R. Muth, F. S. Switzer, and A. Hoover. Head movements and simulator sickness generated by a virtual environment. *Aviation, space, and environmental medicine*, 81(10):929–934, 2010.

B. Walther-Franks, D. Wenig, J. Smeddinck, and R. Malaka. Suspended walking: A physical locomotion interface for virtual reality. In *International Conference on Entertainment Computing*, pages 185–188, 2013.

B. A. Wandell. *Foundations of vision.* 1995.

Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. doi:10.1109/TIP.2003.819861.

Z. Wang, Y. Wang, S. Yan, Z. Zhu, K. Zhang, and H. Wei. Redirected walking on omnidirectional treadmill. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–14, 2023. doi:10.1109/TVCG.2023.3244359.

D. L. Wardman and R. C. Fitzpatrick. *What Does Galvanic Vestibular Stimulation Stimulate?*, pages 119–128. 2002. doi:10.1007/978-1-4615-0713-0_15.

D. L. Wardman, J. L. Taylor, and R. C. Fitzpatrick. Effects of galvanic vestibular stimulation on human posture and perception while standing. *The Journal of Physiology*, 551(3):1033–1042, 2003. doi:https://doi.org/10.1111/j.1469-7793.2003.01033.x.

A. B. Watson and A. J. Ahumada. Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2(2):322–342, 1985.

A. B. Watson, A. J. Ahumada, and J. E. Farrell. Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A*, 3(3):300–307, 1986.

N. A. Webb and M. J. Griffin. Eye movement, vection, and motion sickness with foveal and peripheral vision. *Aviation, space, and environmental medicine*, 74(6):622–625, 2003.

S. Weech, J. Moon, and N. F. Troje. Influence of bone-conducted vibration on simulator sickness in virtual reality. *PLoS ONE*, 13:1–21, 03 2018.

S. Weech, T. Wall, and M. Barnett-Cowan. Reduction of cybersickness during and immediately following noisy galvanic vestibular stimulation. *Experimental Brain Research*, 238(2):427–437, 2020.

M. Weier, T. Roth, E. Kruijff, A. Hinkenjann, A. Pérard-Gayot, P. Slusallek, and Y. Li. Foveated real-time ray tracing for head-mounted displays. *Computer Graphics Forum*, 35(7):289–298, 2016. doi:10.1111/cgf.13026.

T. Weißker, A. Kunert, B. Fröhlich, and A. Kulik. Spatial updating and simulator sickness during steering and jumping in immersive virtual environments. In *IEEE Conf. Virtual Reality and 3D User Interfaces*, pages 97–104. IEEE, 2018.

D. Weyhe, V. Uslar, F. Weyhe, M. Kaluschke, and G. Zachmann. Immersive anatomy atlas—empirical study investigating the usability of a virtual reality environment as a learning tool for anatomy. *Frontiers in surgery*, 5:73, 2018.

A. Williamson and B. Hoggart. Pain: a review of three commonly used pain rating scales. *J. Clin. Nurs.*, 14(7):798–804, 2005.

Y.-H. Wu, C.-H. Tsai, Y.-H. Wu, Y.-S. Cherng, M.-J. Tai, P. Huang, I.-A. Yao, and C.-L. Yang. Breaking the limits of virtual reality display resolution: the advancements of a 2117-pixels per inch 4k virtual reality liquid crystal display. *Journal of Optical Microsystems*, 3(4), 2023.

J. Yang, J. Lee, and S. G. Lisberger. The interaction of bayesian priors and sensory data and its neural circuit implementation in visually guided movement. *Journal of Neuroscience*, 32(49):17632–17645, 2012.

R. W. Young. Pathophysiology of age-related macular degeneration. *Survey of ophthalmology*, 31(5):291–306, 1987.

J. Zabaleta Jiménez, A. Blasco, A. Fernández-Monge, J. Lizarbe, M. Mainer, T. Esnal, J. Baez, and J. Aldazabal. Using virtual reality to simulate an operating room: Satisfaction and results from the first pilot study for nurses. *British Journal of Surgery*, 111(1), 2024.

A. Zare, A. Aminlou, M. M. Hannuksela, and M. Gabbouj. Hevc-compliant tile-based streaming of panoramic video for virtual reality applications. In *Proc. Int. Conf. Multimedia*, pages 601–605, 2016.

BIBLIOGRAPHY

W. Zeng, S. Daly, and S. Lei. Point-wise extended visual masking for jpeg-2000 image compression. In *International Conference on Image Processing*, volume 1, pages 657–660, 2000.

A. Zenner and A. Krüger. Shifty: A weight-shifting dynamic passive haptic proxy to enhance object perception in virtual reality. *IEEE transactions on visualization and computer graphics*, 23(4):1285–1294, 2017.

A. Zenner and A. Krüger. Shifting & warping: A case for the combined use of dynamic passive haptics and haptic retargeting in vr. In *Adjunct Proc. Symp. User Interface Software and Technology*, pages 1–3, 2020.

A. Zenner, K. P. Regitz, and A. Krüger. Blink-suppressed hand redirection. In *IEEE Conf. Virtual Reality and 3D User Interfaces*, pages 75–84, 2021.

A. Zenner, K. Ullmann, O. Ariza, F. Steinicke, and A. Krüger. Induce a blink of the eye: Evaluating techniques for triggering eye blinks in virtual reality. In *Conf. Human Factors in Computing Systems*, page 1–12, 2023.

R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Computer Vision and Pattern Recognition*, pages 586–595, 2018. doi:10.1109/CVPR.2018.00068.

Y. Zhang, K. Chen, and Q. Sun. Toward optimized vr/ar ergonomics: Modeling and predicting user neck muscle contraction. In *ACM SIGGRAPH Conference Proceedings*, pages 1–12, 2023.

S. Zhao, H. Zhang, S. Bhuyan, C. S. Mishra, Z. Ying, M. T. Kandemir, A. Sivasubramaniam, and C. R. Das. Déjà view: Spatio-temporal compute reuse for energy-efficient 360° vr video streaming. In *Int. Symp. Computer Architecture*, pages 241–253, 2020.

M. Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9):25–32, 2005.

Glossary

Accommodation (Eye) The process by which the eye adjusts its focal length to clearly view objects at different distances, primarily through changes in the shape of the lens.

Apparent motion (Perception) The optical illusion that a stationary object is moving, created by rapid sequences of images that simulate movement, such as in films or animations.

Binocular field (Vision) The area of space that can be viewed simultaneously by both eyes without moving the head, contributing to binocular vision and depth perception.

Binocular occlusion conflict (Virtual Reality) A visual phenomenon that occurs when objects are perceived differently by each eye due to partial or complete occlusion in one eye's view, potentially leading to confusion or discomfort in depth perception within virtual environments.

Binocular vision (Vision) The ability to maintain visual focus on an object with both eyes, creating a single visual image and enabling depth perception.

Cerebral cortex (Neuroscience) The outer layer of neural tissue of the cerebrum in the brain, playing a key role in memory, attention, perception, cognition, awareness, thought, language, and consciousness.

Change blindness (Psychology) The perceptual phenomenon that occurs when a change in a visual stimulus is introduced and the observer does not notice it, often due to attentional limitations.

Chiasma opticum (Neuroscience) An X-shaped structure formed by the crossing of the optic nerves in the brain, which allows visual information from both eyes to be partially integrated by each hemisphere of the brain.

Glossary

Ciliary muscle (Eye) A ring of smooth muscle fibers in the eye's middle layer that controls accommodation by altering the shape of the lens to focus on near or distant objects.

Colliculi superiores (Neuroscience) Paired structures located in the mid-brain that act as a hub for coordinating eye and head movements, as well as processing visual, auditory, and somatosensory data.

Cornea (Eye) The transparent front part of the eye that covers the iris, pupil, and anterior chamber, providing most of the eye's optical power by refracting light entering the eye.

Corpus geniculatum laterale (Neuroscience) A major relay center in the thalamus for the visual pathway, receiving sensory input from the retina and sending outputs to the visual cortex.

Cupula (Anatomy) A gelatinous structure in the semicircular canals of the inner ear that responds to fluid movements caused by head rotations, aiding in the sensation of balance.

Cybersickness (Virtual Reality) A form of motion sickness that occurs from exposure to a virtual environment, evoking symptoms such as dizziness, nausea, and disorientation.

Depth of field (Optics) The range of distance within a subject that appears acceptably sharp in an image. The greater the depth of field, the more of the scene appears in focus.

Direct interaction (User Interface Design) An interaction type in which users engage with elements of a digital system directly through physical or virtual means, such as interactions with virtual object on touchscreens.

Disparity (Vision) The difference in the images seen by the left and right eyes, which is the basis for stereopsis, or depth perception.

Endolymph (Anatomy) The fluid contained within the membranous labyrinth of the inner ear, crucial for transmitting mechanical forces to sensory cells during hearing and balance.

Eye drift (Vision) Slow, involuntary motion of the eyes when attempting to fixate steadily on a point, often followed by corrective microsaccades.

Field of view Angular measure for the extent of the observable world that is seen at any given moment, e.g. within an HMD.

Fixation (Vision) The eye mechanism that maintains the gaze in close proximity to a single point. Fixations are crucial for processing detailed visual information and stabilizing the visual environment.

Flicker fusion (Vision) The frequency at which an intermittent light stimulus is perceived to be completely steady to the average human eye.

Fovea centralis (Eye) A small depression in the retina of the eye where visual acuity is highest. It contains a high density of cone cells and only little rod cells and is optimized for seeing fine details.

Foveation (Vision) The decline in visual acuity from a central part to the outside. In the eye, the highest visual acuity is provided in the fovea centralis and decreases towards the periphery.

Head-mounted display virtual reality (VR) display device worn on the head, with the display panel(s) and appropriate optics attached right in front of the eyes.

Immersion (Virtual Reality) The objective level to which a virtual system maps stimuli onto a user's sensory receptors to give the impression of a comprehensive, realistic experience. Immersion is facilitated by engaging multiple senses.

Indirect interaction (User Interface Design) An interaction type where input devices like a mouse or keyboard are used to control a system, without physical contact with the actual elements being controlled.

Inner ear labyrinth (Anatomy) The complex system of fluid-filled chambers and tubes in the inner ear responsible for sensing gravity, acceleration, and rotational movements, crucial for balance and spatial orientation.

Interaction fidelity (Virtual Reality) The degree to which interactions in a virtual environment accurately simulate interactions in the real world, influencing the user's immersion and experience.

Iris (Eye) The colored part of the eye surrounding the pupil, constructed by muscular structures that control the size of the pupil, regulating the amount of light that enters the eye.

Macula lutea (Eye) An approximately circular region in the central part of the retina, densely packed with cone cells and responsible for high-resolution, color vision in bright light conditions.

Microsaccade (Vision) Small, involuntary eye movements made during visual fixation that help refresh the image on the retina to prevent visual fading and maintain visual acuity.

Mixed reality (Virtual Reality) The merging of real and virtual worlds to produce new environments where physical and digital objects co-exist and interact in real time.

Motion sickness (Psychology) A physiological condition characterized by symptoms such as nausea, dizziness, and vomiting that can occur from discrepancies between visually perceived movement and the vestibular system's sense of motion.

Optic array (Vision) The spatial pattern of light rays, varying in brightness and color, which enter the eyes from different directions and provide visual information about the environment.

Optic nerve (Neuroscience) A paired nerve that transmits visual information from the retina to the brain. It consists of the axons of retinal ganglion cells and forms a part of the central nervous system.

Optical flow (Vision) The pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and the scene, providing important information about the structure and movement of objects.

Otoconia (Anatomy) Also known as ear sand, tiny calcium carbonate crystals within the otolith organs of the inner ear that contribute to the body's sense of gravity and motion.

Otolith organs (Anatomy) Structures within the vestibular system of the inner ear, specifically the utricle and saccule, that detect linear accelerations and head tilts in relation to gravity, aiding in balance and spatial orientation.

Periphery (Vision) The large part of the view that surrounds the fovea. It is the outer area seen by the side of the eye when looking straight ahead.

Presence (Virtual Reality) A term describing the phenomenon of users feeling genuinely immersed in a virtual environment, often experiencing it as if they were actually part of that virtual world. Presence is the subjective experience of immersion.

Quick phase (Vision) The rapid, corrective eye movement that occurs during nystagmus to bring the eye back to the fixation point after slow drifts.

Reality-virtuality continuum (Virtual Reality) A spectrum that encompasses all combinations and variations of real and virtual environments, from completely real environments to fully immersive virtual realities.

Retina (Eye) The light-sensitive layer of tissue at the back of the inner eye that contains the photoreceptors. The processes in the retina convert light into electrical signals sent through the optic nerve to the brain for visual recognition.

Saccade (Neuroscience) Rapid, jerky movement of the eyes between phases of fixation, helping to quickly redirect the point of focus so as to scan different parts of the visual scene.

Sclera (Eye) The white outer layer of the eyeball, made of fibrous tissue that protects the inner components of the eye and provides attachment for the extraocular muscles.

Semicircular canals (Anatomy) Three fluid-filled circular tubes in the inner ear, oriented at roughly right angles to each other, which detect rotational movements of the head.

Sensitivity (Vision) The ability of the visual system to detect changes in light intensity by adjacent receptors, specifically referring to the minimum intensity of a stimulus required to elicit a response.

Sensory adaptation (Neuroscience) The process by which sensory receptors become less sensitive to constant stimuli over time, allowing organisms to focus on changes in their environment rather than constants.

Smooth pursuit (Vision) A type of eye movement that allows the eyes to smoothly follow a moving object, maintaining a stable image on the central retina (fovea) during tracking.

User experience (Virtual Reality) The overall experience of a person in a virtual environment, especially in terms of engagement and comfort.

Utricle (Anatomy) One of the two otolith organs located in the vestibule of the inner ear, which detects changes in horizontal movement and the head's position relative to gravity.

Vection (Perception) The illusion of self-motion produced when an individual is stationary but the visual scene around them moves, typically experienced during the viewing of large moving visual fields.

Vergence (Vision) The simultaneous movement of both eyes in opposite directions to obtain or maintain single binocular vision, typically when an object moves closer or further away.

Vergence-accommodation conflict (Psychology) A common issue in stereoscopic displays where the eye's vergence (focus on depth) and accommodation (focus on an object's distance) responses do not match, often causing visual discomfort or fatigue.

Vestibular system (Anatomy) The system in the inner ear that provides the leading contribution to the sense of balance and spatial orientation for the purpose of coordinating movement with balance.

Vestibulo-ocular reflex (Neuroscience) A reflex aiming to stabilize vision by counter-rotating the eyes when the head moves, maintaining a steady gaze during motion.

Visual acuity (Vision) The clarity or sharpness of vision, measured as the ability to discern patterns of a discrete visual angle.

Visual angle (Vision) The angle a viewed object subtends at the eye, typically describing the size of the object in the visual field, which is influenced by both the actual size of the object and its distance from the observer.

Visual-haptic conflict (Virtual Reality) A disparity between what is visually observed and what is felt (haptic), which can disrupt the user's sense of immersion and can lead to discomfort in virtual environments.

Acronyms

CC	control condition.
CDF	Cohen–Daubechies–Feauveau.
CDT	conservative detection threshold.
CGI	computer-generated imagery.
CRF	constant rate factor.
CS	cybersickness.
CSF	contrast sensitivity function.
DCT	discrete cosine transform.
DoF	degree of freedom.
EDA	electrodermal activity.
EMS	electrical muscle stimulation.
FOV	field of view.
fps	frames per second.
FR	foveated rendering.
FWT	fast wavelet transform.
GT	ground truth.
GVS	galvanic vestibular stimulation.
HDR	high-dynamic-range.
HMD	head-mounted display.

Acronyms

HR	heart rate.
HVS	human visual system.
iFWT	inverse fast wavelet transform.
LGT	LeGall-Tabatabai.
MCTS	motion constrained tile sets.
MPEG	Moving Picture Experts Group.
MSQ	pensacola motion sickness questionnaire.
OBMC	overlapped-block motion compensation.
OMAF	omnidirectional media format.
OVR	oculo-vestibular recoupling.
SEM	standard error of the mean.
SotA	state-of-the-art.
SSQ	simulator sickness questionnaire.
SUS	Slater-Usoh-Steed.
tDCS	transcranial direct current stimulation.
VIMS	visually induced motion sickness.
VR	virtual reality.