



Virtual Reality Experience Design Handbook

Heuristics, Patterns, and Practices for
Immersive Applications

Edited by **Rocío Cifuentes-Albeza, Iñigo Cuiñas,
Itziar Goicoechea, Dorota Kamińska,
Francisco-Julián Martínez-Cano,
Rui Raposo, Mário Vairinhos,
and Grzegorz Zwoliński**



CRC Press
Taylor & Francis Group



Virtual Reality Experience Design Handbook

Designing for virtual reality (VR) goes far beyond adapting flat-screen content for a headset. It requires a deep understanding of how the human body and mind engage with immersive environments – so that experiences feel intuitive, natural, and comfortable.

This book is an essential guide for anyone creating VR content that is not only engaging but also inclusive. It offers clear, hands-on support for designing meaningful interactions in virtual spaces.

Aimed at designers, developers, and product teams, it bridges the gap between traditional 2D design and the challenges of embodied interaction. Moving beyond theory, it provides actionable tools – including a VR-specific heuristic framework, checklists, and templates – to simplify complex design decisions.

Readers will learn how to translate familiar 2D patterns into spatial experiences, integrate accessibility, and use elements such as spatial audio, lighting, and animation to guide attention effectively.

As the VR market continues to grow, this handbook equips professionals with the skills and confidence to build high-quality immersive experiences – whether refining existing products or creating entirely new ones.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Virtual Reality Experience Design Handbook

Heuristics, Patterns, and Practices for Immersive Applications

Edited by

Rocío Cifuentes-Albeza, Iñigo Cuiñas,
Itziar Goicoechea, Dorota Kamińska,
Francisco-Julián Martínez-Cano, Rui Raposo,
Mário Vairinhos, and Grzegorz Zwoliński



CRC Press

Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Designed cover image: Firefly

First edition published 2027
by CRC Press
2385 NW Executive Center Drive, Suite 320, Boca Raton FL 33431

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

© 2027 selection and editorial matter, Rocío Cifuentes-Albeza, Iñigo Cuiñas, Itziar Goicoechea, Dorota Kamińska, Francisco-Julián Martínez-Cano, Rui Raposo, Mário Vairinhos, and Grzegorz Zwoliński; individual chapters, the contributors

The Open Access version of this book, available at HYPERLINK "<http://www.taylorfrancis.com>" www.taylorfrancis.com, has been made available under a Creative Commons Attribution-Non Commercial-No Derivatives (CC-BY-NC-ND) 4.0 license. Funded by Erasmus+ Programme of the European Union (Abraxas - 2023-1-PL01-KA220-HED-000157492).

This work was co-funded by the European Union (Project No. 2023-1-PL01-KA220-HED-000157492). The views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System (FRSE). Neither the European Union nor FRSE can be held responsible for them.

Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, access www.copyright.com or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. For works that are not available on CCC please contact mpkbookspermissions@tandf.co.uk

For Product Safety Concerns and Information please contact our EU representative GPSR@taylorandfrancis.com. Taylor & Francis Verlag GmbH, Kaufingerstraße 24, 80331 München, Germany.

Trademark notice: Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

ISBN: [9781041333746] (hbk)
ISBN: [9781041333739] (pbk)
ISBN: [9781003787495] (ebk)

DOI: 10.1201/9781003787495

Typeset in Times
by codeMantra

Contents

<i>Preface</i>	vii
<i>Organization of This Book</i>	ix
<i>About the Editors</i>	xii
<i>List of Contributors</i>	xiv
Introduction to VR Design: Historical Milestones in VR Design	1
<i>Marta Kuźmicz and Dorota Kamińska</i>	
PART I Foundations of Presence	15
1 Presence and Visual Depth Perception in Virtual Reality	17
<i>Mário Vairinhos, Silvino Martins, Rui Raposo, and Pedro Reisinho</i>	
2 Gestalt and Spatial Composition	34
<i>Rocío Cifuentes-Albeza and Francisco-Julián Martínez-Cano</i>	
3 Scenography, Scale, and Lighting in VR	45
<i>Francisco-Julián Martínez-Cano and Rocío Cifuentes-Albeza</i>	
4 Spatial Audio for VR	54
<i>Rui Raposo, Mário Vairinhos, Pedro Reisinho, and Silvino Martins</i>	
PART II Interaction, Locomotion, and UI	69
5 Interaction Modes	71
<i>Iñigo Cuiñas</i>	
6 Locomotion and Comfort	85
<i>Grzegorz Zwoliński</i>	
7 Animation and Motion in VR	102
<i>Mário Vairinhos, Silvino Martins, Rui Raposo, Pedro Reisinho, and Cátia Silva</i>	
8 UI in VR: Diegetic and World-Locked Interfaces	115
<i>Agnieszka Dubiel</i>	

PART III Narrative, Social, Ethics, and Accessibility	127
9 Narrative and Storytelling in Virtual Reality <i>María Cerdeira</i>	129
10 Social Virtual Reality <i>Manuel Caeiro</i>	140
11 Ethics and Sustainability <i>Itziar Goicoechea</i>	150
12 Accessibility and Inclusive Design <i>Tatiana Santos Gonçalves</i>	164
 PART IV Heuristics, Performance, and Validation	 173
13 VR Heuristics <i>Grzegorz Zwoliński</i>	175
14 Performance and Optimization <i>Grzegorz Zwoliński</i>	192
15 Usability Testing and Iterative Development in VR—Methods for Conducting Usability Testing in VR <i>Anna Laska-Leśniewicz and Dorota Kamińska</i>	201
16 Design Translation Patterns for Virtual Reality <i>Grzegorz Zwoliński</i>	217
 <i>Index</i>	 233

Preface

When we first began designing and developing for virtual reality (VR), the landscape felt incredibly vast – and equally unmapped. As researchers, educators, and practitioners scattered across various European institutions, we each found ourselves confronting the exact same hurdles. Whether we were building immersive training simulations, educational tools, or interactive narratives, we repeatedly bumped into the limits of traditional 2D design. The rules we had relied on for years to build intuitive screen-based interfaces suddenly broke down when the user’s entire body became the interface.

We spent countless hours gathering engine documentation, sifting through academic papers on spatial cognition, and learning through trial and error. It became clear that while the hardware for spatial computing was advancing at breakneck speed, the shared language and standardized methodologies for designing these experiences were lagging.

This book is born out of a profound need for a practical, unified, and actionable framework for VR design. It is the result of years of cross-disciplinary collaboration between teams at Lodz University of Technology, Miguel Hernández University, the University of Aveiro, and the University of Vigo. Through our collective work on numerous European research projects and industry applications, we set out to bridge the gap between abstract academic theory and the daily realities of production.

WHAT THIS BOOK IS ABOUT

Our goal with this handbook is not to offer a philosophical meditation on the nature of reality or immersion. Instead, we have written the manual we wished we had when we started. It is a practical tool kit designed to help you navigate the unique challenges of embodied interaction.

In these pages, you will find our *VR Heuristics 1.0* – a structured framework adapted specifically for spatial environments. We introduce the concept of the “comfort budget,” teaching you how to balance visual fidelity and locomotion with human physiological limits. We break down the translation of familiar 2D user interface (UI) patterns into diegetic, world-locked equivalents that make sense in 3D space. Crucially, we have positioned accessibility and inclusive design as a foundational pillar of the design process.

WHO THIS BOOK IS FOR

If you are a traditional UI/UX designer looking to transition into spatial computing, this book will help translate your existing expertise into the Z-axis. If you are a VR developer tired of building technically sound applications that users find confusing or nauseating, you will find actionable checklists and design patterns here to solve those issues. And if you are an educator or human–computer interaction (HCI) researcher, this text provides a structured, reproducible methodology for teaching and testing in VR usability.

HOW TO USE THIS BOOK

We have structured this handbook to be a working companion. You do not need to read it cover to cover to find value. Jump to Chapter 3 when setting up a new scene; consult the UI translation patterns when building an inventory system; or pull from the usability testing checklists when preparing for QA.

Designing for VR is an immense responsibility. We are no longer just asking users to look at our software; we are asking them to step inside it. We hope this handbook equips you with the knowledge, the vocabulary, and the confidence to create virtual worlds that are not only breathtaking but truly comfortable, intuitive, and welcoming for everyone.

Organization of This Book

Grzegorz Zwoliński

Abraxas



Co-funded by
the European Union



Lodz University
of Technology

Universidade de Vigo



universidade de aveiro
theoria poiesis praxis



UNIVERSITAS
Miguel Hernández

This book has been designed as a handbook whose organization follows directly from the specific characteristics of immersive experience design and from the practical educational and design-related needs observed both in academic settings and in the development of virtual reality (VR) applications. Its structure is not based on technological logic, nor does it follow the sequence typically found in manuals focused on individual software tools. Instead, it is built on the assumption that designing VR environments first requires an organized understanding of perception, space, movement, interaction, and the mechanisms that influence user comfort, because these factors shape the user experience long before the choice of a particular technological platform becomes relevant.

This book emerged from the work of an international team collaborating within the project Abraxas – Best Practices for High-Immersion Experiences Design, funded under the Erasmus+ program in the Higher Education sector (Project No. 2023-1-PL01-KA220-HED-000157492). The project has been carried out from December 1, 2023, until November 30, 2026, by a consortium of four universities: Lodz University of Technology, Universidad Miguel Hernández de Elche, University of Vigo, and Universidade de Aveiro. The main objective of the project is to enrich the curricula of all partner institutions through the implementation of an innovative course for 3D experience designers supported by VR-based exercises. Its broader goal is to develop the competencies of future immersive environment designers not only by introducing production tools, but above all by building an understanding of what makes a spatial environment functional, credible, comfortable, and accessible for users with diverse needs.

The educational nature of the project had a direct impact on the organization of this book. Because the material has been developed in parallel with international teaching resources, it was necessary to adopt a structure that would remain relevant regardless of changes in game engines, software libraries, or hardware platforms. In VR, the rapid obsolescence of technical content is particularly evident. Guidelines tied to a specific software version or Software Development Kit (SDK) become outdated much faster than the core design principles related to perception, movement, spatial orientation, or interaction ergonomics. For this reason, this book was intentionally not conceived as a manual explaining how to build a specific system in Unity or how to implement interaction in Unreal Engine. Such information is widely available in software documentation and changes rapidly. The aim of this publication is instead to provide knowledge that remains valuable independently of the implementation environment.

The organization of this book follows the logic of the immersive environment design process. Readers begin with the foundations of perception and presence, because without understanding how the human sensory system functions, it is difficult to make informed decisions about movement, scale, or interaction. In conventional digital interfaces, many design decisions can be made primarily at the visual or functional level without deeply engaging with the physiological aspects of user experience. In VR, this is no longer sufficient. Even fundamental decisions concerning field of view, scale relationships, scene depth, or movement tempo directly affect user comfort and the ability to maintain presence. For this reason, the first part of this book focuses on perception, presence, spatial composition, scenography, and spatial audio. These topics establish the foundation for all subsequent design decisions.

Only after this conceptual basis has been established does this book move into interaction, locomotion, and user interfaces (UIs). This sequence reflects the view that movement and action in VR should not be treated merely as a set of control techniques, but rather as a consequence of how users understand and physically experience space. The chapters devoted to interaction, locomotion, animation, and spatial forms of UI demonstrate that every design decision must be considered simultaneously across ergonomic, cognitive, sensory, and narrative dimensions. Locomotion occupies a particularly important position because movement in VR remains one of the strongest factors determining whether an experience succeeds or fails. In practice, the selected movement model often determines whether users accept the environment or interrupt the experience after only a few minutes because of discomfort.

The third part of this book extends the design perspective beyond interaction mechanics. Contemporary VR experiences increasingly function in educational, social, therapeutic, and cultural contexts, which means that designers cannot limit themselves to ergonomic concerns alone. Understanding narrative structure, social presence, avatar–user relationships, ethical responsibility, and accessibility principles becomes equally important. The placement of these topics after the technical and spatial chapters is intentional: only after understanding how immersive environments operate at a perceptual and interactional level can one meaningfully examine how VR influences social relations, constructs narrative meaning, and affects different groups of users.

The final section of this book has a synthetic and evaluative character. It includes heuristics, criteria for assessing experience quality, performance considerations, and methodologies for usability testing in VR. This structure reflects the natural progression of design competence: first understanding the underlying phenomena, then designing experience components, and only afterward systematically evaluating them. In this sense, this book does not simply communicate knowledge; it also develops a design mindset based on conscious decision-making and iterative verification.

The publication is primarily intended for three groups of readers. The first consists of immersive experience designers, including both professionals already developing VR applications and those transitioning from traditional UX or interaction design into spatial environments. For them, the key objective is to demonstrate that VR requires a fundamentally different way of thinking than screen-based design. The second group consists of academic teachers delivering courses in VR, human–computer interaction (HCI), interaction design, game design, or immersive media. The structure of this book allows individual chapters to function as independent teaching modules, which is particularly valuable when developing new academic curricula. The third group includes students, who often encounter VR first through technological tools without yet possessing the conceptual framework necessary to critically understand design decisions.

One of the key assumptions of this book is its independence from any specific implementation context. This does not mean abandoning practical usefulness, but rather shifting practical value from technical execution to design reasoning. Instead of step-by-step configuration instructions, readers are given a framework that allows them to evaluate independently why a given solution works – or does not work. This approach increases both the longevity of the publication and its usefulness across diverse technological contexts.

In practice, this book is intended to function as a design guide that helps readers ask the right questions before implementation begins. Rather than focusing on how to reproduce a technical effect, it encourages reflection on whether a design choice supports presence, reduces unnecessary sensory load, preserves

spatial logic, and responds to the needs of the intended audience. This organizational logic reflects a contemporary understanding of immersive experience design, in which quality depends not on the number of implemented features, but on the coherence of decisions made across multiple levels simultaneously.

As a result, this book can serve equally well as an academic textbook, a course resource, and a practical guide for teams developing VR applications in education, training, culture, healthcare, or industry. Its structure has been designed so that readers not only acquire knowledge, but gradually develop the ability to understand spatial digital experiences as a distinct and emerging design domain.

About the Editors

Rocío Cifuentes-Albeza, Ph.D., is Associate Professor of Audiovisual Communication and Advertising at Miguel Hernández University (UMH), Spain. She holds a Bachelor's degree in Fine Arts from the Polytechnic University of Valencia (1999) and was awarded the Extraordinary Doctorate Award (UMH) in 2017. Since 2002, she has been a member of the faculty, teaching across the Bachelor's degrees in Fine Arts, Audiovisual Communication, and Journalism. She was Associate Dean of the BA degree program in Audiovisual Communication at UMH (2019–2023).

Iñigo Cuiñas, Ph.D., is a Professor and the Head of the Department of Signal Theory and Communications at the University of Vigo. He teaches courses on Remote Sensing and the Social Links of Engineering at the School of Telecommunication Engineering. Besides, he is involved in promoting Design Thinking methodology among lecturers at different universities, by means of workshops and speeches. He has joined or led different Erasmus+ projects focused on Design Thinking, soft skills, virtual reality in education, Engineering for Peace, outdoor learning activities, and other social and/or transversal issues.

Itziar Goicoechea, Ph.D., is Associate Professor of Project Management in the Department of Engineering Design at Vigo University, Spain. She holds a Bachelor's degree in Industrial Engineering specializing in Enterprise Organization from the School of Engineering at the University of Vigo (1996) and earned her Ph.D. degree from the same institution in 2006. Additionally, she has a Master's degree in Language and Communication in Business and is a member of the Spanish Association of Project Engineering (AEIPRO). She has extensive expertise in project management, with over 15 years of experience in this field in private companies in the automotive industry and construction industries. Her specialties include predictive and agile methodologies in project management as well as design thinking methodology.

Dorota Kamińska, Ph.D., is an Associate Professor and researcher at the Institute of Mechatronics and Information Systems, Lodz University of Technology, and founder of the Voxel Research Lab. Her work focuses on virtual and augmented reality, human–computer interaction, and AI-supported learning, with particular interest in well-being, accessibility, and immersive technologies in education and healthcare.

Francisco-Julián Martínez-Cano, Ph.D., serves as Professor of Audiovisual Direction and Production and Associate Dean of the BA degree program in Audiovisual Communication at Miguel Hernández University, Spain. He has been the recipient of several research fellowships and awards, including the José Castillejo Research Grant from the Spanish Ministry of Universities (2022) and Eng. Fernando Gonçalves Lavrador's Honourable Mention for Best Research Paper at the Avanca Cinema International Conference (2020). He directs the OCAP research group, the Audiovisual Communication and Advertising Observatory at UMH. His research explores the intersection of virtual and augmented reality, AI, cinema, and video games by analyzing the effects and interconnectedness of immersive experiences within pro-social VR, gaming, and narrative structures.

Rui Raposo, Ph.D., is an Associate Professor in the field of Science and Communication Technologies at the Department of Communication and Art from the University of Aveiro. His current research interests focus on issues related to the use of Extended Reality technologies and experiences (VR, AR, MR, and XR) in various contexts of use such as well-being, e-tourism, and cybermuseology. He also teaches, researches, and supervises activities in the field of audiovisual content production.

Mário Vairinhos, Ph.D., is an Associate Professor and researcher in the area of Human–Computer Interaction (HCI) and Extended Reality. He is an integrated member of DigiMedia – Digital Media and Interaction at the University of Aveiro. He is the coordinator of the Research Group in Immersive Play at the host institution of DigiMedia. He is the Co-Director of the Master’s program in Digital Game Development at the University of Aveiro.

Grzegorz Zwoliński, Ph.D., is a researcher at the Institute of Mechatronics and Information Systems, Lodz University of Technology. His research focuses on virtual and augmented reality, immersive systems, interaction design, and usability, with applications in education, training, and simulation-based environments supporting user-centered development of VR technologies.

Contributors

Manuel Caeiro

atlanTTic Research Center for Telecommunication
Technologies
Universidade de Vigo
Vigo, Spain

María Cerdeira

Department of Specific Didactics
Universidade de Vigo
Ourense, Spain

Agnieszka Dubiel

Faculty of Electrical, Electronic, Computer
and Control Engineering
Lodz University of Technology
Lodz, Poland

Tatiana Santos Gonçalves

Department of Social and Human Sciences
Universidad Miguel Hernández
Elche, Spain

Marta Kuźmicz

Faculty of Electrical, Electronic, Computer
and Control Engineering
Lodz University of Technology
Lodz, Poland

Anna Laska-Leśniewicz

Faculty of Electrical, Electronic, Computer
and Control Engineering
Lodz University of Technology
Lodz, Poland

Silvino Martins

Department of Communication and Arts
University of Aveiro
Aveiro, Portugal

Pedro Reisinho

Department of Communication and Arts
University of Aveiro
Aveiro, Portugal

Cátia Silva

Department of Communication and Arts
University of Aveiro
Aveiro, Portugal

Introduction to VR Design

Historical Milestones in VR Design

Marta Kuźmicz and Dorota Kamińska

ABSTRACT

Virtual reality (VR) design has evolved significantly over the past decades, moving from early conceptual ideas of immersive experiences to advanced interactive systems widely used in entertainment, education, healthcare, and training. Understanding the historical development of VR technologies and design principles is essential for analyzing current design approaches and anticipating future directions. This chapter presents an overview of key milestones in the evolution of VR design, beginning with early attempts to create immersive visual experiences and continuing through the emergence of computer-based virtual environments in the 1960s and 1970s. It further discusses the technological and conceptual developments of the 1980s and 1990s, the temporary decline of VR due to technological limitations, and the renewed interest in immersive technologies following the emergence of modern head-mounted displays and game engines in the 2010s. By examining these historical stages, this chapter highlights how advances in hardware, software, and interaction techniques have shaped contemporary VR design practices and contributed to the development of more immersive and user-centered virtual environments.

INTRODUCTION TO VIRTUAL REALITY

Virtual reality (VR) refers to computer-generated environments designed to simulate real or imaginary worlds and provide users with the experience of being physically present within those environments (Sherman & Craig, 2018). Through specialized hardware such as head-mounted displays (HMDs), motion tracking systems, and interactive controllers, VR systems allow users to perceive and interact with digital content in three-dimensional space (Burdea & Coiffet, 2003). The key characteristics of VR experiences include immersion, interaction, and the sense of presence, which describes the psychological perception of being there in a virtual environment (Slater, 2009).

Advances in computing power, graphics technologies, and tracking systems have enabled VR to evolve from experimental research systems into widely available consumer technologies. Today, VR is

used across a wide range of domains including entertainment, education, healthcare, training, and scientific visualization. Modern VR systems can deliver highly immersive experiences by synchronizing visual, auditory, and sometimes haptic stimuli with the user's movements in real time.

Although contemporary VR technologies rely on advanced digital systems, the fundamental ideas behind immersive experiences emerged long before the development of modern computing. Early attempts to create immersive visual environments date back to the nineteenth century and include devices such as stereoscopes and panoramic installations. These early experiments demonstrated that carefully designed visual stimuli could influence human perception and create the illusion of spatial presence. Understanding these historical developments helps explain how modern VR technologies evolved and provides important context for contemporary VR design.

EARLY CONCEPTS OF VR (BEFORE DIGITAL VR)

The idea of creating immersive experiences that simulate alternative realities existed long before the development of modern digital technologies. Early concepts related to VR can be traced back to the nineteenth century, when researchers and inventors began experimenting with visual techniques that attempted to reproduce depth perception and create the illusion of presence in a different environment. One of the most important developments in early immersive visual technologies was the stereoscope (Figure 0.1). The principle of stereoscopic vision was first described by Charles Wheatstone in 1838 (Wheatstone, 1838), and later device designs made the technology widely accessible. One of the most popular devices was the Holmes stereoscope, introduced in 1861 and widely used at the end of the nineteenth century. This device allowed viewers to observe two slightly different images simultaneously, reproducing the disparity between the perspectives of the left and right eyes, and thereby creating the illusion of depth and a three-dimensional scene. The stereoscope demonstrated that the human visual system could be manipulated to perceive spatial environments that do not physically exist, laying an important conceptual foundation for later VR technologies.

Following the development of stereoscopic viewing devices, further attempts were made to create more immersive visual experiences. During the nineteenth century, panoramic paintings became a popular form of entertainment and visual experimentation. Panoramas were large circular paintings displayed inside specially designed rotundas, allowing viewers to stand at the center of the installation and observe a continuous 360° scene surrounding them. This format created an illusion of spatial immersion

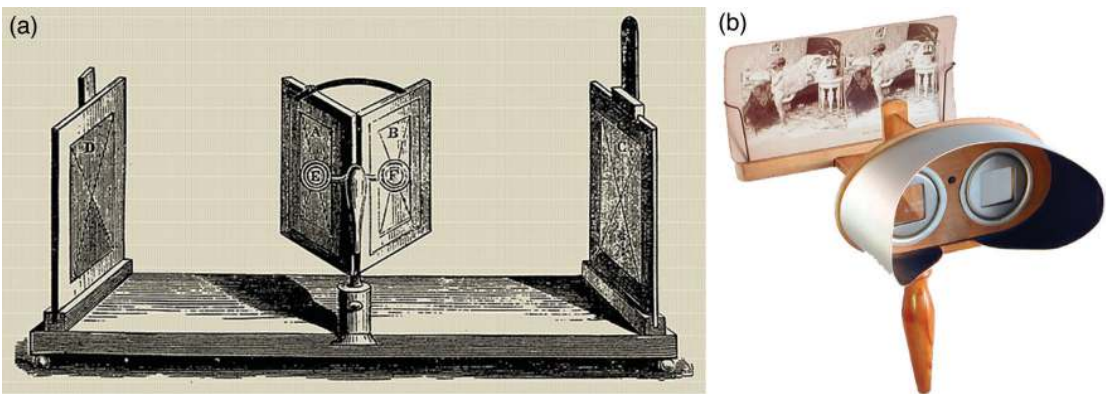


FIGURE 0.1 Early stereoscopic viewing devices. (a) Wheatstone stereoscope (1838), one of the first instruments designed to demonstrate binocular vision and depth perception. (b) Holmes stereoscope (1861), a later and more accessible design widely used in the nineteenth century. (Source: Wikimedia Commons [Left: https://commons.wikimedia.org/wiki/File:Charles_Wheatstone-mirror_stereoscope_XIXc.jpg. Right: https://commons.wikimedia.org/wiki/File:Holmes_Stereoscope_1861.png].)



FIGURE 0.2 Racławice panorama (1893–1894), a large circular panoramic painting displayed in a rotunda and designed to immerse viewers in a surrounding 360° visual environment. (Source: Courtesy of the National Museum in Wrocław.)

by expanding the viewer's field of vision and minimizing external visual distractions. Such installations attempted to immerse the observer within a simulated environment rather than simply presenting an image to be viewed from a distance (Grau, 2004) (Figure 0.2).

Although panoramas were static and lacked interactivity, they introduced important concepts that would later become central to VR design. By surrounding the viewer with visual content and limiting the perception of the external world, panoramas attempted to create a sense of presence within the depicted environment. These early immersive installations demonstrated that carefully designed visual stimuli could significantly influence human perception and create the illusion of being in another place. At the beginning of the twentieth century, further developments in visual media continued to explore similar ideas. Advances in photography and cinematography allowed creators to experiment with stereoscopic images and films designed to enhance the perception of depth. Stereoscopic photography became particularly popular in the late nineteenth and early twentieth centuries, allowing viewers to experience realistic three-dimensional scenes using relatively simple optical devices.

A major step toward multisensory immersive systems occurred in the mid-twentieth century with the work of Morton Heilig. In 1962, he developed the Sensorama, a mechanical device designed to simulate real-world experiences by stimulating multiple human senses simultaneously (Heilig, 1962). The system presented stereoscopic film combined with spatial audio, vibration, airflow, and even scent to enhance the user's sense of presence within the simulated environment (Figure 0.3). The goal of the device was to create a more convincing illusion of being physically present inside the displayed scene.

Although the Sensorama did not achieve commercial success and lacked interactive capabilities, it is widely recognized as one of the earliest attempts to create a multisensory immersive media system. Importantly, it introduced several concepts that later became central to VR design, including multisensory stimulation, immersive viewing conditions, and the attempt to enhance the feeling of presence within a simulated environment.

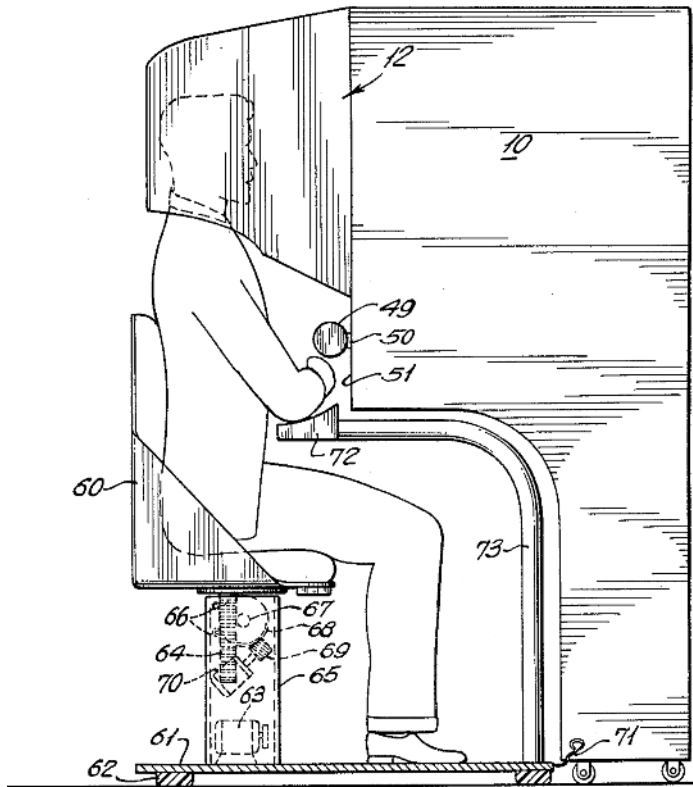


FIGURE 0.3 Sensorama device designed by Morton Heilig (1962). (Illustration from U.S. Patent No. 3050870; Source: Wikimedia Commons (https://commons.wikimedia.org/wiki/File:Sensorama_patent_fig5.png).)

These early experiments, ranging from panoramic environments and stereoscopic media to multisensory cinematic systems, laid out the conceptual foundations for modern VR. Even though they relied on optical or mechanical technologies rather than digital computing, they demonstrated how human perception could be influenced by carefully designed visual and sensory stimuli. The principles developed during this period would later play a crucial role in the emergence of computer-based VR systems.

THE FIRST COMPUTER-BASED VR SYSTEMS (1960s–1970s)

The transition from early optical and mechanical immersive devices to computer-based virtual environments began in the 1960s, when advances in computer graphics and interactive technologies made it possible to generate simulated environments digitally. During this period, researchers started to explore the possibility of combining computer-generated imagery with devices capable of presenting immersive visual experiences to the user. These developments marked the beginning of what is commonly considered the first generation of VR systems. One of the most influential milestones in the history of VR was the work of Ivan Sutherland (Sutherland, 1968). In 1968, he developed what is widely regarded as the first HMD system capable of presenting computer-generated graphics directly to the user's eyes. The device became known as the "Sword of Damocles" due to its large mechanical structure suspended from the ceiling above the user (Figure 0.4). The system used mechanical tracking to adjust the displayed image



FIGURE 0.4 The “Sword of Damocles” head-mounted display developed by Ivan Sutherland in 1968, considered the first computer-based virtual reality system capable of displaying computer-generated graphics aligned with the user’s head movements. (Image courtesy of the Computer History Museum, <https://www.computerhistory.org/revolution/input-output/14/356/1888>)

according to the user’s head movements, allowing the viewer to observe a simple wireframe virtual environment from different perspectives.

At the same time, other researchers explored alternative approaches to immersive interaction. One notable example was the work of Myron Krueger, who developed interactive environments in the 1970s that allowed users to interact with computer-generated visuals using body movements. His system, known as Videoplace, used cameras and projection technologies to capture the user’s silhouette and integrate it into a virtual environment in real time. Krueger described these interactive environments using the term “artificial reality”, emphasizing the experience of direct interaction with computer-generated spaces without the use of HMDs (Krueger, 1983).

Although the graphical capabilities of the system were extremely limited compared to modern VR technologies, the concept introduced by Sutherland represented a breakthrough in immersive computing. For the first time, a computer-generated environment could be visually aligned with the movements of the user’s head, creating a basic form of spatial interaction with a virtual scene. This principle of head-tracked visual feedback later became one of the fundamental components of modern VR systems.

At the same time, research institutions such as the University of Utah and later NASA Ames Research Center began investigating interactive computer graphics and simulation technologies. These research efforts contributed significantly to the development of early virtual environments, flight simulators, and telepresence systems. NASA explored the use of HMDs and motion tracking systems for training purposes and remote operation of equipment.

Despite these advances, early VR systems of the 1960s and 1970s remained largely experimental. The available computing power was limited, and the graphical representations of virtual environments consisted mostly of simple geometric wireframe models. Nevertheless, these early prototypes demonstrated

the feasibility of immersive computer-generated environments and established many of the technical principles that continue to shape VR systems today.

The research conducted during this period laid the groundwork for later developments in immersive technologies, including more advanced HMDs, improved motion tracking, and increasingly realistic virtual environments. As computing technology continued to evolve, these early concepts would eventually lead to the emergence of more sophisticated VR systems in the following decades.

DEVELOPMENT OF VR RESEARCH IN THE 1980s

The 1980s marked a significant period in the development of VR technologies. During this time, advances in computing power, graphics processing, and motion tracking systems enabled researchers to explore more sophisticated immersive environments. VR began to evolve from isolated experimental prototypes into structured research programs conducted by universities, laboratories, and governmental institutions. These efforts focused on improving interaction methods, visual immersion, and the practical applications of virtual environments.

One of the most important research centers contributing to the development of VR technologies during this period was NASA Ames Research Center. Researchers at NASA investigated immersive displays and interactive virtual environments for applications such as pilot training and teleoperation (Fisher et al., 1987). Early systems often combined HMDs with motion tracking and gesture-based interaction devices such as the DataGlove, enabling users to manipulate objects in virtual environments using natural hand movements (Figure 0.5).



FIGURE 0.5 The VIEW (virtual interactive environment workstation) system developed at NASA Ames Research Center in the late 1980s. The operator uses a head-mounted display together with DataGlove controllers to interact with a virtual environment. (Source: Wikimedia Commons [[https://commons.wikimedia.org/wiki/File:THE_VIEW_\(Virtual_Reality\).jpg](https://commons.wikimedia.org/wiki/File:THE_VIEW_(Virtual_Reality).jpg)].)

An important technological innovation of the 1980s was the development of the DataGlove, an input device designed to capture hand movements and gestures. The glove contained sensors that measured finger bending and hand orientation, allowing the system to translate physical hand movements into actions within the virtual environment. This technology significantly improved the naturalness of interaction with virtual objects and became one of the early foundations of gesture-based interaction in VR systems (Zimmerman et al., 1986).

The 1980s also saw significant progress in the development of simulation systems and telepresence technologies. Researchers explored how immersive environments could allow users to control machines or observe remote locations as if they were physically present. Such systems were particularly relevant in areas where direct human presence was difficult or dangerous, such as space exploration, deep-sea operations, or hazardous industrial environments. These developments introduced the concept of telepresence, where users could experience and interact with a remote environment through technological mediation.

Although VR systems during this period were still limited by hardware constraints and high costs, the research conducted in the 1980s established many of the foundations of modern VR. Advances in interaction devices, immersive displays, and simulation technologies demonstrated the potential of virtual environments for training, research, and human–computer interaction. These developments also prepared the ground for the commercialization attempts and technological expansion that would occur in the following decades.

COMMERCIALIZATION ATTEMPTS IN THE 1990s

The 1990s marked an important phase in the history of VR, characterized by the first attempts to commercialize VR technologies. Advances in computer graphics, display technologies, and interaction devices encouraged companies and research institutions to explore the possibility of bringing VR systems to a broader audience. During this period, VR began to transition from research laboratories to commercial environments such as entertainment centers, arcades, and consumer electronics markets.

At the same time, research laboratories continued to develop alternative immersive technologies that did not rely on HMDs. One of the most influential systems was the CAVE (cave automatic virtual environment), introduced in 1992 at the University of Illinois (Cruz-Neira et al., 1992). The CAVE system projected stereoscopic images onto the walls and floor of a room-sized cube, allowing users wearing stereoscopic glasses to experience immersive virtual environments without a HMD (see Figure 0.6). This approach enabled multiple users to observe the same virtual scene simultaneously and became widely used in scientific visualization, engineering design, and research applications.

The beginning of the decade marked a breakthrough in the creation of entirely synthetic, shared environments. In 1990, Blanchard et al. introduced “Reality Built for Two”, which allowed multiple users to share a virtual space simultaneously in real time. To interact with this environment, users employed the DataGlove for gestural manipulation alongside the EyePhone HMD (Blanchard et al., 1990). This system represented a significant conceptual shift from single-user applications to multi-user virtual communication. However, the technology was not brought to the public due to high costs. A single DataGlove was priced at \$10,000 and required a \$250,000 computing system to function (Hine, 2025).

Other pioneers turned their attention to making such experiences accessible to the mainstream. One of the most notable companies involved in early commercial VR development was Virtuality Group, a British company that introduced arcade-based VR systems in the early 1990s. These systems typically consisted of HMDs, motion tracking systems, and interactive controllers that allowed users to participate in multiplayer virtual environments. Virtuality machines were installed in arcades and entertainment venues, offering experiences such as virtual sports games or multiplayer combat simulations. Although technologically limited compared to modern VR systems, these installations represented one of the first attempts to provide immersive interactive experiences to the public (La Viola et al., 2017).

Another significant attempt to introduce VR-like technology to the consumer market was the Nintendo Virtual Boy (see Figure 0.7), released in 1995. The device was marketed as a portable VR gaming console



FIGURE 0.6 CAVE (cave automatic virtual environment), an immersive projection-based virtual reality system in which stereoscopic images are displayed on the walls of a room-sized cube, allowing users inside the environment to explore virtual scenes without wearing a head-mounted display. (Source: Wikimedia Commons [https://commons.wikimedia.org/wiki/File:CAVE_Crayoland.jpg].)

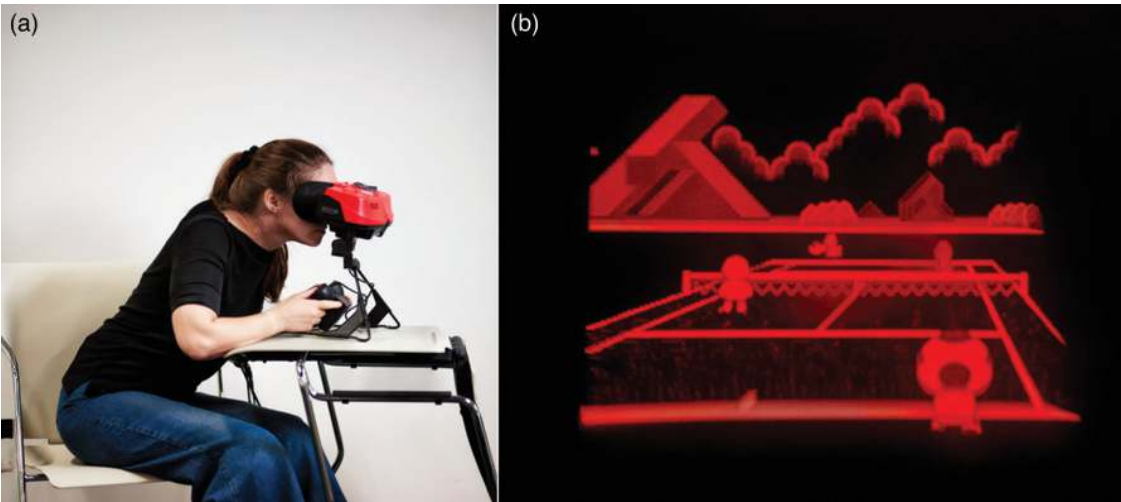


FIGURE 0.7 Nintendo Virtual Boy (1995), an early consumer device designed to deliver stereoscopic gaming experiences. (a) User interacting with the Virtual Boy console. (b) Example of the stereoscopic monochromatic display used by the system.

and used stereoscopic displays to create the perception of depth. However, unlike modern VR headsets, the Virtual Boy did not provide head tracking or fully immersive environments. Users interacted with the system while looking into a fixed viewing device that displayed monochromatic red graphics. Despite strong expectations, the product failed commercially and was discontinued shortly after its release. The failure was largely attributed to technological limitations such as low display resolution, lack of head tracking, limited graphical capabilities, and reports of user discomfort during prolonged use (Kent, 2001).

Many early VR products of the 1990s faced significant technological limitations that restricted their widespread adoption. Hardware capabilities at the time were insufficient to generate complex virtual environments in real time with high visual fidelity. Low display resolution, limited field of view, and high system latency often reduced the quality of the experience and could lead to discomfort for users. Additionally, VR systems were typically expensive and required specialized equipment, making them impractical for large-scale consumer markets.

As a result, the enthusiasm surrounding VR technologies during the early 1990s gradually declined. Although the commercialization attempts of this period were largely unsuccessful, they played an important role in demonstrating the potential of immersive technologies and highlighted the technical challenges that needed to be overcome. Lessons learned from these early systems would later contribute to the development of more advanced and accessible VR technologies in the following decades.

THE VR RENAISSANCE (2010–2016)

The period between 2010 and 2016 is often described as the renaissance of VR, marking the transition of VR from experimental research environments into the mainstream technology market. This revival was largely driven by advances in computing power, graphics processing, and display technologies, which made immersive systems significantly more accessible and affordable than earlier attempts in the 1990s.

A major turning point occurred in 2012, when Palmer Luckey launched the Oculus Rift headset through a Kickstarter crowdfunding campaign (Figure 0.8). The project quickly gained international attention and demonstrated strong demand for affordable, high-quality VR devices (Jerald, 2015). In 2014, Oculus VR was acquired by Facebook (now Meta), which significantly accelerated investment and development in the VR industry. Following this momentum, several major technology companies entered the VR market. Companies such as HTC, Sony, and Samsung introduced their own VR devices, while Google explored mobile-based VR solutions. This influx of investment and competition helped drive rapid improvements in hardware quality, tracking technology, and user experience (La Viola et al., 2017).

Mobile VR solutions also played an important role in expanding access to VR technologies. One of the most widely known initiatives was Google Cardboard (see Figure 0.9), introduced in 2014. This simple and inexpensive viewer used a smartphone display together with a pair of lenses to create stereoscopic VR experiences. Although technologically limited compared to dedicated VR headsets, Google Cardboard significantly lowered the barrier to entry and allowed millions of users to experience basic VR applications using devices they already owned.

Shortly afterward, Samsung introduced the Gear VR headset (see Figure 0.9) in collaboration with Oculus. Unlike Google Cardboard, Gear VR incorporated additional sensors and a more advanced interaction interface, providing improved head tracking and a more comfortable user experience. The device used Samsung smartphones as both the display and computing unit, making it one of the first widely distributed mobile VR systems and an important step toward the broader adoption of immersive technologies (La Viola et al., 2017).

Another important milestone of this period was the introduction of the HTC Vive system in 2016, developed by HTC in cooperation with Valve. Unlike earlier VR systems, HTC Vive enabled room-scale VR, allowing users to physically move within a tracked space. The system used external tracking devices known as Lighthouse base stations, which emit infrared signals detected by sensors located on the headset and controllers.



FIGURE 0.8 Oculus rift development kit 1 (2013). (Source: Wikimedia Commons [https://commons.wikimedia.org/wiki/File:Oculus_Rift_-_Developer_Version_-_Front.jpg].)

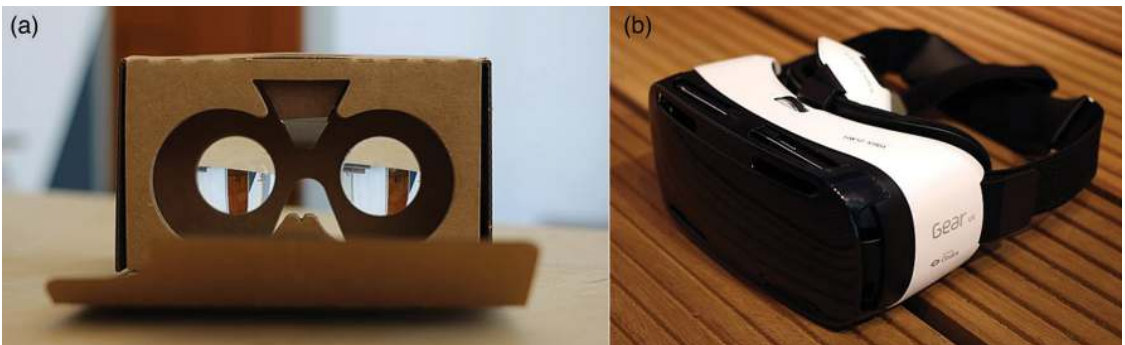


FIGURE 0.9 (a) Google Cardboard (2014), a low-cost smartphone-based VR viewer. (b) Samsung Gear VR (2015), a mobile virtual reality headset developed in collaboration with Oculus. (Source: Wikimedia Commons [Left: [https://commons.wikimedia.org/wiki/File:Google_Cardboard_2_\(18634310535\).jpg](https://commons.wikimedia.org/wiki/File:Google_Cardboard_2_(18634310535).jpg). Right: https://commons.wikimedia.org/wiki/File:Samsung_Gear_VR.jpg].)

(Figure 0.10). This technology enabled precise tracking of the user's position and movements within the room, significantly increasing the sense of immersion. The HTC Vive also relied on a wired connection to a powerful computer, ensuring high graphical fidelity and low latency. Although the system required additional equipment and setup, it provided one of the most immersive VR experiences available at the time and became widely used in research laboratories, gaming, and professional applications (La Viola et al., 2017).

At the same time, the development of powerful and accessible game engines greatly supported VR content creation. Platforms such as Unity and Unreal Engine introduced dedicated VR support, allowing developers to build immersive applications more efficiently. These tools significantly lowered the



FIGURE 0.10 HTC Vive headset and Lighthouse tracking system enabling room-scale VR.

barrier to entry for creating VR experiences and played a crucial role in expanding VR beyond gaming into areas such as education, healthcare, and training. Together, these technological and market developments marked the beginning of the modern VR ecosystem and laid the foundation for the rapid growth of immersive technologies in the following years (Radianti et al., 2020).

STANDALONE VR AND MAINSTREAM ADOPTION (2019–PRESENT)

The next major stage in the evolution of VR began with the introduction of standalone VR headsets, which no longer required an external computer, cables, or external tracking systems. These devices integrated displays, processors, sensors, and tracking systems directly into the headset, making VR significantly more accessible to a wider audience.

A key milestone was the release of the Oculus Quest in 2019. Unlike earlier systems such as Oculus Rift or HTC Vive, the Quest operated entirely independently and used inside-out tracking, where built-in cameras on the headset track the user's position and hand controllers without the need for external sensors. This greatly simplified setup and allowed users to freely move within their environment.

Subsequent generations of standalone headsets, including Meta Quest 2 and Meta Quest 3 (see Figure 0.11), further improved display resolution, processing power, and interaction capabilities. These devices also introduced hand tracking, allowing users to interact with virtual environments without controllers.

The availability of standalone VR systems significantly expanded the range of VR applications beyond gaming. VR began to be widely adopted in education, healthcare, training, design, and research,

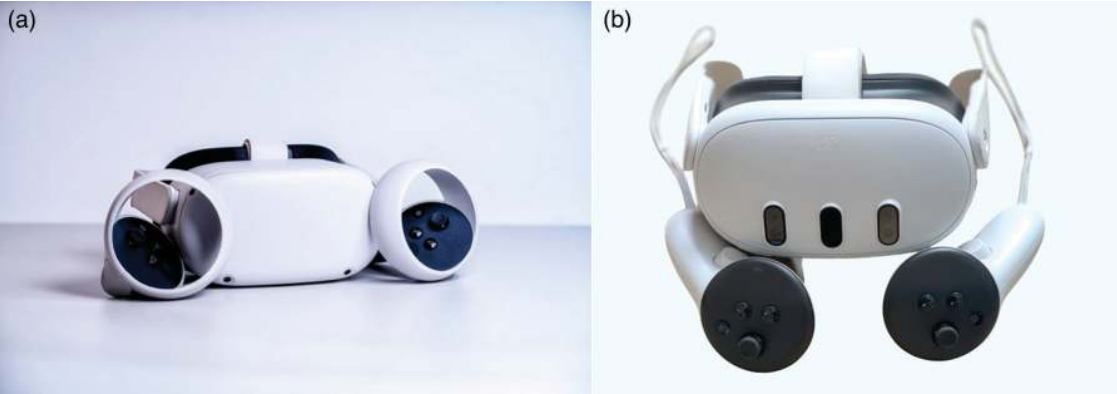


FIGURE 0.11 Standalone VR headset [(a) Meta Quest 2, (b) Meta Quest 3] using inside-out tracking, which eliminates the need for external sensors or wired connections.

TABLE 0.1 Key milestones in the development of virtual reality technologies

PERIOD	KEY DEVELOPMENTS	EXAMPLES/SYSTEMS
Nineteenth century	Early experiments with immersive visual perception	Wheatstone stereoscope (1838); Holmes stereoscope (1861); panoramic installations
1950s–1960s	Early multisensory immersive media systems	Sensorama (Heilig, 1962)
1960s	First computer-based head-mounted display systems	Sword of Damocles (Sutherland, 1968)
1970s	Interactive environments and early concepts of artificial reality	Videoplace (Krueger, 1970s)
1980s	Development of immersive displays, tracking systems, and gesture-based interaction	NASA Ames VR research; DataGlove
Early 1990s	Projection-based immersive environments and early commercial VR systems	CAVE (Cruz-Neira et al., 1992); virtuality arcade systems
Mid-1990s	Early consumer virtual reality attempts	Nintendo Virtual Boy (1995)
2010–2016	Renaissance of VR driven by modern head-mounted displays and game engines	Oculus Rift; HTC Vive; PlayStation VR
2014–2016	Mobile virtual reality solutions	Google Cardboard; Samsung Gear VR
2019–present	Standalone VR systems and wider mainstream adoption	Meta Quest series

as the technology became easier to deploy and more affordable. As a result, VR moved from a niche technology used primarily in laboratories and specialized facilities to a widely accessible platform supporting immersive experiences across many domains (Pallavicini & Pepe, 2020). To summarize the evolution discussed in this chapter, Table 0.1 outlines the key milestones in the development of VR from early experiments to modern standalone systems.

FROM VR TECHNOLOGY TO VR DESIGN PRINCIPLES

As VR technologies became more mature and widely available, the focus of research and development gradually shifted from hardware capabilities to the design of effective and meaningful user experiences. While early VR systems primarily explored technological possibilities, contemporary VR applications

increasingly emphasize how users interact with virtual environments and how these environments influence perception, behavior, and emotional responses.

This shift brought attention to key concepts such as immersion, presence, and usability, which are fundamental for understanding how users experience virtual environments. Immersion refers to the technological properties of the system that enable a convincing virtual environment, including display quality, field of view, tracking accuracy, and system latency. Presence, on the other hand, describes the psychological sensation of “being there” within the virtual space, where users perceive the virtual environment as if they were physically located inside it.

In addition to immersion and presence, interaction design has become a crucial component of VR experience development. Unlike traditional graphical interfaces, VR environments require spatial interaction techniques that allow users to navigate, manipulate objects, and perform tasks within three-dimensional spaces. Designers must therefore carefully consider input devices, gesture-based interaction, controller design, and spatial navigation methods to ensure that interaction remains intuitive and natural.

At the same time, usability in VR became an important research topic, as designers began to investigate how interface elements, interaction techniques, and spatial navigation affect user performance, comfort, and cognitive load. VR interfaces introduce unique challenges that are not present in traditional desktop systems, such as spatial orientation, depth perception, and physical movement within the virtual environment. Poorly designed interfaces may lead to confusion, increased cognitive effort, or even physical discomfort.

Another important aspect of VR design concerns user comfort and physiological responses. Extended exposure to virtual environments may cause symptoms such as eye strain, fatigue, or cybersickness, particularly when there are inconsistencies between visual motion cues and the user’s physical movement. As a result, designers must consider factors such as motion design, frame rate stability, latency reduction, and appropriate locomotion techniques when developing VR applications.

Furthermore, VR experiences often involve a strong sense of embodiment, where users perceive a virtual body or avatar as representing themselves within the digital environment. This sense of embodiment can significantly influence user behavior, emotional engagement, and social interaction in virtual spaces. Consequently, designing effective VR systems requires attention not only to technical performance but also to psychological and perceptual aspects of user experience.

Designing effective VR experiences therefore requires not only advanced technology but also carefully developed design principles that support intuitive interaction, maintain user comfort, and minimize potential usability issues. Understanding how users perceive and interact with virtual environments has become a central concern for researchers and designers working with immersive technologies.

Today, VR is increasingly considered part of a broader ecosystem of immersive technologies that also includes augmented reality (AR) and mixed reality (MR), often collectively referred to as extended reality (XR).

These concepts form the foundation for contemporary research on VR usability and user experience, which is discussed in the following sections.

REFERENCES

- Blanchard, C., Burgess, S., Harvill, Y., Lanier, J., Lasko, A., Oberman, M., & Teitel, M. (1990). Reality built for two: a virtual reality tool. In *Proceedings of the 1990 Symposium on Interactive 3D Graphics*, Snowbird, UT, United States (pp. 35–36). Association for Computing Machinery.
- Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology*. John Wiley & Sons.
- Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6), 64–73.
- Fisher, S. S., McGreevy, M., Humphries, J., & Robinett, W. (1987, January). Virtual environment display system. In *Proceedings of the 1986 Workshop on Interactive 3D Graphics*, Chapel Hill, NC, United States (pp. 77–87).

- Grau, O. (2004). *Virtual Art: from illusion to immersion*. MIT press.
- Heilig, M. L. (1962). *Sensorama Simulator*, US Patent No. 3050870. US Patent and Trademark Office. <https://patents.google.com/patent/US3050870A/en>.
- Hine, E. (2025). Virtual reality, cyberspace, and embodiment: a historical debate with contemporary resonance. *Virtual Reality*, 29, 63.
- Kent, S. L. (2001). *The ultimate history of video games: From Pong to Pokemon_Revisited*. Three Rivers Press.
- Krueger, M. W. (1983). *Artificial reality*. Addison-Wesley.
- Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.
- La Viola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: Theory and practice*. Addison-Wesley Professional.
- Pallavicini, F., & Pepe, A. (2020). Virtual reality games and the role of body involvement in enhancing positive emotions and decreasing anxiety: Within-subjects pilot study. *JMIR Serious Games*, 8(2), e15635.
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & education*, 147, 103778.
- Sherman, W. R., & Craig, A. B. (2018). *Understanding virtual reality: Interface, application, and design*. Morgan Kaufmann.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549–3557.
- Sutherland, I. E. (1968, December). A head-mounted three-dimensional display. In *Proceedings of the Fall Joint Computer Conference, part I*, December 9–11, 1968, San Francisco, California, USA (pp. 757–764).
- Wheatstone, C. (1838). Contributions to the physiology of vision. *Philosophical Transactions of the Royal Society*, 128, 371–394.
- Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S., & Harvill, Y. (1986). A hand gesture interface device. *ACM Sigchi Bulletin*, 18(4), 189–192.

PART I

Foundations of Presence



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Presence and Visual Depth Perception in Virtual Reality

Mário Vairinhos, Silvino Martins, Rui Raposo,
and Pedro Reisinho

ABSTRACT

This chapter examines the critical role of depth perception in facilitating presence and immersion within virtual reality (VR) environments. It provides a comprehensive taxonomic analysis of depth cues, categorized into pictorial, oculomotor, and dynamic sources, detailing their integration within the human visual system.

Furthermore, this chapter offers practical design guidelines for immersive environments, addressing technical challenges and the strategic use of perceptual anchors to maintain spatial coherence. Ultimately, the work argues that a rigorous application of these visual mechanisms is a prerequisite for achieving narrative and ludic immersion, ensuring that the virtual environment functions as a credible egocentric reference frame for the user.

Human experience of the world is fundamentally grounded in perception. As John Berger famously observed, “seeing comes before words”, emphasizing the role of perception in how individuals situate themselves in relation to the surrounding world (Berger, 1972). Before we interpret or describe the world, we first perceive it through sensory processes that allow us to situate ourselves spatially in relation to objects, environments, and other people. This role of perception becomes particularly significant when designing immersive technologies such as virtual reality (VR), where users interact with simulated environments through perceptual information.

PRESENCE AND IMMERSION

One of the central concepts used to describe perceptual engagement with virtual environments is presence. Presence refers to the subjective psychological sensation of “being” in a virtual environment, despite the user’s awareness that the experience is mediated by technology (Jerald, 2015; Slater, 2003). When

presence occurs, users tend to respond to virtual environments in ways that resemble their responses to real-world situations, a phenomenon often described as a “perceptual illusion of non-mediation” (Lombard & Ditton, 1997). This experience is often dependent on the formation of a spatial situation model (SSM), a mental representation of the environment based on processed spatial cues and personal cognition (Wirth et al., 2007). According to this view, presence is achieved when the user adopts the SSM as their primary egocentric reference frame, successfully prioritizing virtual spatial cues over those of the physical world (Skarbez et al., 2017; Wirth et al., 2007).

Although closely related, presence should be distinguished from immersion. Immersion generally refers to the objective technological properties of a VR system that determine how effectively it delivers coherent sensory stimuli (Slater & Wilbur, 1997). Presence, by contrast, describes the user’s subjective experience that may emerge from interacting with such systems (Slater, 2003). However, Skarbez et al. (2017) argue that “immersion” alone is insufficient to describe the requirements for presence. They propose a model where presence is a function of two distinct system characteristics: Immersion and Coherence. In this framework, immersion is defined by the sensorimotor contingencies the system supports (e.g., “how” the system tracks and displays depth), while coherence refers to the extent to which the virtual world behaves in a logically and physically consistent manner (“what” is being simulated). For depth perception, this implies that a system must provide not only the technical cues for stereopsis (immersion) but also ensure those cues are consistent with the user’s physical expectations of space (coherence).

From this perspective, the capacity of a VR system to support convincing spatial experiences relates to its immersive potential. Factors such as the number of sensory channels, the fidelity of stimuli, and the quality of visual and auditory simulation all contribute to this potential (Slater, 2003). Systems that effectively integrate these elements support what Slater (2009) and Skarbez et al. (2017) identify as Place Illusion (PI), the strong sense of being in a virtual place. This illusion is particularly dependent on the system’s ability to provide depth information that matches the user’s head movements, a key sensorimotor contingency.

Furthermore, the technological affordances of immersion do not act in isolation, providing the sensory framework that facilitates other layers of engagement, including psychological dimensions of the immersive experience (Ermi & Mäyrä, 2005). In a 3D mediated environment, accurate visual depth perception enhances spatial coherence, ensuring that the virtual world aligns with the user’s sensorimotor expectations (Skarbez et al., 2017). Building upon previous research (Martins & Vairinhos, 2023), we argue that this coherence is a prerequisite for spatial immersion, one of the core dimensions of narrative immersion as defined by Ryan (2015), which characterizes the user’s connection with the setting and environment of a story. While narrative or ludic immersion can exist independently of spatial depth (as in literature or traditional media), in VR, any inconsistency in depth perception can disrupt the Plausibility Illusion (Skarbez et al., 2017). Such disruptions create a cognitive friction that may hinder the user’s ability to become fully absorbed in the story (Murray, 2016; Ryan, 2015) or the game’s mechanics (McMahan, 2003). Additionally, the stabilization of the virtual experience is further mediated by individual susceptibility, where traits like “immersive tendencies” determine how effectively a user integrates these spatial, narrative, and ludic layers into a coherent psychological whole (Witmer & Singer, 1998).

Within this broader framework, perceptual processes remain the essential foundation for achieving a sense of presence in VR. Vision provides the primary cues through which users interpret environmental structure and object relationships, making the perception of depth a bridge between the system’s perceptual immersive affordances and the internal construction of a stable SSM (Wirth et al., 2007). By satisfying the sensorimotor requirements for PI, depth perception allows the virtual environment to be recognized as a credible egocentric reference frame (Skarbez et al., 2017). Consequently, this spatial foundation serves as the necessary stage upon which narrative, ludic, and social layers of immersion can flourish without cognitive friction. Understanding how VR succeeds in placing the user in a mediated world thus requires a detailed examination of the specific visual mechanisms, physiological and optical, that direct human depth perception.

PERCEPTIVE IMMERSION AND DEPTH PERCEPTION

The world is around me, not in front of me.

Maurice Merleau-Ponty's "Eye and Mind" (1961)

From a perspective in cognitive psychology, visual depth perception can be analyzed as a modular process that, while possessing a degree of functional autonomy, is deeply integrated into a broader multimodal architecture. Although the early stages of depth extraction—such as binocular disparity and monocular cues—operate through relatively specialized sensory channels, the resulting phenomenal experience is the product of continuous cross-modal integration and top-down cognitive modulation.

Depth perception does not function in a vacuum; it is fundamentally interdependent with the vestibular and proprioceptive systems, which provide the essential egocentric frame of reference required to map visual space. For instance, the perception of an object's distance is constantly calibrated against the body's posture and balance, ensuring that spatial representations are motor-relevant. Furthermore, this process is heavily influenced by executive functions: selective attention filters the visual field to prioritize specific depth planes, while long-term memory provides the schematic templates necessary for size constancy and familiar size cues. Consequently, what appears to be a direct sensory intake is, in fact, a complex inferential process where bottom-up retinal signals are synthesized with multi-sensory data and prior cognitive expectations to construct a stable three-dimensional environment.

According to the taxonomic framework proposed by Sekuler and Blake (2006), see Figure 1.1, depth cues are conceptually organized into three primary dimensions: pictorial vs. oculomotor, monocular vs. binocular, and static vs. dynamic. These classifications reflect how the visual system utilizes modular sensory inputs and physiological feedback to construct a coherent three-dimensional representation of the environment.

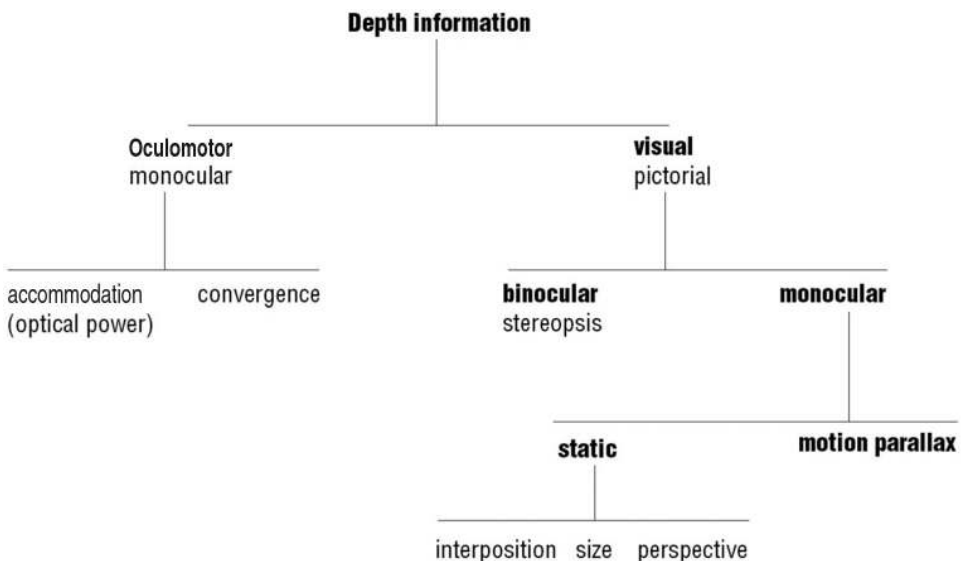


FIGURE 1.1 Depth information cues. (Adapted from Sekuler & Blake 2006.)

Pictorial vs. Oculomotor Cues

Pictorial cues are information-rich features available in a two-dimensional retinal image, such as linear perspective, occlusion, and texture gradients. These are essentially heuristic-based, allowing the brain to infer distance through visual and environmental patterns. In contrast, oculomotor cues are non-visual in origin; they rely on kinesthetic feedback from the muscles controlling the eyes. This involves accommodation (the changing shape of the lens) and convergence (the inward rotation of the eyes), providing the cognitive system with physical data regarding the proximity of an object.

Monocular vs. Binocular Cues

The distinction between monocular and binocular cues highlights the difference between single-channel processing and stereoscopic integration. Monocular cues can be processed using only one eye and are sufficient for perceiving depth in distant landscapes or static images. Binocular cues, specifically binocular disparity, rely on the slight horizontal shift between the images captured by each retina. The cognitive system performs a complex cross-correlation of these two disparate inputs to calculate depth through stereopsis, which is crucial for fine motor tasks and near-space navigation.

Static vs. Dynamic Cues

Finally, cues are categorized by their temporal properties. Static cues provide depth information from a stationary viewpoint, whereas dynamic cues—most notably motion parallax—depend on the relative movement between the observer and the environment. As the observer moves, objects at different distances displace across the retina at varying velocities. The visual cortex processes these velocity gradients to create a vivid sense of depth, a process that is highly integrated with the vestibular system to maintain spatial orientation during locomotion.

PICTORIAL CUES

Pictorial cues are static, monocular sources of visual information derived from the two-dimensional retinal projection. These cues serve as the primary heuristic set that the visual system employs to reconstruct a three-dimensional environment from a flat, proximal stimulus. This process involves complex inferential mechanisms where the brain interprets environmental features to form impressions of depth and spatial layout. The perception of depth derived from these cues is fundamentally organized within a spatial coordinate system known as an egocentric frame of reference. In this system, the origin or “zero point” is the observer themselves, meaning that distance and direction are calculated relative to the subject’s own body and point of view. This is critical for motor-action planning, such as reaching for an object or navigating through a doorway.

In contrast, the allocentric frame of reference specifies the location of objects independently of the observer’s current position, encoding spatial relationships between objects within the environment (e.g., “the chair is next to the table”). Cognitive processing involves a constant translation between these two frameworks: while pictorial cues provide the raw data for egocentric depth judgments, the brain integrates this information with memory and environmental landmarks to maintain a stable, allocentric map of the world.

Consequently, depth perception inherently necessitates a judgment of distance relative to the observer as well as the relative distance between two distinct objects; that is, it involves a comparative assessment of the spatial interval separating elements within the visual field.

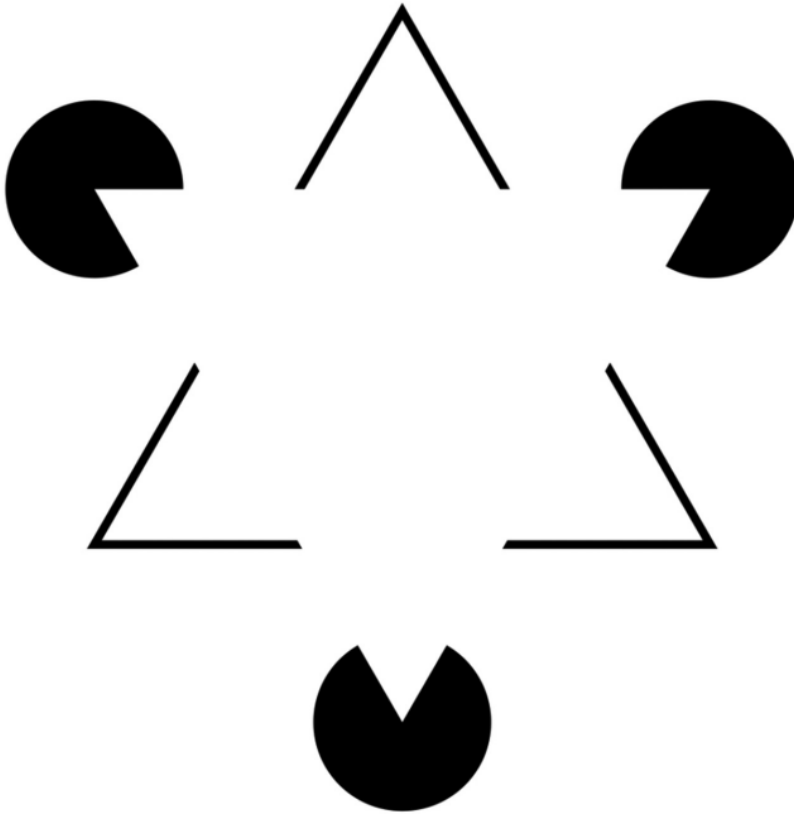


FIGURE 1.2 Kanizsa's triangle optical illusion.

Interposition

Occlusion (or interposition) refers to the phenomenon where an object positioned closer to the observer partially or completely obscures the view of another object situated further away along the same line of sight. From a cognitive perspective, occlusion is one of the most robust monocular depth cues where the contours of the foreground object intersect with those of the background, to infer a clear depth ordering. This “depth-from-occlusion” mechanism allows the brain to establish a relative spatial hierarchy, even in the absence of other cues like shading or perspective.

Figure 1.2 presents the Kanizsa Triangle, a renowned optical illusion that demonstrates the potent influence of occlusion as a depth cue. This phenomenon illustrates the visual system's tendency to perform modal completion, where the brain perceives subjective or “illusory” contours. In this case, the arrangement of “Pac-Man” shapes and V-shaped lines generates the vivid impression of a non-existent white triangle superimposed on the other elements. This illusory foreground object is perceived as being closer to the observer, effectively blocking the shapes beneath it and triggering a powerful sense of depth and stratification in a two-dimensional image.

Perspective

Linear perspective (or conical perspective) serves as a geometric projection that maps a three-dimensional environment onto a two-dimensional pictorial plane. From a cognitive and physiological standpoint, this projection closely approximates the image formation process of the human ocular system, which functions as a converging lens system projecting light onto the retina. Consequently, this geometric model

underpins most modern representational technologies, ranging from traditional photography to sophisticated VR systems. The primary characteristic of this perspective is the convergence of parallel lines toward a single vanishing point on the horizon.

It is interesting to note that some animals, such as insects like flies, bees, and dragonflies, possess compound eyes composed of thousands of individual units called ommatidia, each of which captures light from a specific direction. The resulting visual output is an angular mosaic of the environment, rather than a single central projection formed by a lens. In these cases, the outcome is not a classical linear (conical) perspective image, but rather a directional sampling of the visual field.

However, in conical projection, the human crystalline lens converges light to form a retinal image where the convergence of parallel lines toward a vanishing point provides a critical geometric cue. Through this linear perspective gradient, the brain computes the relative distance of objects by interpreting their scale and placement along these converging vectors as indicators of their depth within the three-dimensional environment.

Design guidelines for VR

When designing the scenography of a virtual space, it is essential to prioritize the alignment of objects parallel to the converging vectors that lead toward the primary vanishing points. This strategic spatial arrangement reinforces the observer's sense of depth by leveraging the brain's innate sensitivity to linear perspective. In outdoor urban scenes, for instance, one should avoid irregular alignments of streets and buildings, as these can disrupt the geometric cues the visual system uses to map distance (see Figure 1.3). Similarly, in interior environments, the placement of furniture, floor patterns, or structural beams should be organized to form



FIGURE 1.3 Street from Industrial facility (VR environment).

clear geometries rather than irregular clusters. By reducing the number of competing or conflicting vanishing points, the designer prevents the scene from becoming visually disorganized and confusing. From a cognitive perspective, this architectural clarity minimizes the computational load required for spatial mapping, allowing the user to navigate and perceive the environment with greater ease and immersion.

In the absence of competing depth cues, linear perspective can become the dominant factor in spatial perception, often leading the visual cortex to assign different apparent sizes to physically identical objects.

A classic example of this extensively documented phenomenon that highlights how the brain relies on environmental geometry to calibrate size-distance relationships that exploit this geometric ambiguity is the Ames Room (Wang et al., 2022). In this illusion, a specially constructed trapezoidal room is viewed through a small monocular peephole. This specific viewing condition is critical as it deliberately neutralizes alternative depth cues, such as stereopsis (binocular disparity) and motion parallax, which would otherwise reveal the room's true distorted shape (see Figure 1.4).

The illusion functions by manipulating the scales of the floor, windows, and interior objects to align perfectly with the perspective expected of a standard rectangular room when viewed from a single fixed point. Because the observer's brain assumes the back wall is parallel to them, it incorrectly attributes the difference in the retinal image size of two individuals to their physical scale rather than their actual distance from the viewer. Consequently, one person appears to be a giant while the other appears tiny, demonstrating that when linear perspective is the sole source of information, the visual system prioritizes geometric consistency over physical reality.

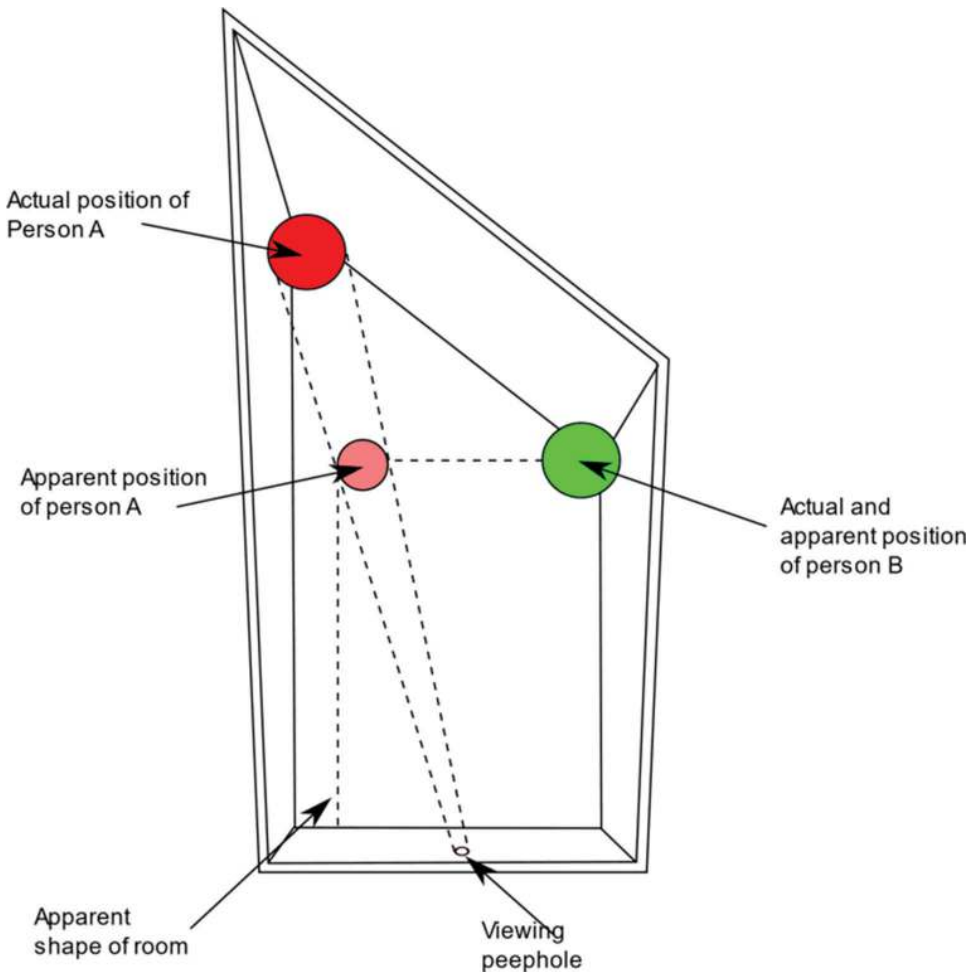


FIGURE 1.4 Ames room optical illusion.

Size

A critical emergent property of linear perspective is the systematic reduction in the retinal image size of objects as their distance from the observer increases.

In cognitive psychology, this relates to the size-distance invariance hypothesis, which suggests that the perceived size of an object is determined by both its retinal size and its perceived distance. While the geometric projection dictates that objects appear smaller as they recede, the human visual system often utilizes these scale changes as a primary depth cue to maintain size constancy, allowing us to perceive an object's true physical dimensions despite its shrinking representation on the retinal mosaic.

Under this mechanism, the visual system retrieves stored knowledge from long-term memory regarding the standard dimensions of a known object. If an object is unfamiliar, the brain lacks reliable clues for comparison, making it significantly more difficult to judge its precise distance. Conversely, if the scale of a familiar object is intentionally distorted, in certain conditions, the brain often misjudges the actual distance; it prioritizes its prior knowledge over the actual sensory input, incorrectly attributing the object's retinal size to its distance rather than its altered scale.

This highlights the inferential nature of depth perception, where the “percept” is a compromise between immediate retinal information and established cognitive templates.

An interesting and well-consolidated empirical phenomenon, though one that still lacks a consensus explanation, is the fact that both size and distance are judged with significantly less accuracy in VR environments compared to identical situations in the real world (Rzepka et al., 2023). This study suggests that poorer reliability of oculomotor cues in VR evokes a relatively greater reliance on familiar size. The brain compensates by shifting its reliance toward higher-level cognitive heuristics. Consequently, the familiar size mechanism gains greater functional relevance in VR contexts, as the visual system leans more heavily on stored mental templates to resolve spatial ambiguities that the digital rendering fails to clarify.

Design guidelines for VR

Given the pivotal role that familiar size plays in depth perception, particularly within VR, designers must exercise extreme precision when assigning scales to 3D assets of immersive environments. This is especially critical when integrating 3D models from distinct asset packs, which often arrive with inconsistent visual styles and varied default scales. Without careful adjustments, these discrepancies can disrupt the user's spatial coherence and break immersion. Furthermore, in scenarios where users encounter unusual architecture or culturally unfamiliar artifacts, the environment can feel spatially ambiguous. To mitigate this, designers should intentionally introduce familiar reference objects, such as common tools, furniture, or vegetation, to serve as “perceptual anchors”. These recognizable elements provide the necessary cues for the brain to accurately estimate size and distance, facilitating a more intuitive and stable recognition of the digital environment.

Texture Gradients

Another vital source of depth information, also emerging from the geometry of linear perspective and the reduction of an object's retinal size over distance, is the texture gradient. Extensively studied by J.J. Gibson in the 1950s, this cue is fundamental to our understanding of spatial layout (Sekuler & Blake, 2006).

Texture gradients occur when surfaces possess repeating pictorial elements—such as consistent sizes, shapes, or orientations—distributed across a scene. As these surface patterns recede into the distance, their density appears to increase while the individual elements become smaller and more closely packed. This systematic transformation provides the brain with rich information regarding the volumetry, curvature, and depth of environmental features (see Figure 1.5). By processing these gradients, the visual system can accurately perceive the slant of a ground plane or the three-dimensional contours of complex objects, even in a static, two-dimensional representation.

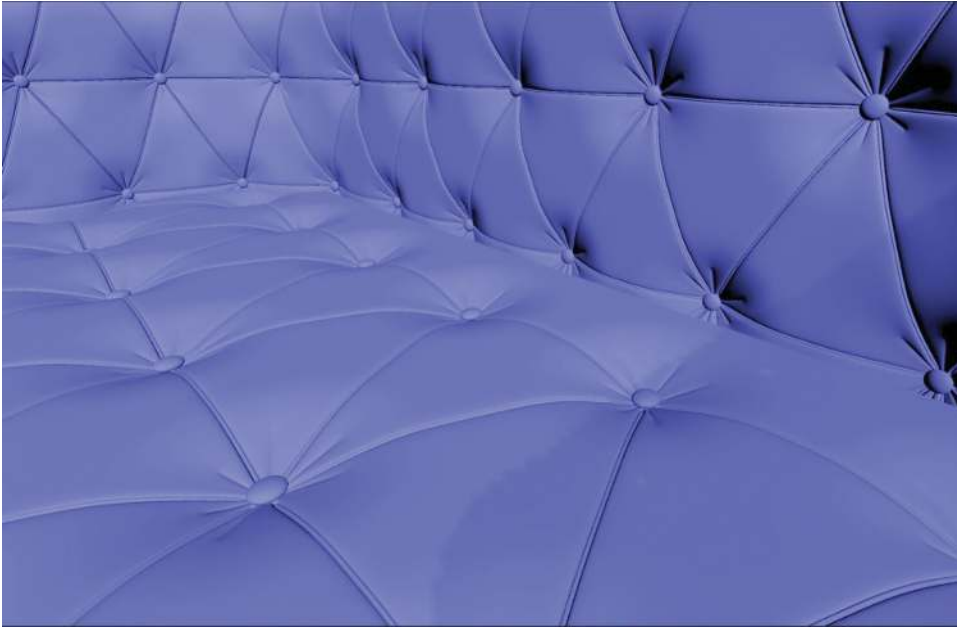


FIGURE 1.5 3D model with a repetitive geometry creates a strong perceptual texture gradient.



FIGURE 1.6 Virtual environment of an old Portuguese traditional home.

Design guidelines for VR

When designing immersive environments, it is essential to prioritize rich, detailed textures with uniform patterns to maximize depth perception. Repetitive patterns are particularly effective, as they provide the visual system with a continuous stream of information regarding an object's volume and underlying geometry.

For instance, in Figure 1.6, we can see that the intricate texture of a fabric cushion does more than define the material; it explicitly describes its volumetric form through the compression and expansion of the pattern. Similarly, a woolen bedspread utilizes its woven texture to reveal wrinkles and subtle surface protrusions, allowing the brain to map the topography of the soft surface. This principle also applies to

rigid architectural elements, such as a wooden accent wall, where the grain and the minute misalignments between individual planks provide essential data regarding the wall's relief and physical presence.

To complement high-quality texture mapping, it is equally important to implement normal maps within the material design process. By simulating how light interacts with surface irregularities, normal maps add a critical layer of realism and detail. From a cognitive perspective, these maps enhance the scene by providing consistent data for both binocular and dynamic depth cues, ensuring that surface depth remains convincing even as the observer moves or shifts their gaze.

Atmospheric Gradients

The Earth's atmosphere is not a perfectly transparent medium. While its presence is negligible at close range, its subtle opacity becomes increasingly evident over greater distances. Solar radiation, comprising the full spectrum of visible light, interacts with the various gas molecules and chemical elements within the atmosphere through a physical process known as Rayleigh scattering. During this interaction, shorter wavelengths, specifically those in the blue component of the spectrum, are scattered more significantly than others, which is why the sky appears blue.

In outdoor environments, particularly in expansive landscapes where the field of view extends toward the horizon, light reflected from distant objects must travel through a greater volume of atmosphere. As a result, these objects lose saturation and contrast, progressively acquiring a desaturated, bluish tint (Aerial Perspective).

Shadows

Shadows serve as a fundamental pictorial depth cue, playing a pivotal role not only in perceiving an object's distance and spatial localization but also in determining critical morphological properties such as orientation, convexity, and concavity (Bruce et. al., 2003).

The visual system utilizes two distinct types of shadows to reconstruct three-dimensional layouts, shading and cast shadows.

Attached Shadows and Shading occur on the surface of the object itself and provide essential information about its 3D volume. By interpreting the gradient of light and dark, the brain can distinguish between a convex surface (bulging outward) and a concave one (hollowed inward), often operating under the "light-from-above" as illustrated in Figure 1.7.

Cast Shadows

These are projected by the object onto another surface. The displacement between the object and its cast shadow is a powerful metric for the brain to calculate the object's height above the ground or its relative distance from background surfaces.

By integrating these shadow cues, the visual cortex can resolve ambiguities in a two-dimensional image, transforming flat shapes into tangible, three-dimensional forms with a clearly defined position in space (see Figure 1.8).

Design guidelines for VR

When developing a digital scene, special attention must be paid to both material properties and shadow configuration. To achieve a high level of spatial fidelity, materials should integrate maps beyond the basic Albedo (color) texture, specifically incorporating normal maps and Ambient Occlusion (AO).

AO is a grayscale map that simulates how exposed a specific part of a surface is to ambient lighting. By darkening occluded areas, such as crevices, holes, corners, and contact points, it generates subtle shadows that reinforce the perception of high and low relief. This process significantly enhances the visual weight and realism of an object within its environment.

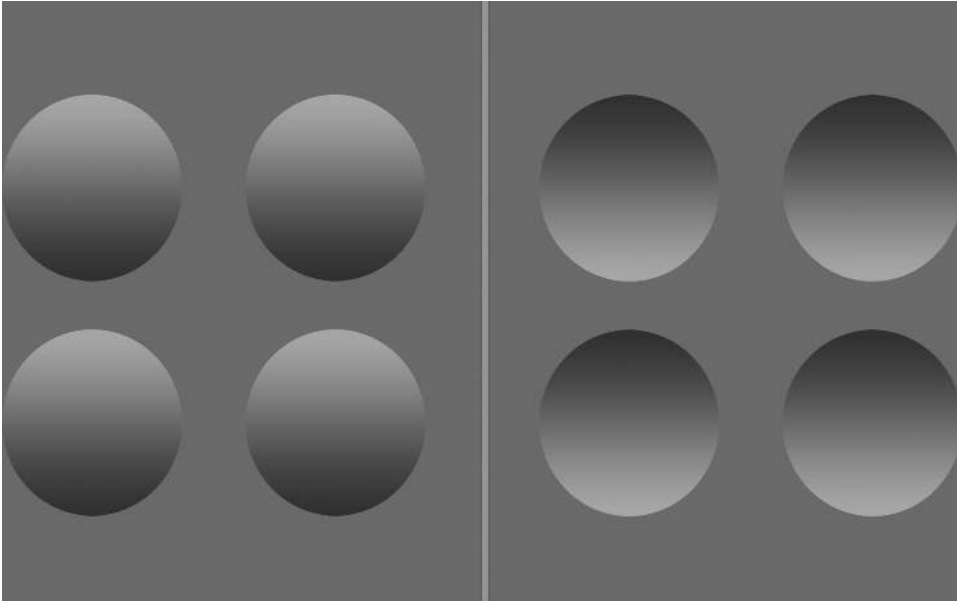


FIGURE 1.7 Shading effect on convexity.



FIGURE 1.8 Shadows produce a strong sense of depth, helping the user locate an object's position, especially under ambiguous conditions.

Furthermore, designers should exercise caution regarding the number of light sources that project shadows. Allowing too many active shadow-casting lights can result in multiple, overlapping shadows for a single object. This redundancy creates a cluttered and ambiguous visual field, which disrupts the user's ability to clearly perceive the object's orientation and spatial relationship to the ground plane. By streamlining lighting and shadow behavior, the scene remains visually coherent and maintains its immersive quality.

MOTION PARALLAX

As we move through the world, objects shift positions relative to one another, creating a phenomenon known as motion parallax. This effect is a constant part of our daily experience, whether we are walking, traveling by train, or looking out of a car window. It serves as a powerful dynamic depth cue, allowing the brain to calculate distances based on the relative velocity of objects in the visual field.

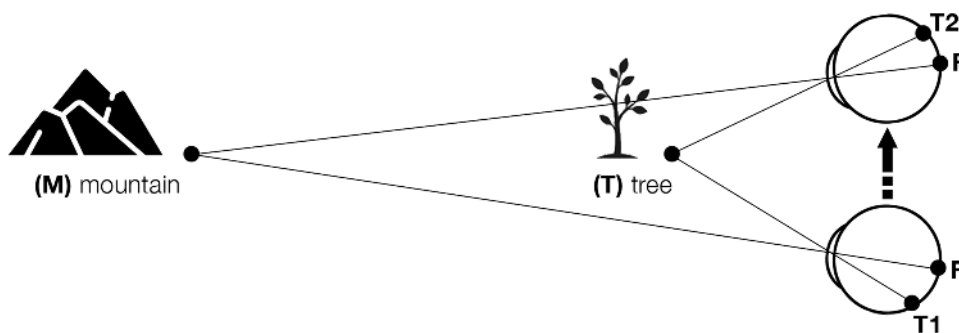


FIGURE 1.9 Motion parallax.

The core principle of motion parallax is that objects closer to the observer appear to move faster across the field of vision (such as trees along a roadside), while more distant objects appear to move much more slowly (such as mountains on the horizon). One particular aspect of this mechanism occurs when the observer fixes their gaze on a specific point in the moving landscape: objects located beyond that fixation point appear to move in the same direction as the observer, whereas objects closer than the fixation point move in the opposite direction.

These concepts are illustrated in Figure 1.9, which depicts an observer looking out of a side window while traveling in a car.

If the observer fixes their gaze on the mountains on the horizon, these distant objects remain stationary on the retina, corresponding to Point F (the foveal region). In contrast, closer objects—such as the tree labeled T—are projected onto different retinal positions (T1 and T2) as the observer moves. The magnitude of this retinal displacement is inversely proportional to the object's distance: the further the tree is from the observer, the smaller the resulting displacement. By processing these shifts in retinal position over time, the brain accurately estimates the relative distances between various moving elements within the scene.

This dynamic process provides highly precise information about the relative location and volume of objects. It is so effective that we often instinctively trigger it; when encountering an unfamiliar object, we frequently move our heads from side to side or circle around it. This behavior is a subconscious attempt to generate new perspectives and utilize motion parallax to better understand the object's three-dimensional structure.

The expressive power of this cue has been widely exploited by visual artists and animators. Disney Studios pioneered the use of the multiplane camera to simulate these layers of depth in traditional animation. Today, this technique remains a staple in 2D platforming video games, where multiple background layers scroll at different speeds to create a rich, immersive sense of space within a two-dimensional medium.

Design Guidelines for VR

Modern VR Head-Mounted Displays (HMDs) integrate advanced tracking systems that monitor head movement with six degrees of freedom (6DoF), encompassing both rotational and translational motion. This technological capability allows users to naturally engage in exploratory behaviors, such as leaning in or moving laterally, to trigger motion parallax as they would in the physical world.

However, VR environment design must be approached with caution when simulating scenarios where the virtual scene moves independently of the user's physical body. A prime example is the previously described experience of a passenger looking out of a car or train window. In such cases, the visual system perceives high-velocity movement and depth gradients, while the vestibular system (the inner ear's balance mechanism) remains stationary.

This discrepancy creates a profound sensory conflict, as the brain receives contradictory data regarding the body's state of motion. Such conflicts are a primary trigger for motion sickness (or cybersickness), a topic that will be explored in greater depth. To maintain user comfort, designers must balance the richness of dynamic depth cues with the physiological constraints of the human sensory system.

BINOCULAR DEPTH PERCEPTION—STEREOPSIS

Stereopsis is a binocular depth cue derived from the simultaneous integration of visual information from both eyes. This phenomenon was famously investigated in the nineteenth century by Charles Wheatstone (1838), who developed the first stereoscope. His device demonstrated how the brain processes two distinct monoscopic images—one from each eye—and fuses them into a single, cohesive three-dimensional percept. Wheatstone realized that because our eyes are horizontally separated, light reflected from a 3D object falls on slightly different locations on each retina. He termed the resulting process of mental fusion stereopsis, from the Greek words *stereos* (“solid”) and *opsis* (“vision”). The fundamental principle behind stereograms is to ensure that each eye perceives only its corresponding two-dimensional monoscopic view. As shown in Figure 1.10, a simple stereogram can demonstrate this effect without the need for specialized hardware. These “stereo pairs” can be viewed using two primary manual techniques:

- **Cross-Eyed Viewing (Cross-View):** This technique involves crossing the eyes to focus on an imaginary point between the observer and the image. By aligning the two reference dots, the observer forces the brain to fuse the two images into a central, three-dimensional composite. In this configuration, the depth is perceived vividly; for example, the letter C may appear significantly closer than the letter P.
- **Parallel Viewing (Wall-Eyed):** In this method, the observer holds the image at a distance and relaxes their gaze as if focusing on the horizon (infinity). By looking “through” the image, a third central image emerges. Once the reference dots are overlapped and the images are fused, the depth relationship is often reversed compared to the cross-eyed method: the letter P becomes prominent and appears closer, while the letter C recedes into the background.

Through these techniques, the brain’s ability to calculate depth from retinal disparity—the minute differences between the two images—is made consciously apparent, providing a vivid sense of “solid” volume.

Humans possess a variable interpupillary distance (IPD), which averages approximately 65 mm in adults. While this distance is smaller in children, it typically begins to stabilize around the age of four. This physical separation between the eyes is what enables the brain to receive two distinct perspectives of a given object, providing the slight horizontal offsets necessary for depth calculation. Following the principles of other stereoscopic technologies, VR HMDs mimic this natural interpupillary offset. By positioning independent displays (or discrete regions of a single screen) within the field of view of each eye, these devices deliver two unique perspectives that correspond to the user’s binocular vision.

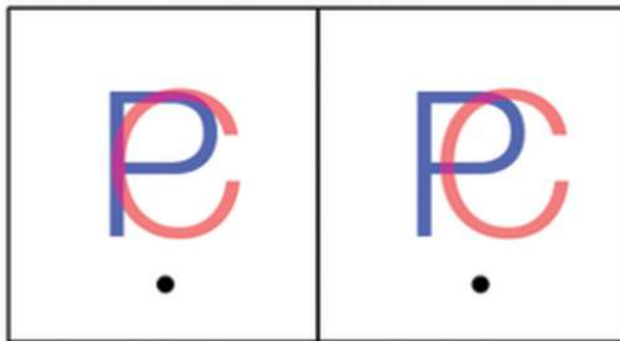


FIGURE 1.10 Free view stereogram.

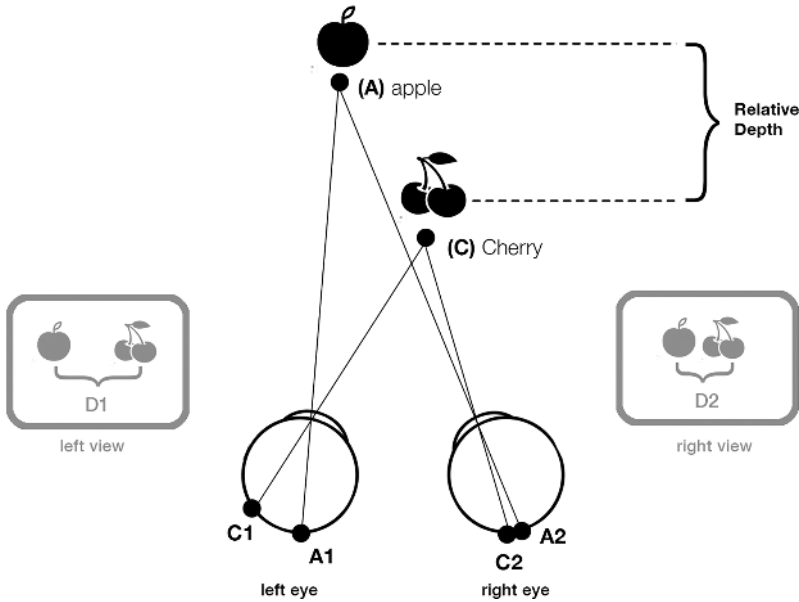


FIGURE 1.11 Retinal disparity is proportional to relative depth.

Retinal Disparity

Retinal disparity refers to the difference in the lateral separation of the same object as perceived by the left and right eyes. As illustrated in Figure 1.11, the magnitude of this disparity is directly proportional to the relative depth of the objects within the visual field.

At the bottom of the diagram, the left and right eyes are represented alongside their respective monocular views. In the upper portion, two objects, an apple and a cherry, are positioned at different distances from the observer. Suppose the observer fixes their gaze on the apple (Point A). In this case, the apple is projected onto the foveal regions (A1 and A2) of both the left and right retinas. Because the cherry is positioned closer to the observer than the apple, its image falls on peripheral retinal locations C1 in the left eye and C2 in the right eye.

Crucially, the lateral separation between the two fruits in the left eye's view (D1) is significantly larger than the separation in the right eye's view (D2). This geometric difference demonstrates that the images captured by each eye are slightly distinct; the resulting retinal disparity, the mathematical difference between these two projections, provides the brain with the precise data needed to calculate the relative depth and three-dimensional layout of the scene.

When two objects exhibit zero retinal disparity, it indicates they are situated at the same relative distance from the observer. This imaginary spatial curve, which encompasses all points that project onto corresponding retinal locations, is known as the horopter.

In Figure 1.12, when an observer fixes their gaze on Point A, all objects situated at that same depth, including Point B, will possess zero retinal disparity. This occurs because the lateral separation of the object on the left retina (D1) is identical to its separation on the right retina (D2). This principle applies to any point plotted along the arc of the horopter.

Spatial depth is further categorized by how objects sit relative to this plane. Objects located closer to the observer than the horopter create crossed disparities, as their images shift outward on the retinas, while objects located further away result in uncrossed disparities, where their images shift inward. By processing these specific directional offsets, the brain can instantly determine whether an object is positioned in front of or behind the current point of fixation (Sekuler & Blake, 2006).

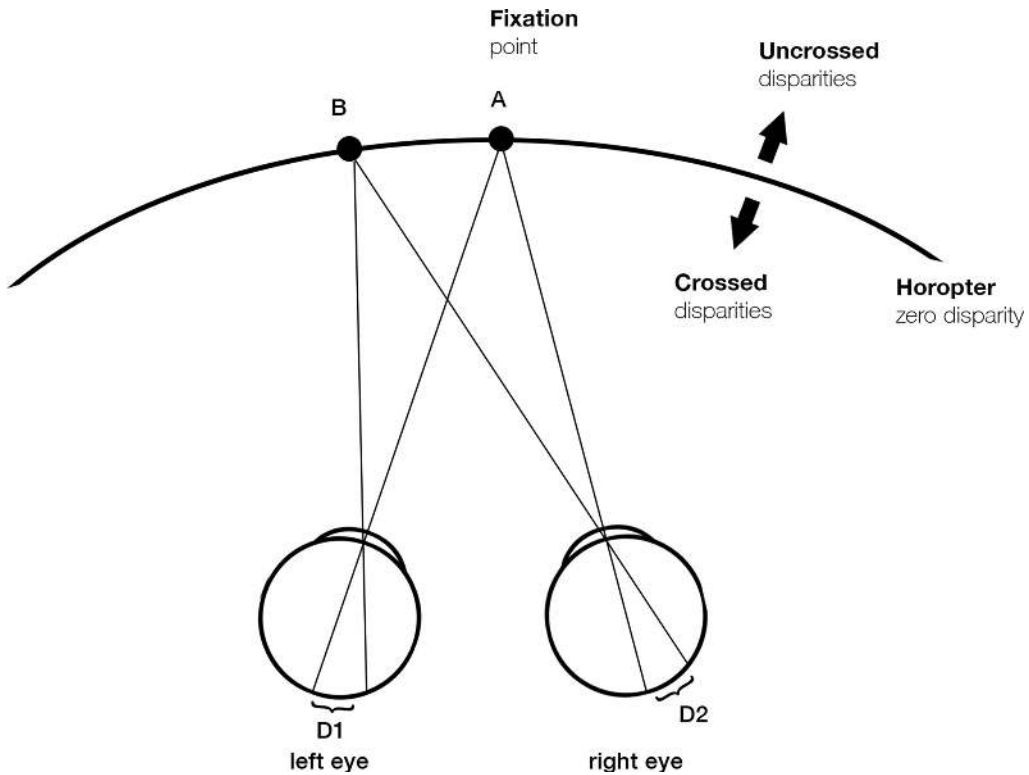


FIGURE 1.12 Horopter plane.

Design guidelines for VR

As suggested by Allison et al. (2009), the effective range of stereopsis as a depth cue remains a subject of debate within the cognitive psychology community. Nevertheless, a broad consensus across various studies indicates that stereopsis serves as a critical source of depth information for distances up to approximately 30m, beyond which its contribution becomes increasingly negligible.

This threshold is a vital consideration for the design of immersive virtual environments. It establishes a practical boundary for the “action space”—the zone where the user is most likely to interact with objects. Within this 30-m radius, designers should prioritize high-fidelity 3D modeling and precise scenography to reinforce the user’s sense of spatial volume and presence. Conversely, for elements located at greater distances, more resource-efficient two-dimensional assets can be utilized. This includes the use of skyboxes, level-of-detail optimization techniques or distant matte paintings, which provide essential environmental context without the computational overhead of true 3D geometry, as the brain’s ability to detect binocular disparity at those ranges is significantly diminished.

OCULOMOTOR DEPTH CUES

Finally, in the study of depth perception, oculomotor cues represent a unique category of spatial information derived directly from the physical sensations of the eye muscles. Unlike pictorial cues, which are based on the visual properties of an image, these cues rely on the brain’s ability to sense the mechanical state of the visual apparatus.

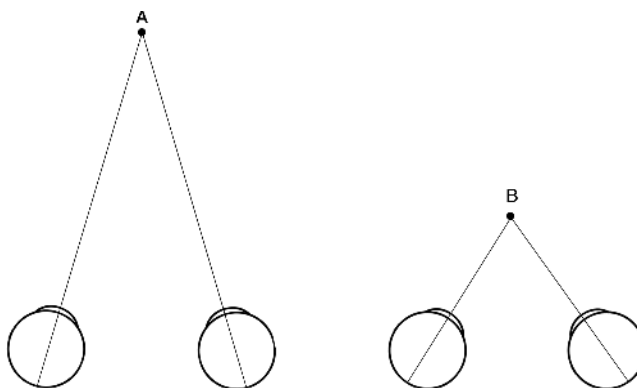


FIGURE 1.13 Vergence oculomotor depth cue.

Accommodation

Accommodation is the process by which the ciliary muscles adjust the shape of the eye's crystalline lens to bring an object into sharp focus on the retina. When an observer looks at a nearby object, the lens must become more curved (thicker) to increase its refractive power; conversely, for distant objects, the lens flattens. The brain monitors the amount of muscular effort required to achieve this focus, using it as a biological metric to estimate the object's distance. However, this cue is only effective at relatively short ranges, typically within 2–3 m.

Convergence

Convergence refers to the inward rotation of the eyes required to maintain a single binocular image of an object as it approaches the face. As we can see in Figure 1.13, the closer an object is, the greater the degree of inward turning performed by the extraocular muscles.

The brain interprets the tension from these muscles as a highly reliable indicator of depth. Like accommodation, convergence is most effective for objects within close proximity, loses its precision beyond 6 m, and becomes virtually negligible as a depth cue at distances exceeding 10 m.

REFERENCES

- Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, 9(1), 10–10. <https://doi.org/10.1167/9.1.10>.
- Berger, J. (1972). *Ways of seeing*. Penguin Books.
- Bruce, V., Green, P. R., & Georgeson, M. A. (2003). *Visual perception: Physiology, psychology and ecology* (4th ed.). Psychology Press. <https://doi.org/10.4324/9780203427248>.
- Ermi, L., & Mäyrä, F. (2005). Fundamental components of the gameplay experience: Analysing immersion. In *DiGRA '05-Proceedings of the 2005 DiGRA International Conference: Changing Views: Worlds in Play*, Vancouver, British Columbia, Canada.
- Jerald, J. (2015). The VR book: Human-centered design for virtual reality. In M. T. Ozsu (Ed.), *The VR book* (1st ed.). Association for Computing Machinery. <https://doi.org/10.1145/2792790>.
- Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2). <https://doi.org/10.1111/J.1083-6101.1997.TB00072.X>.

- Martins, S., & Vairinhos, M. (2023). Ludic and narrative immersion in virtual reality exposure therapy to animal phobias: A systematic literature review. *Virtual Worlds*, 2(4), 303–325. <https://doi.org/10.3390/virtualworlds2040018>.
- McMahan, A. (2003). Immersion, engagement, and presence: A method for analyzing 3-D video games. In M. J. P. Wolf & B. Perron (Eds.), *The video game, theory reader* (pp. 67–86). Routledge, Taylor & Francis Group.
- Murray, J. H. (2016). *Hamlet on the holodeck: The future of narrative in cyberspace*. The MIT Press.
- Ryan, M.-L. (2015). *Narrative as virtual reality: Revisiting immersion and interactivity in literature and electronic media* (2nd ed.). Johns Hopkins University Press.
- Rzepka, A. M., Hussey, K. J., Maltz, M. V., Babin, K., Wilcox, L. M., & Culham, J. C. (2023). Familiar size affects perception differently in virtual reality and the real world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1869) 20210482.
- Sekuler, R., & Blake, R. (2006). *Perception* (5th ed.). McGraw Hill Higher Education.
- Skarbez, R., Brooks Jr. F. P., & Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Computing Surveys (CSUR), ACM Journals*, 50(6), 1–39. <https://doi.org/10.1145/3134301>.
- Slater, M. (2003). A note on presence terminology. *Presence Connect*, 3. <http://publicationslist.org/data/melslater/ref-201/a%20note%20on%20presence%20terminology.pdf>.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549. <https://doi.org/10.1098/RSTB.2009.0138>.
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6), 603–616. <https://doi.org/10.1162/pres.1997.6.6.603>.
- Wang, S., Sanches de Oliveira, G., Djebbara, Z., & Gramann, K. (2022). The embodiment of architectural experience: A methodological perspective on neuro-architecture. *Frontiers in Human Neuroscience*, 16, 833528.
- Wheatstone, C. (1843). Contributions to the physiology of vision—Part the first. On some remarkable and hitherto unobserved phenomena of binocular vision. *Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London*, 4, 76–77. <https://doi.org/10.1098/rspl.1837.0035>.
- Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F. R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., & Jäncke, P. (2007). A process model of the formation of spatial presence experiences. *Media Psychology*, 9(3), 493–525. <https://doi.org/10.1080/15213260701283079>.
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3), 225–240.

Gestalt and Spatial Composition

2

Rocío Cifuentes-Albeza and
Francisco-Julián Martínez-Cano

ABSTRACT

This chapter explores the application of visual organisation principles within the 3D immersive environment of virtual reality (VR): proximity and similarity, among other principles, to support group formation; the figure-ground relationship in volumetric staging; the articulation of depth cues and lighting as a guide; framing, enunciation, and spatial hierarchy; as well as legibility, distance, scale, and typography in VR. This content is structured by defining the classical laws of Gestalt to analyse their practical application in the design of 3D immersive VR content. The findings conclude that effective spatial organisation and the ‘Principle of Invariance’ are crucial for maintaining structural coherence. Ultimately, applying these psychological foundations allows designers to construct stable, legible, and intuitive virtual realities, ensuring that the architecture of attention enhances the sense of presence while optimising the user’s mental resources in complex digital ecosystems.

INTRODUCTION

Virtual reality (VR) constitutes not merely a technological advancement but a fundamental challenge to our understanding of human perception. By immersing a user within an artificial environment, we are not simply projecting imagery; rather, we are constructing a novel reality, distinct from the known, which the brain must interpret in a coherent and stable manner. The designer shapes the architecture of the user’s attention, optimising mental resources to facilitate effective immersion and a sense of presence (Mazhar et al., 2025). To achieve this, immersive design finds its foundation in Gestalt Psychology—specifically in the profound understanding of its described principles (Figure 2.1)—applying them within three-dimensional (3D) design and VR. VR possesses the potential to enhance cognitive capacities for comprehending complex data sources through immersion, the sense of presence, spatial awareness, natural interaction, and sensory feedback (Widjojo et al., 2017).

In conjunction with Gestalt laws, visual language serves as the cornerstone of creation within the field of design, comprising principles and rules that the designer must master to foster their capacity for visual organisation (Wong, 1993). Gestalt theory acts as a basis for multisensory design because it offers a robust framework for creating environments where the global perception of information is clear and efficient, regardless of the sensory channel employed for its transmission (Chang et al., 2007).

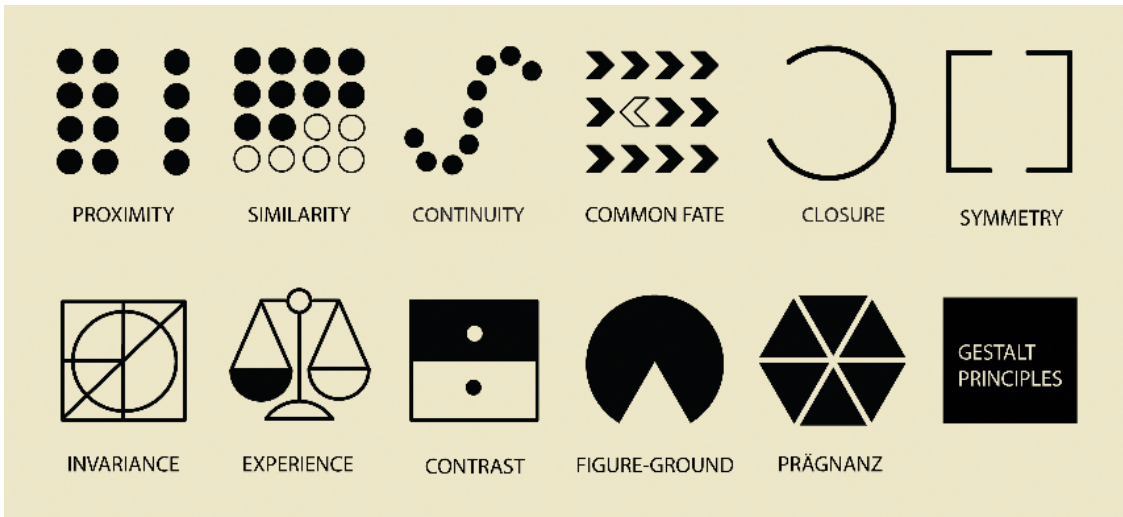


FIGURE 2.1 Gestalt principles. (Credit: Rocío Cifuentes-Albeza.)

GESTALT IN 3D: GROUPING

The human brain does not process isolated data points; rather, through perceptual processes, it seeks patterns that provide coherence to our experience, thereby averting any form of 3D chaos. The application of Gestalt principles within 3D space can guide user attention, enhance navigation, and cultivate a significantly more robust sense of presence without the requirement for explicit instructions. The mechanisms of sight involve two fundamental perceptual processes: unification and segregation (Jerald, 2016, p. 231). The former, unification, is mediated by perceptual grouping, which enables the consolidation of disparate elements into complete objects, or ‘cognitive units’ (Rosli & Cabrera, 2015). The latter serves to demarcate the boundaries of an object or cluster, distinguishing the entity or group from its surrounding environment.

Regarding group formation, several Gestalt laws elucidate this phenomenon: the laws of proximity, symmetry, similarity (MacNamara, 2017), and continuity.

Proximity and Symmetry

The Law of Proximity posits that elements situated in close spatial relation tend to be perceived as a collective whole. In the visual grouping of individual elements, the perceptual mechanism tends to interpret them as a unitary structure; although the constituent 2D or 3D forms may be individually discernible, the perceptual result is the sum of these parts.

Similarly, the Law of Symmetry acts as an integrating principle, as humans more readily discern symmetrical elements. Symmetry may arise from the internal structure of a figure or object, or as a consequence of the spatial arrangement of multiple objects articulated compositionally. Even when these elements are separated or interspersed among other objects, symmetrical components are perceived as a grouped unit. This pursuit of pattern-based harmony facilitates the comprehension of perceived stimuli. This is further described by the Law of Prägnanz (the principle of good figure), which suggests that we organise reality in the simplest, most stable, and most organised manner possible. Attributes such as symmetry, closure, and contrast enhance ease of appreciation and recall, playing a significant role in the design of consolidated structures. The brain consistently opts for interpretations involving lower information loads and minimal cognitive effort.

Similarity and Continuity

The Principle of Similarity states that elements sharing common characteristics are perceived as a group. Scene design should be guided by this principle, categorising elements by shared attributes (Liang et al., 2024). These attributes range from morphological similarity (shape, geometric vs. organic forms) to chromatic similarity. The latter may involve various dimensions of colour, such as surface distribution, chromatic key, light intensity, or spectral zones (e.g., cool vs. warm tones). Similarity may also stem from the level of iconicity (realistic, stylised, or abstract), semantic dimensions, or the iconic specimen of the 3D object (sign, symbol, icon, or index). Furthermore, the Principle of Continuity involves grouping objects arranged in alignment, perceived as a continuous line or row.

Practical Application in 3D Design

The formation of groups via proximity, similarity, symmetry, and continuity aims for cohesion, based on the premise that the global configuration transcends the individuality of its components. In 3D content design, these principles enhance usability and user experience (UX) while establishing visual hierarchies.

Consider the design of a controller comprising various buttons. These repetitive elements are arranged not randomly, but according to a governing norm of repetition, often subordinated to a structure such as a basic grid. Both the element (the basic unit of repetition) and the norm (the ordering logic) facilitate group formation. Ordered composition and constant spatial intervals are critical to the perceptual process of grouping.

In addition to spatial proximity, semantic information—such as numbering or related operational clusters—must follow a logical sequence for the user. For instance, a numeric keypad follows conventions dictated by the reading system of a specific social context. In Western contexts, this follows a left-to-right and top-to-bottom order. Similarity in form and colour can further establish ‘kinships’ or differentiated groups. Ultimately, grouping inventory tools within a single panel facilitates efficient and rapid localisation.

FIGURE-GROUND IN VOLUMETRIC STAGING

The figure-ground principle posits that every figure is situated against a background, or, more precisely, every object is positioned within an environment. The human brain tends to categorise these elements—figure and ground—individually and sequentially, as it is perceptually impossible to perceive a single element as both figure and ground simultaneously. Within the visual field, information is typically organised into two planes: the figure, which captures perceptual attention, appears in the foreground and possesses well-defined contours; whereas the ground comprises everything surrounding the figure, appears to recede behind it, and lacks distinct boundaries. We rely on specific cues to discern which element functions as the figure and which as the ground, a relationship that is, moreover, reversible. These cues include size, contrast, position, and simplicity, among other attributes. In the context of VR, the segregation of figure and ground must be managed through depth and volume.

Figure-Ground Segregation by Depth

In a volumetric VR scene, spatial distance facilitates the distinction between figure and ground. The figure must occupy a depth plane distinct from that of the background. For instance, when an object in close proximity to a wall shares a similar texture, the brain requires increased processing time to achieve discrimination. Conversely, when an object occludes part of the background, the brain instantaneously confirms the object’s predominant role over the environment, resulting in immediate figure-ground segregation.

Contrast

Contrast is essential for highlighting objects (figures) within a given environment (ground), establishing the ‘foreground-background’ effect and thereby enhancing figure-ground segregation. In this process, the characteristics of both figure and ground must be precisely defined to ensure the visual hierarchy is correctly established (for example, highlighting the figure through specific illumination and high-intensity chromatic tones against a monotonous background). The greater the contrast between two areas, the higher the probability that one will emerge as the figure and the other will recede as the ground (Schmidt & Delazari, 2013).

Conversely, intentional similarity between figure and ground can blur this distinction, creating a form of camouflage that integrates them into a single plane. This rhetorical visual play can be achieved through colour, texture, and lighting. Ultimately, contrast plays a fundamental role in the aforementioned perceptual process.

Volumetric Negative Space

In a volumetric scene, negative space is constituted by the ‘air’—an empty space devoid of content that surrounds the 3D object. For a 3D object (figure) to stand out within its environment (ground), a 3D ‘buffer zone’ or ‘area of respect’ is required around it. In contrast, an environment over-saturated with objects leads to increased perceptual fatigue, as the distinction between figure and ground results in a loss of hierarchical order. This may be further exacerbated by visual noise, potentially inducing cognitive overload.

In conclusion, for effective figure-ground discrimination, the 3D object must maintain a clear visual hierarchy through physical depth, luminous contrast, and appropriate spatial isolation.

DEPTH CUES AND LIGHTING AS GUIDANCE

The central visual field is the area of human vision corresponding to the centre of the gaze, where visual acuity and colour perception are at their maximum. Unlike peripheral vision, the frontal field is processed in the fovea centralis, a small region of the retina containing a dense concentration of cones—the photoreceptor cells responsible for light and colour. Eye tracking serves as the technical bridge ensuring that central vision consistently receives the necessary information to validate object invariance, regardless of the complexity of the viewing angle. VR systems leverage eye tracking to maintain structural invariance through foveated rendering, thereby ensuring that the critical details of a 3D object—much like the intricate textures of sculptures by the Baroque master Gian Lorenzo Bernini—are always processed with maximum acuity by the user’s central vision.

The relationship between the gaze and Gestalt theory constitutes the core of selective attention in VR. While Gestalt laws describe how elements, objects, and the environment are organised at a perceptual level, the gaze is the active mechanism that determines which elements enter this organisational process. Consequently, the gaze functions as an exploratory mechanism, whereas Gestalt laws are intrinsically linked to the system of perceptual organisation.

The gaze acts as a selector for perceptual organisation between figure and ground and activates the visual hierarchy. Given that it is impossible to perceive all stimuli simultaneously, the gaze represents a selective form of visual organisation. It triggers the perceptual process centred on figure-ground discrimination, bringing the figure into the foreground while the ground recedes perceptually, moving to the secondary level of the visual hierarchy, as detailed previously in the ‘Figure-Ground in Volumetric Staging’ section.

The Perception of Depth

The construction of depth perception must be addressed with rigorous consistency to avoid fracturing the sense of presence within the 3D immersive VR experience. Several critical factors require consideration, including atmospheric perspective, inter-object occlusion, and motion parallax.

- **Atmospheric Perspective and Contrast:** Atmospheric perspective assists the brain in processing distances; through the gradation of colour or texture, the sensation of perspective is reinforced. Regarding colour, this gradation may arise from shifts in saturation: highly saturated colours appear proximal to the observer, whereas desaturated tones suggest distance. Concerning texture gradient, the level of relief and detail diminishes as an object recedes from the observer, losing visual prominence. Furthermore, the deployment of haze or fog can effectively simulate this effect in virtual landscapes.
- Contrast serves as a powerful instrument for both the discrimination of 3D objects and the creation of depth. As human beings function within a natural medium where contrast is a primary cue, its application in VR environments is vital for establishing spatial depth.
- **Occlusion and the Law of Closure:** Inter-object occlusion is a fundamental tool for depth synthesis (Jerald, 2016). It involves the visual blocking of one entity by another, thereby limiting the view of the partially or totally concealed object. Visibility is only regained through user displacement or a shift in the viewpoint. Perceptually, when one object partially occludes another, the brain confirms that the former is proximal while the latter is situated at a greater distance. This effect, referred to in general image theory as overlapping or *traslapo* (Villafañe & Mínguez, 1996), is essential for generating the sensation of volume.
- Partial concealment does not impede object recognition; however, binocular-occlusion conflicts may arise (MacNamara, 2017) if occlusion is incorrectly implemented, potentially breaking immersion and inducing user discomfort. In high-density VR environments, the recognition of partial objects is supported by the Gestalt Principle of Closure. While the system conserves computational resources by simplifying peripheral objects, the user's mind unconsciously completes these fragmentary structures, maintaining the perception of a coherent, solid world without the need for full 360° high-fidelity rendering. This reconstruction is a product of the human preference for the most parsimonious interpretation of environmental stimuli.
- **Motion Parallax and Directional Vectors:** Motion parallax dictates that proximal objects appear to move at a higher velocity than distal ones. Designers must manage this factor carefully; failure to do so may compromise depth perception and induce kinesis (motion sickness). There is a marked perceptual preference for continuous trajectories over abrupt visual transitions; even when lines are interrupted, the mind tends to bridge the gaps to maintain visual continuity and fluidity.

The Principle of Common Fate refers to synchronised movement; when a group of objects shares a displacement vector, velocity, and direction, they are perceived as a single entity, regardless of their individual alignment or morphological similarity. This directional vector can construct a linear perspective; for instance, when two lines converge towards a vanishing point, the Principle of Continuity compels us to perceive them as a single entity extending towards infinity, generating a sensation of longitudinal depth.

- **Comparison of Continuity and Common Fate:** The principles of Continuity and Common Fate are pivotal in immersive interface design, as they facilitate the creation of mental pathways and reduce cognitive load. Their shared characteristics include:
 - **Directional Grouping:** Both principles assist in interpreting elements that converge towards a singular objective, enhancing the perception of depth and traversal.
 - **Predictive Movement Patterns:** These principles allow the user to anticipate forthcoming events by pre-visualising the path. The distinction lies in the fact that continuity guides the visual trajectory (passive exploration), whereas common fate aids in the comprehension of active events.

- **Reduction of Visual Noise:** In a 360° environment, the user organises information more efficiently by focusing on the ‘common thread’—be it a continuous line or a moving cluster.

In summary, continuity is a static and structural property (e.g., a trail of particles or a curved interface panel), while common fate is kinetic and temporal. In the latter, synchronised movement towards a shared point ensures elements are perceived as a functional unit. Thus, continuity acts as a map to guide the gaze, whereas common fate serves as a ‘call to action’.

Lighting in 3D Environments

Lighting within 3D environments functions as a critical visual perception tool that contributes to the sense of spatiality, depth, and the hierarchisation of content. Surface Gestalt suggests that a user cannot perceive a surface in isolation from its surrounding space (Assadi, 2001).

- **Modelling and Depth Cues:** Illumination provides indispensable visual support both in the modelling of 3D objects and in the generation of depth cues (Jerald, 2016; Gaggioli & Breining, 2001; Schmidt & Delazari, 2013; Liang et al., 2024); this applies to both surface shading and shadows cast upon the environment. Specular highlights serve as potential indicators of dynamic realism in VR environments. These must account for the user’s head movements; if highlights remain static during displacement, the sense of realism is significantly diminished. Specularity facilitates the distinction between the material properties of 3D objects and the environment itself, aiding spatial interpretation. For instance, on a metallic surface, the point of maximum luminance—coupled with diffusion gradients and shading—identifies the trajectory of the light source while defining the object’s boundaries and edges. Both luminance and shading must contribute to depth synthesis and adapt according to the observer’s position and movement to ensure the user perceives a coherent sense of materiality.
- **Cast Shadows and Ambient Occlusion:** Cast shadows are instrumental in mitigating the ‘floating object’ effect, providing objects with spatial coherence within the virtual realm. These shadows provide the observer with data regarding the position and distance of the light source and help validate the physics of the environment when an object intersects with its cast shadow. Furthermore, ambient lighting—referring to light bouncing off walls and floors—contributes to the construction of depth and spatiality. This encompasses techniques such as ambient occlusion, where shading in corners or crevices reinforces the perception of three-dimensionality.

FRAMING, STAGING, AND SPATIAL HIERARCHY

In immersive VR environments, in contrast to cinema and photography, the user is granted total autonomy regarding movement and gaze direction. Nonetheless, through intentional environmental design, it is possible to guide the user’s attention. This objective is achieved through dynamic framing, staging, and spatial hierarchy.

Dynamic Framing: Architectural Framing and 360° Composition

To direct user attention towards specific scene elements, it is imperative to establish a clear visual hierarchy and distinct focal points (Liang et al., 2024). Data information layers must be organised hierarchically, comprising levels of information derived from natural distribution patterns (Moloney et al., 2018).

Environmental elements, such as doorways, windows, or caverns, can function as primary centres of interest. Furthermore, variables such as lighting act as cues to guide the user and establish visual accents within a specific narrative framework. These elements constitute a clear invitation for the observer's gaze, occasionally functioning as directional vectors for locomotion.

Staging: Points of Interest, Guiding Lines, and Comfort Zones

The environment must, by its own merit, communicate the scene to the user. The arrangement of objects and the organisation of spaces must possess a logic capable of elucidating narrative information within the immersive VR space. The Law of Continuity can facilitate the design of pathways that prompt action and encourage intuitive displacement patterns. To mitigate user fatigue, objects should be positioned within the comfort zone; by placing key elements frontally relative to the spectator, the necessity for constant head rotation is minimised, thereby avoiding uncomfortable and non-ergonomic movements.

Spatial Hierarchy: Scale and Dominance, Luminous and Chromatic Contrast, and Interaction Zones

- **Scale and Dominance:** Scalar elements of the image, such as size, scale, and proportion (Villafañe & Mínguez, 1996), establish relational meaning and influence environmental comprehension. In an immersive environment, user displacement functions as an interactive metrical act; it is a process whereby the cognitive system 'measures' and reconstructs the object's geometry through movement (Diettrich, 2001).

Regarding size, objects of greater dimensions typically command a higher sense of importance within the 360° compositional order compared to smaller entities, generally occupying a position of dominance. Hierarchy may be categorised as primary or secondary: Primary Hierarchy: Established in relation to the focal object—such as one intended for user interaction—and the remaining environmental elements. It addresses the initial impact and user interest. Secondary Hierarchy: Established once the user has identified the dominant object and begins to perceive peripheral elements, such as menus, status indicators, or text. It addresses the functional question: 'How can I interact with this dominant object?'

- **Hierarchical levels are constructed through several mechanisms:**
 - **Luminous and Chromatic Contrast:** Highlighted objects typically possess increased illumination or chromatic accents. These are created via saturation, complementary colour contrasts, colour temperature, or luminance contrast (white-to-black ratios).
 - **Interaction Zones:** Egocentric space is organised into three distinct zones based on distance from the user: personal space, interaction space, and horizon space (Jerald, 2016). Personal space (0.5–2 m) implies a high interaction hierarchy, containing objects the user can touch or grasp. Interaction or action space (2–10 m) suggests locomotion and navigation. Finally, horizon space (beyond 10 m) constructs atmospheric hierarchy and spatial orientation.

READABILITY, DISTANCE, SCALE, AND TYPOGRAPHY IN VR

Distance, Scale, and Readability

The Principle of Invariance dictates that an object or figure can be recognised regardless of its spatial disposition; scale, translation, rotation, or elastic deformation do not impede recognition. Humans tend to discriminate between form and structure; the brain does not register static objects but rather abstract geometric structures. Structure underlies the memory of known entities, providing meaning to observations. While the observer's relative position may modify the perceived shape, the internal structure remains constant. Thus, invariance refers to the perceptual capacity to recognise complex patterns.

Invariance facilitates an illusion of solidity in immersive VR, allowing a virtual object to be perceived as a physical body because the brain functions as a model processor. It enables us to extract constant characteristics (e.g., vertices, axes of symmetry, proportions). Certain viewpoints provide more direct information regarding internal structure, facilitating identification. In VR, as the user approaches or recedes from an object, the scale provides the illusion of growth or shrinkage, emulating the natural patterns of stereoscopic vision.

The omnidirectionality of a 360° environment is a fundamental challenge for 3D content. Drawing a parallel with the works of the Baroque sculptor Gian Lorenzo Bernini, a 3D object must maintain structural coherence regardless of the user's viewpoint. Bernini's sculptures offer a basic structure recognisable from any position while simultaneously inviting spatial exploration through intricate detail (Walk Inside Rome, 2019). Consequently, 3D objects in VR can be developed using similar sculptural strategies.

Finally, distance and scale are pivotal. In VR, camera height and object scale are critical; a camera height exceeding 2 m may induce a sense of gigantism, whereas a height below 1.2 m may evoke the perspective of a child. High-frequency detail textures, such as wood grain, serve as essential cues for the user to confirm the veridical scale of an object.

Typography and Readability in VR

In VR, text must be treated as a 3D object; consequently, its legibility from oblique angles and its integration within the spatial environment, among other factors, must be meticulously considered:

- **Selection of Typefaces:** Sans-serif typefaces—particularly those featuring wide apertures (reading fonts) and substantial stroke weights—are well suited for virtual environments. Conversely, serif typefaces are less functional, as the phenomenon of aliasing significantly impairs the legibility of serifs.
- **Chromatic Optimisation of Typography and Background:** Interactive elements must be clearly distinguished from non-interactive ones. Colours must be readily discernible; MacNamara (2017) suggests a hierarchy of preference—red, green, yellow, and white—for highlighting information.
- **Reading Conditions:** The specific conditions under which text is consumed directly influence legibility. These conditions are determined by several factors:

- **Static versus Dynamic Reading:** Reading while stationary versus reading during locomotion.
- **Spatial Orientation and Placement:** To mitigate character distortion resulting from oblique viewing angles—which occur when the observer is not positioned frontally to the text—reading surfaces should be oriented towards the user.
- **Field of View (FoV) and Ergonomics:** optimal text placement should be centralised and slightly elevated within the FoV. It should remain consistent relative to head movement, allowing information to be processed almost subconsciously without disrupting immersion in the primary task (MacNamara, 2017).
- **Reading Distance and Illumination:** Short-, medium-, and long-range distances necessitate specific ergonomic adjustments regarding text size and vertical placement within the environment, alongside consideration for ambient lighting conditions.

The Construction of the Learned Object and Legibility

The Principle of Experience and the Principle of Invariance share a common objective: cognitive survival within VR environments. Memory serves to nourish invariance, as the recognition of a spatial geometric structure is facilitated by past experiences and their subsequent recollection. Memory enables the storage of structural data pertaining to both objects and environments. Furthermore, the Principle of Recall assists the user in validating the underlying structure of an object, permitting a comparison between the perceived structure and previously acquired knowledge.

While invariance performs a verification of proportions and distances, recall is responsible for validating the accuracy of that information. Collectively, these laws establish that form is the variable element that shifts according to the viewpoint (invariance), whereas structure is the invariant component that persists in memory through learning and recollection.

CONCLUSIONS

The analysis of Gestalt within VR design demonstrates that the quality of immersion is not proportional to computational power per se, but rather to the perceptual intelligence applied to spatial composition. By structuring the virtual environment in accordance with the innate principles of visual organisation, the designer optimises technological resources to maximise the sense of presence and shape the UX. Based upon this premise, the following conclusions are presented:

The Prevalence of Invariant Structure

It has been demonstrated that the human brain functions as a processor of structural models rather than a mere recipient of superficial details. In accordance with the Principle of Invariance and the Law of Prägnanz, the stability of an object within VR resides in its essential configuration—specifically its vertices, axes of symmetry, and proportions—rather than its photographic realism. Structural coherence from multiple perspectives constitutes the sine qua non for immersion. Provided the internal structure remains robust, the user maintains the ‘ontological security’ of the object, even when the system economises resources on peripheral textures. Design must, therefore, prioritise the legibility of form over ornamentation.

The Designer as Architect of Attention and Hierarchy

The transition from two-dimensional media to a 360° environment eliminates traditional framing control, thereby delegating autonomy to the user. Nevertheless, this research confirms that Gestalt-based design re-establishes narrative direction through spatial hierarchisation. The strategic application of the Law of Continuity and Common Fate vectors enables the guidance of the foveal gaze without recourse to explicit instructions. By aligning environmental elements—such as architectural frames, vanishing lines, and high-contrast lighting—the designer reduces the user’s cognitive load, transforming navigation from an exhaustive search process into one of intuitive recognition.

The Reduction of User Cognitive Effort (UX)

One of the most critical findings is the direct correlation between perceptual organisation and the mitigation of fatigue. A virtual environment that disregards the laws of visual perception compels the brain into a constant effort of segregation and reconstruction, which leads to stress and visual fatigue. The implementation of ‘buffer zones’ or volumetric negative space, in conjunction with the coherent management of cast shadows and occlusion, validates the physics of the environment for the brain. By minimising sensory discrepancies through visual organisation, the sense of presence is fortified, and the duration of comfortable engagement within the virtual experience is extended.

Readability and Ergonomics in 3D Space

Finally, the integration of textual information and User Interface (UI) elements under Gestalt principles reveals that readability within VR is a purely geometric and contextual phenomenon. The efficacy of visual communication in VR depends upon treating typography as a physical entity. The selection of sans-serif typefaces, the optimisation of chromatic contrast, and, crucially, the orientation of reading surfaces towards the user are imperative to prevent the disruption of immersion caused by decoding difficulties.

In summary, designing for VR is, ultimately, designing for the human mind. The application of Gestalt laws serves as the necessary bridge between software architecture and the psychology of perception. By aligning the capabilities of rendering engines with the innate rules of visual organisation, we ensure that the user does not merely ‘view’ an artificial environment, but inhabits it with the same naturalness and immediacy with which they navigate the physical world.

REFERENCES

-
- Assadi, A. H. (2001, November). Perceptual geometry of space and form: Visual perception of natural scenes and their virtual representation. In R. A. Navarro & J. L. Artal (Eds.), *Vision geometry X* (Vol. 4476, pp. 59–72). SPIE. <https://doi.org/10.1117/12.447288>.
- Chang, D., Nesbitt, K. V., & Wilkins, K. (2007, March). The Gestalt principle of continuation applies to both the haptic and visual grouping of elements. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, Tsukuba, Japan (pp. 15–20). IEEE. <https://doi.org/10.1109/WHC.2007.113>.
- Diettrich, O. (2001, January). Virtual reality and cognitive processes. In G. Joubert (Ed.), *Virtual reality: Cognitive foundations, technological issues & philosophical implications* (pp. 203–219). Springer.

- Gaggioli, A., & Breining, R. (2001). Perception and cognition in immersive virtual reality. In G. Riva & F. Davide (Eds.), *Communications through virtual technology: Identity, community and technology in the internet age* (pp.71–86). IOS Press.
- Jerald, J. (2016). High-level concepts of content creation. In *The VR book human-centered design for virtual reality* (pp. 225–236). Morgan & Claypool.
- Liang, H., Xu, H., Wang, Y., & Pan, J. (2024, July). Integrating design psychology in VR scenarios to improve virtual reality experiences. In *2024 10th International Conference on Virtual Reality (ICVR)*, Bournemouth, United Kingdom (pp. 37–46). IEEE. <https://doi.org/10.1109/ICVR62393.2024.10868876>.
- MacNamara, W. (2017). *Evaluating the effectiveness of the Gestalt principles of perceptual observation for virtual reality user interface design* (Master's thesis, Technological University Dublin). TU Dublin Arrow. <https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1113&context=scschcomdis>.
- Mazhar, A. A., Abduljabbar, O. J., Daoudi, O., Abdallah, M., & Alkhateeb, J. H. (2025). Enhancing virtual communication by exploring user experience design in VR interfaces. *Journal of Information Hiding and Multimedia Signal Processing*, 16(1), 414–422.
- Moloney, J., Spehar, B., Globa, A., & Wang, R. (2018). The affordance of virtual reality to enable the sensory representation of multi-dimensional data for immersive analytics: From experience to insight. *Journal of Big Data*, 5(1), 53. <https://doi.org/10.1186/s40537-018-0158-z>.
- Rosli, M. H. W., & Cabrera, A. (2015). Gestalt principles in multimodal data representation. *IEEE Computer Graphics and Applications*, 35(2), 80–87. <https://doi.org/10.1109/MCG.2015.29>.
- Schmidt, M. A. R., & Delazari, L. S. (2013). Gestalt aspects for differentiating the representation of landmarks in virtual navigation. *Cartography and Geographic Information Science*, 40(3), 159–164. <https://doi.org/10.1080/15230406.2013.807031>.
- Villafañe, J., & Mínguez, N. (1996). *Principios de la Teoría General de la Imagen*. Pirámide.
- Widjojo, E. A., Chinthammit, W., & Engelke, U. (2017, November). Virtual reality-based human-data interaction. In *2017 International Symposium on Big Data Visual Analytics (BDVA)*, Adelaide, Australia (pp. 1–6). IEEE. <https://doi.org/10.1109/BDVA.2017.8114627>.
- Wong, W. (1993). *Principles of form and design*. John Wiley & Sons.

Scenography, Scale, and Lighting in VR

3

Francisco-Julián Martínez-Cano and
Rocío Cifuentes-Albeza

ABSTRACT

This chapter examines the synthesis of scenography design required to construct immersive virtual reality (VR) environments. VR represents a fundamental shift from traditional media, as the user's physical body becomes the absolute metric for the virtual world. Maintaining place illusion requires real-world scaling aligned with perceived affordances. Designers must prioritize the reach envelope. To sustain the plausibility illusion, the text explores the roles of environmental storytelling, semantic expectancy, and a lived-in aesthetic in establishing world coherence. Lighting is analysed as a dual-purpose tool for wayfinding and ocular ergonomics. Finally, this chapter advocates for performance-aware art direction, integrating spatial optimization to maximize visual density within hardware frame budgets.

INTRODUCTION

The organization and design of theatrical settings that integrate lighting, sound, costumes, and set design to create an engrossing story environment have historically been referred to as “scenography” (Howard, 2009). Scenography is a key component of virtual reality (VR) and affects how users interact with and understand the virtual environment (VE). By establishing the place's emotional tone and directing the user's attention, it guarantees that significant plot points are not missed in a fully explorable setting.

Unlike static, non-interactive media where framing is predetermined, VR affords users complete control over their viewpoint. Consequently, the design of scenography must account for the user's agency. Spatial dynamics within VR facilitate exploratory behaviour while maintaining narrative clarity using motion cues, contrast, visual hierarchy, and the strategic placement of landmarks (Grau, 2004). Empirical evidence indicates that users' sense of presence is significantly enhanced by environments that are thoughtfully constructed (Slater, 2018), enabling them to focus on salient elements without experiencing cognitive overload (Steinicke, 2016). As VR technology continues to evolve, the ongoing refinement of these compositional strategies remains crucial for the development of compelling and immersive content.

HUMAN SCALE, REACH ENVELOPES, AND SPATIAL AFFORDANCES

The shift from traditional screen media to immersive VR marks a fundamental change in spatial design. Whereas traditional media relies on camera framing, forced perspective, and lens choices to convey spatial relations, VR eliminates the frame entirely, surrounding the user with the environment. The user's own body—specifically interpupillary distance (IPD) and camera height—serves as the definitive scale reference. Designing VR scenography thus involves creating for an embodied presence, where the user is both participant and spatial metric. This move from screen-space to world-space requires that virtual assets be created at a precise 1:1 real-world scale, except when intentional psychological distortions are intended. This accurate scaling underpins the “place illusion” (PI), defined as the compelling sensation of being present in a place—the sense of “being there”—despite awareness of its virtuality (Slater, 2009, p. 3551). This illusion depends on sensorimotor contingencies. If the scale is mismatched, the user's brain detects a subtle dissonance that shatters presence, replacing it with a sense of “uncanny valley” spatiality. Scale perception is also a function of the virtual avatar; embodying a child-sized avatar causes users to overestimate object sizes compared to using an adult avatar (Banakou et al., 2013, p. 12847).

Beyond visual scale, VR scenography must be constructed around the physiological constraints of the human body, specifically concerning *peri personal space*, the space immediately surrounding the user within which they can comfortably reach and manipulate objects without straining or moving their base. In ergonomic and VR interaction design, this is operationalized as the reach envelope.

When designing interactive virtual sets, the placement of vital interactive elements and objects of importance must respect the user's primary (forearm sweep) and secondary (full arm extension) reach zones. Forcing a user to repeatedly reach outside this envelope not only causes physical fatigue known as the “gorilla arm” effect (Jerald, 2015; La Viola et al., 2017, p. 129) but also disrupts cognitive immersion.

Mine, Brooks, and Sequin highlighted the critical role of proprioception—a person's subconscious sense of their body's position in space—in virtual interactions. They noted that “though proprioception greatly enhances virtual-environment interaction, precise manipulation is still harder in virtual spaces than in real space” (Mine et al., 1997, p. 7) due to the lack of haptic feedback in immersive VE. When a scenographer positions an interactive element just beyond the user's natural reach, the resulting physical strain disrupts immersion by shifting focus to the constraints of the room-scale setup. Therefore, VR scenography must consider the user's reach envelope; failure to allow access to seemingly reachable objects undermines affordance, causing frustration and diminishing immersion (Jerald, 2015).

The physical layout of the virtual set communicates what actions are possible within it. This relies heavily on the concept of affordances, pioneered by Gibson (1979). An affordance is a relation between an object and an organism that affords the opportunity for an action. In a VE, scenography provides spatial affordances. A flat, rigid surface at knee height affords sitting; a brightly lit button affords pressing.

However, in VE, objects do not have the characteristics of physical mass and resistance. Therefore, VR designers depend upon what Donald Norman (1999) then further developed as perceived affordances. Norman clarified that in digital design, “the designer cares more about what actions the user perceives to be possible than what is true” (Norman, 1999, p. 39).

For the VR scenographer, this distinction is critical. The visual design (the perceived affordance) must perfectly align with the interactive programming (the actual affordance). If an environment features a meticulously rendered, photorealistic chair, the user will instinctively perceive the affordance of sitting. If the user attempts to lean on or sit on it, only to fall right through the polygon mesh because no collision bounds were programmed, the psychological contract of the virtual world is broken. This mismatch creates severe “breaks in presence” (BIPs) (Slater et al., 2003, p. 1; Sánchez-Vives & Slater, 2005, p. 335). Consequently, VR scenography demands restraint: objects that cannot deliver their implied affordances should be visually abstracted, positioned beyond reach, or omitted to maintain environmental coherence.

SET DRESSING AND WORLD COHERENCE

In VR, set dressing transcends mere aesthetic decoration; it is the primary vehicle for “plausibility illusion” (Psi), the second pillar of presence. Unlike PI, Psi is “the illusion that what is apparently happening is really happening (even though you know for sure that it is not)” (Slater, 2009, p. 3553). For a world to be coherent, set dressing must satisfy the user’s internal logic and provide a consistent narrative grammar.

Digital environments are navigable spaces (Murray, 1997), making the user an active explorer. The set must be readable, a concept rooted in environmental storytelling where “the story element is infused into the physical space” (Carson, 2000). Every object serves as a data point for the user to construct a mental model of the world’s history. However, information density must be calibrated. Increased visual fidelity enhances presence only when sensory information is consistent (Cummings & Bailenson, 2015, p. 4). An overly dense area next to a sparse one triggers a cognitive dissonance that highlights the simulation’s artifice.

The Logic of Object Placement and the Lived-In Aesthetic

From a functionalist standpoint, world coherence is upheld through “spatial syntax”, the set of rules governing how spaces and objects relate to human movement and social interaction (Hillier et al., 1976, pp. 179–180). In VR scenography, object placement primarily manages users’ cognitive load. According to Sweller’s Cognitive Load Theory (1988), if a VE is disorganized or lacks a recognizable logic in its set dressing, it forces users to expend excessive extraneous cognitive resources on navigation, thereby limiting resources available for narrative engagement or task performance.

In their study on spatial navigation in VE, Ruddle, Payne, and Jones found that “spatial knowledge is acquired most effectively when the environment contains a consistent set of landmarks and a coherent structural hierarchy” (1998, p. 191). For the VR scenographer, this means set dressing must follow a functional hierarchy. A messy desk in a virtual office should be clustered in a way that reflects human activity (e.g., the “mouse” area remains clear, while “clutter” accumulates in the periphery). This logical clustering allows the user to use chunking (Fountain et al., 2022, p. 1402), a cognitive process where a group of objects is perceived as a single unit rather than individual, distracting polygons.

In their research, Sánchez-Vives and Slater defined immersion as “the technical capability of the system to deliver a surrounding and convincing environment with which the participant can interact” (2005, p. 333). The degree of presence is therefore related not only to the extent to which the environment responds to the user but also to the internal consistency of the environment itself. This consistency is achieved through a lived-in aesthetic. Scenographers employ what is known in computer graphics as weathering—digital entropy representation—to avoid the sterility associated with pristine virtual spaces, which often evoke the uncanny valley. Introducing subtle imperfections, such as scuff marks or dust accumulation, enhances plausibility by simulating the effects of time and physical interaction.

The lived-in aesthetic serves less as visual appeal and more as proprioceptive grounding. Jerald highlights that “consistent environmental cues provide the user with a sense of stability” (2015, p. 62). Logical arrangements—such as positioning a trash can near a door or placing light switches within reach—activate users’ pre-existing mental schemas, reducing search latency. Conversely, incongruent placements (e.g., a chair in a hallway’s centre) induce cognitive friction, revealing the environment as artificial. Thus, coherence stems from predictability, grounded in lifelong internalization of socio-spatial norms.

Semantic Coherence and Narrative Priming

Prop selection and placement must correspond to users’ semantic expectancy, whereby the brain anticipates scene content based on categorical frameworks. In VR, deviations from these expectations absent

narrative rationale disrupt world coherence, consistent with schema theory, which posits that environmental processing relies on stored knowledge shaped by beliefs and expectations (schemata) (St. George et al., 1994, p. 70).

Object recognition plays a major role in comprehension of the environment. In their study on scene perception, Biederman, Mezzanotte, and Rabinowitz identified that object recognition is significantly hindered when objects are placed in “violation” of their typical context. They noted that semantic relations are accessed simultaneously with physical properties and object identification, not deferred until after initial perceptual processing is complete (1982, p. 144). The relations of an object to its setting are accessed as part of the initial perception of the object. For the VR scenographer, this means a “semantic mismatch”—such as a modern fire extinguisher in a medieval dungeon—is not just an aesthetic error; it is a cognitive speed bump that forces the user to switch from a state of experiential processing to an evaluative processing, effectively ending the state of “flow” required for deep immersion.

This coherence is further reinforced through narrative priming. In VR, the environment must act as a cognitive map that primes the user for intended interactions. If a scene implies a specific era or function, every asset must reinforce that knowledge. Priming operates as a nonverbal communication. For instance, the scenographer should set the scene with high interaction zones, marking them by higher prop density and sharper textures, but also establishing boundaries delineated by natural collision blocks, debris, and furniture that might feel narratively appropriate rather than using invisible walls that don’t help in the construction of the spatial experience.

Drawing on Ryan’s analogy of literary immersion (2001, cited in Dalby, 2016, p. 61), VR immersion entails user surrender to the virtual world. However, the suspension of disbelief (Coleridge, 1817) in VR is fragile, complicated by the user’s awareness of mediation through head-mounted displays (HMDs). Therefore, immersive success depends on scenographic alignment with content design, enabling users to accept technological mediation and engage fully. Thus, unreal or magical elements can be trusted, but not the improbable. When a user enters a virtual medical lab, they expect a specific “density of clinical detail”. If the environment fails to meet this semantic density—or uses assets that feel generic and out of place, for instance, a modern office chair in a 1920s noir detective scene—it creates a temporal dissonance through which the user’s internal model of the world collapses. This mismatch also triggers BIPs. They argue that any sensory or semantic inconsistency serves as a reminder of the mediated nature of the experience, pulling the user out of the virtual world.

Consistency of Physics and Materiality

World coherence is maintained through the consistency of material properties. If the set dressing includes a heavy-looking iron anvil and a light plastic cup, the user expects them to behave differently if bumped or moved.

Jason Jerald argues that VR’s objective is not strict replication of reality. “Simple worlds consisting of basic structures that provide a sense of spatial stability can be extremely compelling, and making worlds more photorealistic does not necessarily increase presence” (Zimmons & Panter, 2003, cited in Jerald, 2015, pp. 50–51). High immersion relies on multiple VR continua shaped by the designer’s aims and vision.

Representation fidelity is the level of a VR experience to convey a place, but this might not relate to the real world; it might reference conveying emotions. Therefore, the most important aspect of representation instead of realism in VR is not the number of polygons but the consistency of the world’s rules. The second criterion is interaction fidelity, which refers to the degree to which interactions with objects correspond to the equivalent real-world task. The last one is experiential fidelity, which refers to the extent to which the user’s personal experience matches the virtual one.

Balancing the three fidelity continua necessitates clarity in interactive affordances. If some objects are “static” (baked into the environment) while others are “dynamic” (interactable), visual language must clearly distinguish them. Scenographers must use visual cues—such as subtle highlights or distinct shadows—to signal which parts of the dressed set are interactive, aligning the visual representation with the underlying physics.

LIGHTING FOR COMFORT, MOOD, AND WAYFINDING

The transition from conventional screen-based media to HMDs fundamentally shifts the role of lighting from a primarily aesthetic consideration to a critical factor influencing perceptual stability. Within VR environments, the scenographer must carefully balance the intensity and distribution of virtual illumination against the physiological constraints of the human visual system, particularly addressing the divergence between simulated lighting cues and the eye's physical responses.

Two key concepts remain essential to this balance: brightness and lightness. The first relates to “the apparent intensity of light that illuminates a region of the visual field” (Jerald, 2015, p. 90), and it's not only the quantity of light that reaches the user's eye, but also the different coloured lights that are going to offer different brightness properties. The second relates to “the apparent reflectance of a surface, with objects reflecting a small proportion of light appearing dark and objects reflecting a small proportion of light appearing light/white” (Jerald, 2015, p. 90). The objects should appear to maintain the same lightness and darkness even if the light reaching the user's glance is changing; this rule is known as lightness constancy.

Ocular Ergonomics and the Vergence-Accommodation Conflict

The primary challenge in VR lighting is the vergence-accommodation conflict (VAC). In the real world, the eyes converge and accommodate (focus) on the same point. In VR, the eyes accommodate a fixed focal plane (the screen) while converging on virtual objects at varying distances. When vergence and focal distance do not change together, more time is required to fuse stereoscopic stimuli (Hoffman et al., 2008, pp. 10–11). Lighting techniques, such as implementing depth-of-field effects or diffusing illumination on objects situated beyond a comfortable convergence range, serve to alleviate cognitive load and enhance sustained visual comfort.

A recurring issue within VR art direction is the occurrence of shimmering or inconsistent reflections. Disparities where specular highlights appear in one eye but not the other induce binocular rivalry, undermining visual coherence. Photometric asymmetries in luminance or contrast have been shown to diminish user comfort significantly (Kooi & Toet, 2004, pp. 105–106). To preserve the perceptual integrity of the VE, lighting calculations must ensure binocular consistency. Highly reflective surfaces should be treated with caution; if the “Fresnel effect” (Akenine-Möller et al., 2018, p. 319) or specular reflections do not align with the user's pupillary distance, the resulting rivalry causes the object to appear to vibrate, shattering PI.

Visual Hierarchy and Attentional Guidance

The intentional placement of visual components in VR to direct the user's gaze and aid in environment comprehension is known as visual hierarchy. Because viewers may look in any direction, VR necessitates dynamic composition techniques. According to Jenkins (2004), the spatial storytelling of digital environments relies on the strategic placement of elements that direct user focus without restricting agency.

Visual components must also be positioned in a way that supports the main idea of the story/experience. According to Jerald (2015), when story-critical parts are visually emphasized through hierarchical design, consumers are more likely to interact with them. As they explore the surroundings, users are certain to stay engaged with the story because of this alignment. Establishing a clear visual hierarchy relies on strategies involving motion, scale, and contrast. Directional lighting can accentuate salient elements, while contrast in colouration directs user focus towards key components. Additionally, motion—whether through dynamic objects or subtle environmental alterations—functions as an effective attentional cue.

In omnidirectional VR settings, users may experience choice paralysis due to the abundance of visual information. Here, lighting becomes the principal modality for wayfinding by structuring visual hierarchy according to luminance contrast. Objects or regions with elevated contrast naturally draw gaze more reliably than mere absolute brightness, emphasizing contrast as the dominant factor in directing visual

attention. Application of techniques such as rim lighting to delineate a focal hero object against a dark backdrop, or employing spotlights to mark thresholds like doorways, generates perceived affordances. These cues facilitate navigation within the virtual space without reliance on extradiegetic interface elements such as floating arrows or minimaps, which might compromise immersion.

The Physics of Presence: Global Illumination

Despite its substantial computational demands, lighting remains the foremost contributor to depth perception in VR. Real-time rendering systems endeavour to emulate light behaviour efficiently while preserving perceptual plausibility (Akenine-Möller et al., 2019, p. 386).

For an environment to be experienced as present, illumination must simulate a physically closed system, necessitating Global Illumination (GI)—the indirect lighting resulting from inter-surface light reflection. This indirect component is critical for anchoring objects within a scene; absent such cues, objects appear detached or levitating, undermining realism and scene coherence.

Given the computational cost of real-time GI at the high refresh rates required by HMDs (exceeding 90 Hz), practitioners frequently employ ambient occlusion (AO) to approximate shading effects in crevices and occluded regions. In VR contexts characterized by heightened depth sensitivity, precomputed (baked) lightmaps incorporating AO often outperform dynamic lighting approaches by providing stable, subtle shadows that assist users in spatial judgements—such as estimating distances between their virtual hands and objects, fundamental to reach envelope perception. Baking high-resolution AO into textures thus enables complex environments with a grounded appearance while mitigating runtime performance costs associated with dynamic light calculations.

DENSITY, OCCLUSION, AND LEVEL OF DETAIL STRATEGY

Unlike conventional gaming, where frame rate drops typically result in minor visual disruptions, similar performance lapses in VR can induce physical discomfort, including sickness. Consequently, scenographers must employ spatial optimization methodologies, balancing visual density—the perceived detail complexity—and computational efficiency—the processing load imposed on hardware—to sustain both performance and perceptual fidelity.

While high object density enhances the PI, it also increases the draw call count, which is the primary bottleneck for VR performance. To combat this, scenographers utilize static batching—combining multiple set-dressing props into a single mesh, which needs only one draw call to render the set (Akenine-Möller et al., 2019, p. 797). However, this creates a conflict with affordances: batched objects cannot be individually interactive. Therefore, the “density” of a scene must be strategically distributed, placing high-fidelity, interactive meshes within the reach envelope and using batched, lower-fidelity geometry for the scenic backdrop. At the same time, “using instancing minimizes the number of draw calls needed. Level of detail techniques are also used, such as rendering impostors for distant models” (Akenine-Möller et al., 2019, p. 800). The geometry stage’s performance is influenced by the quantity of light sources. More calculation is required when there are more light sources. Disabling or reducing local lights and using an environment map is a frequent approach to reduce computational work requirements.

Occlusion Culling: The Art of Hiding

One of the tools for maintaining performance in complex sets is occlusion culling: the process of not rendering objects that are currently blocked by other objects (occluders). “It is the most complex culling

technique, as it requires computing how objects affect each other” (Akenine-Möller et al., 2019, p. 830). Together with back face culling and view frustum culling, they are the specific techniques to keep a balance between what is on the scene and what needs to be rendered for the user’s eye.

In VR scenography, the environment should be designed with performance-aware layouts. Large architectural elements (walls, large cabinets, and hills) act as occluders. By intentionally designing a set with “Z-corridors” or broken lines of sight, the artist can dress a world with extreme detail, knowing that the engine only renders the specific “slice” the user is currently looking at. This allows for the illusion of a dense, expansive world without overwhelming the GPU.

Level of Detail and Geometric Fidelity

The Level of Detail (LOD) strategy involves creating multiple versions of a 3D model with decreasing complexity. As the user moves away from an object, the engine swaps the high-poly version for a low-poly version, and usually high-poly versions are generated with normal maps. However, simplified geometry and lighting produced using screen-space techniques, standard maps, textures, billboard sprites, some shaders and shadows (depending on implementation), and other 3D-like illusions in games are not truly 3D and frequently appear flat or odd in VR. Fortunately, many of these methods are effective across longer distances, but they should not be used up close. Close things need to be detailed to be as convincing as possible.

In the context of VR, LOD transitions must be significantly more subtle than in flat-screen media. Because the user has stereoscopic depth perception, “popping” (the visible jump between LOD levels) is highly disruptive to presence. The transition between different levels of detail should be imperceptible to the user to maintain the continuity of the virtual experience. To achieve this, scenographers use dithered transitions or crossfading, where the two LODs are blended over several frames.

Finally, the virtual world must be organized into spatial partitions. This allows the engine to quickly frustum cull anything that is not within the user’s HMD regular 110-degree field of view. For the VR designer, this means the world coherence is a series of loaded and unloaded “cells.” A well-dressed set ensures that these transitions are hidden behind narrative “choke points”—such as elevators, airlocks, or long hallways—allowing the hardware to swap high-density assets in and out of memory without the user perceiving the BIPs.

PERFORMANCE-AWARE ART DIRECTION

Performance-aware art direction in VR is the practice of designing a visual language that respects the “frame budget” (the strictly allotted time—typically 11.11 ms for 90 Hz—the GPU must render each frame). When an art director fails to account for these constraints, the resulting “judder” or dropped frames cause a catastrophic failure of both the PI and the user’s physical well-being.

Stylized vs. Photorealistic

The choice of art style represents a significant strategic decision in VR scenography. Although photorealism is often associated with higher immersion, it can prove counterproductive within VR environments due to issues such as the uncanny valley effect and increased computational demands. More crucial than striving for realism is maintaining a consistent visual style; a uniformly stylized environment facilitates the user’s cognitive acceptance of the virtual world’s underlying rules.

Adopting a stylized aesthetic—such as cel-shading, low-polygon models, or painterly techniques—allows art directors to minimize dependence on complex shaders and dense geometric models.

The computational resources conserved through this approach can be redirected towards elements with greater impact on presence and interactivity, including elevated frame rates or enhanced physics-based interactions, which arguably contribute more to user experience and PI than high-resolution texture details.

Material Efficiency and the “Hero” Object Strategy

In VR, every viewed surface is perceived through stereoscopic vision, which complicates the use of traditional two-dimensional approximations like parallax mapping. Consequently, shader efficiency becomes a paramount consideration in art direction. Performance optimization often involves employing master materials that utilize shared texture atlases, thereby reducing the number of draw calls. For instance, rather than applying 50 distinct textures within a cluttered scene, efficient art direction might implement a single, high-quality trim sheet covering multiple objects, preserving visual complexity without overburdening CPU-GPU communication.

Considering the spatial constraints inherent in VR interaction—specifically reach envelopes—art directors frequently implement a hero asset strategy. This strategy focuses polygonal detail and texture resolution on objects within the user’s immediate tactile range. This approach parallels the foveated nature of human vision, where only a small central region of the visual field is perceived in high detail, thus optimizing computational efforts towards the areas of greatest perceptual importance.

In practice, this means that a virtual object such as a door handle (a hero object) may possess a polygon count comparable to that of the entire surrounding environment. Such hierarchical allocation ensures that tactile interactions, which constitute a core component of VR experience, exhibit high fidelity and realism, while peripheral and occluding elements remain computationally lightweight.

The VR scenographer serves as a mediator between aesthetic considerations and technical constraints. By integrating theories of spatial affordances (Gibson, 1979), the physiological effects of luminance perception (Kooi & Toet, 2004), and the principles governing real-time rendering physics (Akenine-Möller, 2018), the designer crafts environments that are not merely visual but experientially immersive. A performance-conscious art direction maintains the coherence and stability of the virtual experience, balancing hardware capabilities with human perceptual and physiological limitations.

REFERENCES

- Akenine-Möller, T., Haines, E., Hoffman, N., Pesce, A., Iwanicki, M., & Hillaire, S. (2018). *Real-time rendering* (4th ed.). CRC Press. <https://doi.org/10.1201/b22086>
- Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110(31), 12846–12851. <https://doi.org/10.1073/pnas.1306779110>
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2), 143–177. [https://doi.org/10.1016/0010-0285\(82\)90007-X](https://doi.org/10.1016/0010-0285(82)90007-X)
- Carson, D. (2000). Environmental storytelling: Creating immersive 3D worlds using lessons learned from the theme park industry. *Game Developer*. <https://www.gamedeveloper.com/design/environmental-storytelling-creating-immersive-3d-worlds-using-lessons-learned-from-the-theme-park-industry>.
- Coleridge, S. T. (1817). Biographia literaria (chapter XIV). In V. B. Leitch (Ed.), *The Norton anthology of theory and criticism* (pp. 677–682). W. W. Norton & Company.
- Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2), 272–309. <https://doi.org/10.1080/15213269.2015.1015740>
- Dalby, J. R. (2016). Immersed in difficulty: The problem of suspension of disbelief in transmedia and VR experiences. *Online Journal of Communication and Media Technologies*, 7(SI), 67–85.

- Fountain, S. B., Sharp, J. L., Jackman, C. C., & Dyer, K. H. (2022). Chunking. In J. Vonk & T. K. Shackelford (Eds.), *Encyclopedia of animal cognition and behavior* (pp. 1402–1406). Springer.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
- Grau, O. (2004). *Virtual art: From illusion to immersion*. MIT Press.
- Hillier, B., Leaman, A., Stansall, P., & Bedford, M. (1976). Space syntax. *Environment and Planning B: Planning and Design*, 3(2), 147–185. <https://doi.org/10.1068/b030147>
- Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3), 33, 1–30. <https://doi.org/10.1167/8.3.33>.
- Howard, P. (2009). *What is scenography?* (2nd ed.). Routledge.
- Jenkins, H. (2004). Game Design as Narrative Architecture. In N. Wardrip-Fruin & P. Harrigan (Eds.), *First person: New media as story, performance, and game*, 118–130. MIT Press.
- Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.
- Kooi, F. L., & Toet, A. (2004). Visual comfort of binocular and 3D displays. *Displays*, 25(2–3), 99–108. <https://doi.org/10.1016/j.displa.2004.07.004>
- La Viola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: Theory and practice*. Addison-Wesley Professional.
- Mine, M. R., Brooks, F. P., Jr., & Sequin, C. H. (1997). Moving objects in space: Exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, Los Angeles, CA, USA (pp. 19–26). <https://doi.org/10.1145/258734.258747>
- Murray, J. H. (1997). *Hamlet on the holodeck: The future of narrative in cyberspace*. Free Press.
- Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, 6(3), 38–43. <https://doi.org/10.1145/301153.301168>
- Ruddle, R. A., Payne, S. J., & Jones, D. M. (1998). Navigating large-scale “desk-top” virtual buildings: Effects of orientation aids and familiarity. *Presence*, 7(2), 179–192. <https://doi.org/10.1162/105474698565668>.
- Ryan, M. L. (2001). *Narrative as virtual reality: Immersion and interactivity in literature and electronic media*. The Johns Hopkins University Press.
- Sánchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332–339. <https://doi.org/10.1038/nrn1651>.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549–3557. <https://doi.org/10.1098/rstb.2009.0138>
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, 109(3), 431–433. <https://doi.org/10.1111/bjop.12305>
- Slater, M., Brogni, A., & Steed, A. (2003, October). Physiological responses to breaks in presence: A pilot study. In *Presence 2003: The 6th Annual International Workshop on Presence*, Aalborg, Denmark (Vol. 157). Citeseer.
- St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, 6(1), 70–83. <https://doi.org/10.1162/jocn.1994.6.1.70>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Zimmons, P., & Panter, A. (2003). The influence of rendering quality on presence and task performance in a virtual environment. In *Proceedings of the IEEE Virtual Reality 2003 (VR'03)*, Los Angeles, CA, USA (pp. 293–294). IEEE.

Spatial Audio for VR

4

Rui Raposo, Mário Vairinhos, Pedro Reisinho,
and Silvino Martins

ABSTRACT

This chapter investigates sound's critical role in virtual reality (VR), positioning audio as a primary driver of spatial presence. Unlike traditional media, VR audio requires dynamic, 360° responsiveness to user agency to maintain the illusion of reality. We explore acoustic fundamentals alongside human spatial hearing mechanisms like Interaural Time and Level Differences. The text compares spatialization technologies, specifically binaural rendering (Head-Related Transfer Functions, HRTFs) for “ego-centric” simulation and Higher-Order Ambisonics (HOA) for “world-centric” perspectives. Technical implementations are examined, emphasizing sub-50 ms latency constraints to mitigate cybersickness. We should highlight the sound designer's evolution into a world-builder using multisensory integration, such as audio-haptic feedback. Finally, this chapter addresses emerging ethical challenges, including AI-generated deepfakes, and stresses the necessity of inclusive, accessible design for users with diverse sensory abilities.

INTRODUCTION TO SOUND

In the architectural construction of virtual reality (VR), sound acts as the primary driver of spatial presence, complementing visual stimuli to anchor and immerse users within a synthetic environment. While traditional media like cinema or television utilize soundscapes in symbiosis with fixed visual frames, VR sound design transitions from linear, non-reactive audio to complex, 360° soundscapes that seamlessly augment the user's experience through interactive modalities. This fundamental difference lies in user agency: as a participant navigates a virtual space, the auditory environment must respond dynamically to their orientation and location. This responsiveness maintains the psychological and physiological illusion of reality, thereby sustaining immersion and presence (Kern & Ellermeier, 2020; Mavridou et al., 2025). Fundamentally, VR sound is the digital reconstruction of mechanical vibrations designed to foster presence in a virtual experience. The sound designer's role has evolved from a passive storyteller to a world-builder. They are tasked with crafting environments where every rustle of leaves or distant echo reinforces the user's belief in the digital world. This chapter explores foundational concepts spanning basic acoustic principles to the integration of advanced spatial audio technologies, providing a comprehensive framework for understanding the sonic dimension of VR.

Fundamentals of Acoustics and Human Audio Perception

Understanding sound in VR begins with the physics of acoustics and the biology of hearing. The sonic experience is built upon the fundamental physics of vibration: sound begins as a mechanical disturbance propagating through a medium such as air (Figure 4.1).

In VR, these vibrations are digitally reconstructed to simulate reality, with their effectiveness governed by three primary physical properties: frequency, amplitude, and timbre.

Frequency measured in cycles per second (Hz) serves as the mathematical foundation for perceiving pitch, dictating whether a user hears the low rumble of an engine or the high chirp of a bird. For VR designers, manipulating frequency is essential for both realism and spatial awareness. High-frequency sounds are more easily localized by the human ear but are highly susceptible to being blocked by virtual obstacles, a phenomenon known as acoustic shadowing.

Amplitude translates the physical energy of a vibration into the psychological experience of volume or loudness. In VR, amplitude functions as a dynamic variable dictating the perceived “weight” and distance of an object. Controlling sound wave intensity allows developers to signal the proximity of a threat or the vastness of a warehouse. The interplay between frequency and amplitude enables users to feel the pressure of an explosion or the intimacy of a whisper, establishing a cornerstone of physiological immersion.

Finally, *Timbre* allows listeners to distinguish between sounds sharing the same pitch and loudness. Resulting from a complex hierarchy of overtones and harmonics, timbre gives a sound its unique identity and serves as the primary tool for material identification in VR (e.g., distinguishing metal from water or cobblestone from a hollow wood deck). While frequency and amplitude inform the “where” and “how much”, timbre informs the “what”. Together, these properties form the DNA of a soundscape.

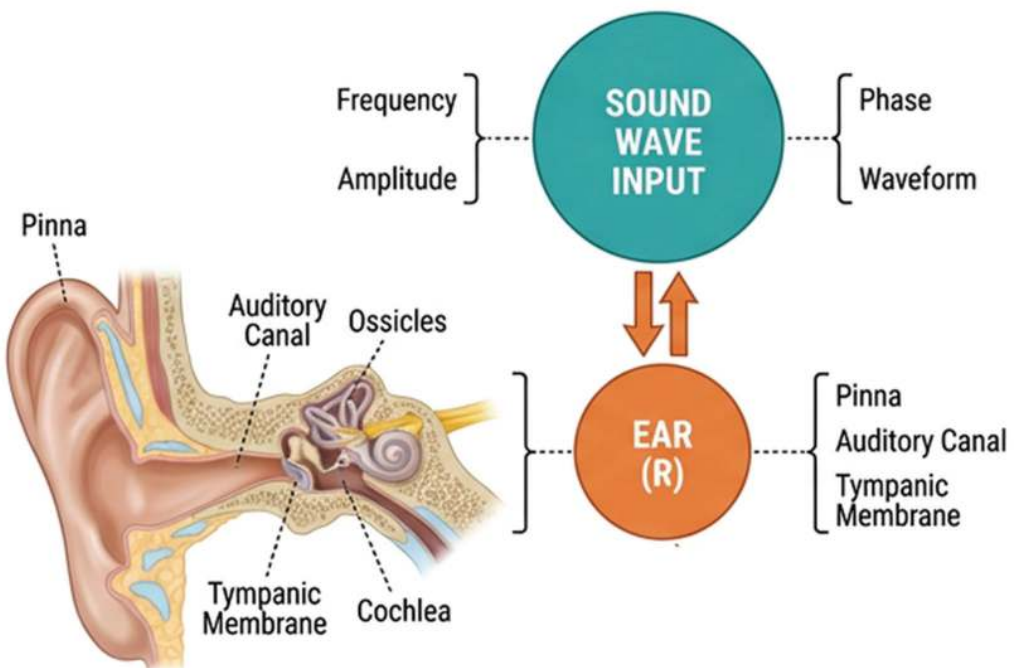


FIGURE 4.1 A visual overview of the auditory pathway, illustrating how sound wave properties are processed through the anatomical structures of the ear.

Mastering these properties ensures the user's brain accepts the synthetic environment as authentic. When frequency, amplitude, and timbre align coherently with spatial movement, sound design transcends simple playback to become a believable world the user inherently feels part of.

Emotional and Psychological Responses to Sound

Sound exerts profound psychological effects on human experience; musical elements, acoustic characteristics, and design choices can evoke specific emotions and influence mood. In VR, audio design is leveraged strategically to guide emotional responses, establish dramatic pacing, and reinforce narrative themes. Music and sound effects trigger emotional engagement through direct responses, cultural associations, and physiological changes like altered heart rate, breathing cadence, and skin conductance. Integrating interactive audio creates opportunities for dynamic emotional engagement where the soundscape responds to user actions and states. For example, adaptive music systems in VR games that modulate tempo, harmonic content, and orchestration in response to gameplay create powerful feedback loops, enhancing emotional immersion. Extensive research confirms the effectiveness of strategically designed audio in modulating user affect and creating engaging experiences.

Sound Design Strategies for Engagement

VR audio designers employ various strategies, ranging from naturalistic designs representing accurate environmental acoustics to stylized, abstract audio that establishes specific moods without prioritizing realism. Effective design often layers multiple audio elements, creating rich, complex soundscapes that reward contextual listening with multiple levels of perceptual detail. Using silence and dynamic contrast rather than filling all sonic space fosters dynamic listening experiences.

Interactive audio feedback is a particularly powerful engagement strategy; immediate sonic responses to user actions create cause-and-effect relationships that enhance agency and control. Well-designed feedback reinforces the success or failure of actions, communicates system states, and delivers satisfying payoffs. However, designers must ensure feedback is noticeable without becoming repetitive or annoying over time, and that distinct feedback types are easily differentiated by users.

FUNDAMENTALS OF SPATIAL AUDIO IN VR

The pursuit of true VR immersion extends beyond visual fidelity, relying heavily on replicating how humans naturally perceive sound. Understanding the intersection of psychoacoustics and spatial audio technology is critical to grasping the biological mechanisms behind 3D acoustic localization. By mastering concepts like binaural audio, ambisonics, Interaural Time and Level Differences, and Head-Related Transfer Functions (HRTFs), developers can translate human hearing principles into dynamic digital environments. Executing these spatial audio elements effectively prevents cognitive dissonance and solidifies the desired sense of presence and immersion.

Spatial Audio Technologies: Binaural vs. Ambisonics

To achieve true 3D immersion, developers must navigate two primary spatial audio technologies. While both aim to envelop the user, they differ significantly in their mathematical approaches and playback requirements.

- **Binaural Audio: The Ego-Centric Perspective:** Optimized for headphones, binaural audio provides highly accurate localization. Often recorded using physical “dummy head” models to capture the natural acoustic filtering of the human body, it relies on HRTFs to simulate how sound reaches a listener’s ears from specific XYZ coordinates. This intimacy is perfect for localized Sound Objects (e.g., a ticking clock), but the perspective is “baked-in”. To remain interactive in VR, the system must use real-time digital signal processing to re-spatialize the sound as the user’s head rotates.
- **Ambisonics (The World-Centric Perspective):** Ambisonics is a multichannel format capturing audio as a complete 360° sphere using a tetrahedral microphone. This “B-format” records overall pressure and directional vectors, making it inherently flexible. Because it is a directional field, it can be rotated in real time to match user orientation without complex re-rendering. This makes it the industry standard for Ambient Soundscapes.

Modern VR workflows often combine both, using ambisonics for the environmental layer and binaural point-sources for specific interactive elements.

PSYCHOACOUSTICS, SPATIAL HEARING, AND HRTFs

Psychoacoustics researches how humans perceive the physical properties of sound. Spatial hearing allows the brain to map an acoustic environment, determining a source’s origin, distance, and movement. Localization relies heavily on two binaural cues (Figure 4.2):

- **Interaural Time Difference:** This refers to the micro-delay between sound reaching the ear closer to the source vs. the ear further away.
- **Interaural Level Difference:** This is the variation in loudness caused by the “head shadow” effect, where the mass of the listener’s own head attenuates higher frequencies as they travel toward the further ear.

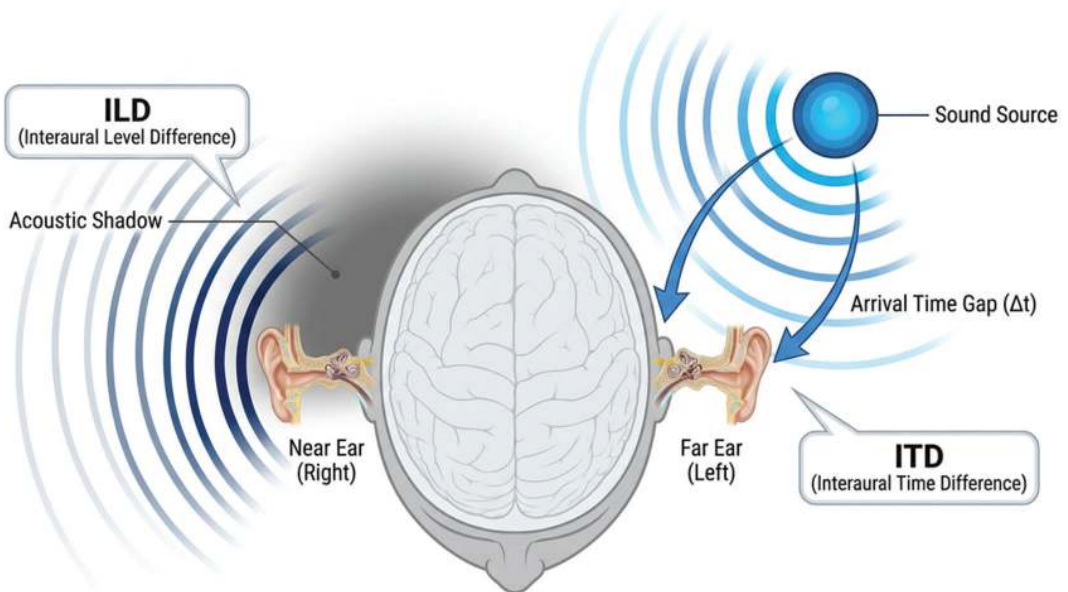


FIGURE 4.2 Illustration of ITD and ILD binaural cues for sound localization.

Binaural Audio and HRTFs

Beyond horizontal cues, humans use HRTFs to perceive elevation and front-back placement. The physical architecture of the outer ear (pinna), head, and torso filters sound waves depending on their angle of incidence. Binaural rendering utilizes algorithms to simulate these filters, shifting audio in real time to maintain consistent spatial coordinates as the user moves. Because anatomical features vary, generic HRTFs do not produce optimal localization for everyone, leading advanced VR systems to employ personalized HRTF measurements (Figure 4.3).

Spatial audio representation forms the foundation of immersive VR experiences, with binaural rendering utilizing mathematical representations of these HRTFs to simulate how sounds are filtered before reaching the inner ear (Richard et al., 2023). These functions enable the precise localization of sound sources in 3D space, creating the perception that audio originates from specific locations around the listener rather than from stereo speakers. In VR, designers use sophisticated algorithms to simulate these filters, ensuring that as a user tilts their head, the audio shifts in real time to maintain a consistent spatial coordinate. Contemporary research demonstrates that proper HRTF implementation significantly enhances the authenticity of virtual auditory environments. However, the effectiveness of binaural rendering depends heavily on the accuracy of HRTF measurements and the degree to which they can be personalized (Richard et al., 2023). Different individuals have varying anatomical features that affect how sound is filtered, meaning that generic HRTFs may not produce optimal localization for all users. Advanced VR audio systems increasingly employ personalized HRTF selection or measurement procedures, recognizing that individual hearing differences significantly impact the immersive experience.

Spatial Positioning and 3D Sound Field Design

Creating convincing three-dimensional sound fields requires careful consideration of how sounds are positioned, moved, and perceived within virtual spaces. The ability to place sound sources at arbitrary positions around the listener enables designers to create auditory environments that respond dynamically to user

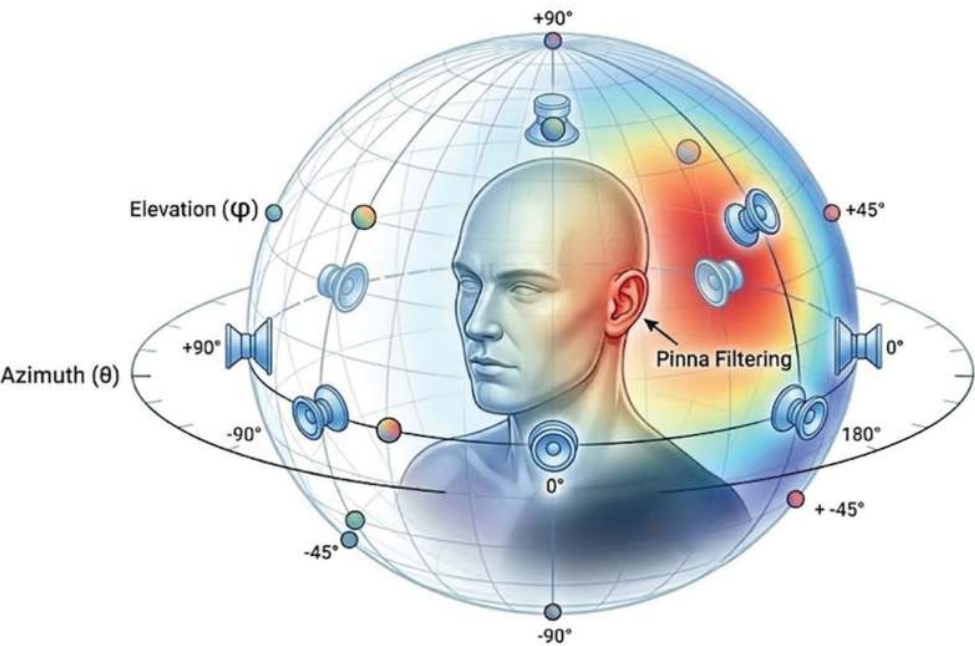


FIGURE 4.3 A 3D visualization of HRTF spatial mapping and pinna filtering.

movement and interaction. The psychological impact of accurate spatial positioning cannot be overstated for it provides a sort of “spatial glue” that complements and validates much of the visual stimuli presented to the user. When a user hears a sound object, such as a voice-over or a crackling fire, their brain instinctively expects the audio to move perfectly synchronized with their own physical movements. If these cues are even slightly misaligned, with a lag or inconsistency between what is seen and what is heard, the illusion of presence is broken and may even lead to cases of simulator sickness or cognitive dissonance. To prevent this, the technical implementation of spatial sound involves complex calculations to determine how audio should be filtered, delayed, and rendered. Real-time processing requirements mean that VR audio systems must efficiently compute these spatial transformations with incredibly low latency. By mastering the nuances of spatial hearing, from simple vibration to neural perception, VR developers can create environments that feel less like a digital projection and more like a tangible reality the user may feel immersed in.

AMBISONICS AND MULTICHANNEL AUDIO FORMATS

Ambisonics is a systematic approach to recording and reproducing spatial audio using spherical harmonic expansions. Unlike traditional surround sound relying on fixed speakers, ambisonics captures the 3D audio field as decodable components for any configuration. This format-independent approach is ideal for VR, where user movement is unpredictable. The difference between First-Order Ambisonics (FOA) and Higher-Order Ambisonics (HOA) is analogous to the audio resolution of a 3D sound sphere. FOA acts as a broad brush, using only four channels to capture general volume and basic directional differences, creating a spatially “blurry” soundscape akin to a low-resolution image. To achieve sharper localization, HOA incorporates additional harmonic components, increasing the channel count to 9, 16, or more. These overlapping mathematical shapes eliminate spatial blurriness, allowing users to pinpoint the exact location of a sound source (Figure 4.4). Selecting the ambisonics order involves a trade-off between spatial accuracy and computational load (Shivappa et al., 2016), with VR applications typically employing second- or third-order ambisonics as a practical balance. This high-resolution acoustic precision ensures users can distinctly separate overlapping sounds.

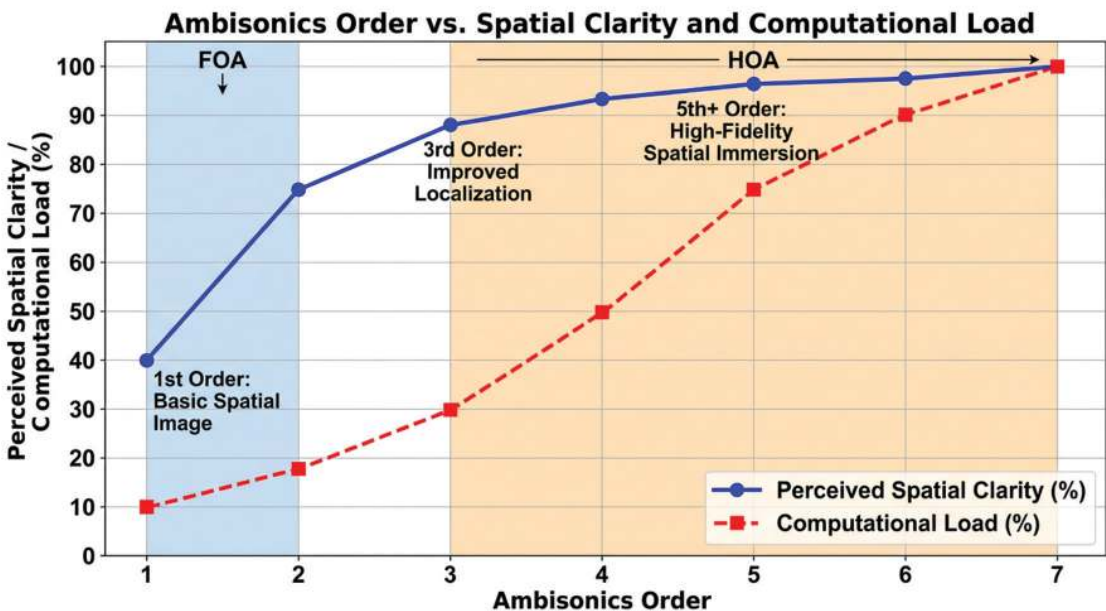


FIGURE 4.4 Relationship between ambisonics order, perceived spatial clarity, and computational load.

This high-resolution acoustic precision is essential for creating increasingly realistic VR environments, ensuring that users can distinctly separate overlapping sounds and are drawn into a truly immersive, pinpointed auditory reality.

Audio Signal Processing Techniques

Effective VR audio design requires sophisticated real-time signal processing to manipulate, filter, and render audio. Modern digital processing enables dynamic control over filtering, spatialization, and effects, enhancing auditory authenticity. As processing demands decrease with advanced consumer hardware, systems must operate under strict latency constraints, typically requiring processing and output delays below 50 ms to prevent noticeable audiovisual asynchronization (Cortés et al., 2024). Achieving this demands optimized signal processing pipelines, including efficient convolution algorithms, GPU-accelerated processing, and multicore parallelization.

AUDIOVISUAL INTEGRATION AND SYNCHRONIZATION

Integrating audio and visual modalities creates a substantially more engaging VR experience than either modality alone, as the brain automatically combines these inputs to construct unified perceptual experiences (Covaci et al., 2022). In VR, this multisensory integration is crucial; audiovisual coherence directly influences the user's sense of presence. The effectiveness of this integration depends on the perceptual compatibility and precise temporal alignment of the content (Zhang et al., 2023). When audio and visual events lag or convey conflicting environmental information, perceptual disruption occurs, and immersion diminishes. Therefore, audio design must support and complement visual content rather than compete with it.

Temporal Synchronization Effects

The temporal relationship between sensory events profoundly impacts how users integrate multisensory information. While users typically tolerate audiovisual delays up to 200 ms in standard media, small asynchronies can sabotage immersion in VR. Interactive applications requiring immediate feedback demand much stricter latency thresholds to preserve the perception of direct cause-and-effect. Implementing temporally synchronized media requires careful system design to minimize latencies and render both modalities to a common time reference (Zhang et al., 2023). Achieving the aforementioned sub-50-ms latency often requires specialized hardware, optimized architectures, and careful trade-offs between audiovisual fidelity and computational efficiency to prevent cybersickness symptoms (Figure 4.5).

MULTISENSORY INTERACTION DESIGN

Beyond audiovisual integration, VR experiences increasingly incorporate haptic, olfactory, and vestibular stimuli to engage users across multiple sensory channels. These mulsemmedia systems create synergistic effects where each modality reinforces the perceptual impact of the others. However, coordinating multiple modalities amplifies existing challenges, requiring meticulous attention to temporal synchronization, spatial coherence, and perceptual compatibility. For example, audio-haptic feedback systems combining sound effects with synchronized tactile sensations create interactive loops for virtual object manipulation or gameplay. Designing

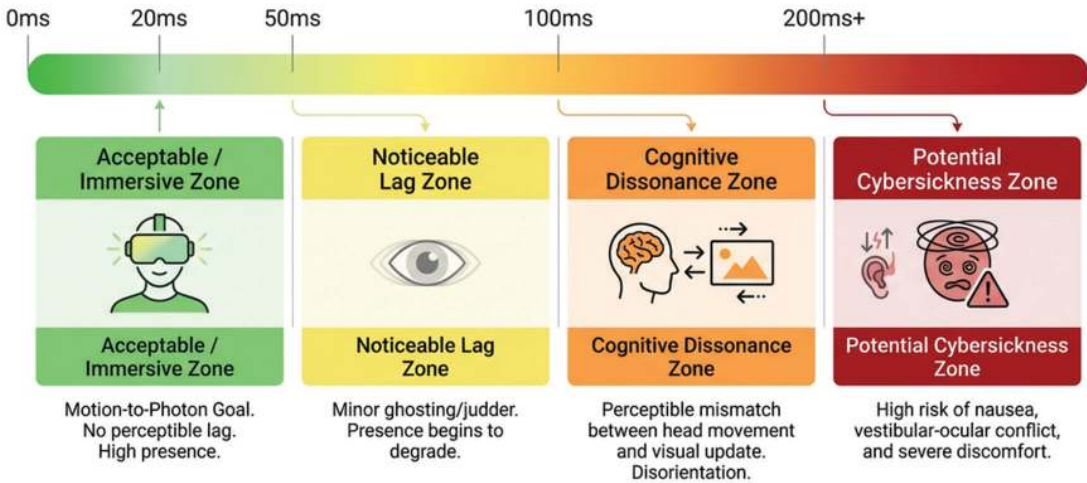


FIGURE 4.5 Critical latency thresholds in virtual reality.

these multisensory experiences requires a deep understanding of how different sensory systems interact to leverage their complementary characteristics without risking perceptual conflicts that shatter user immersion.

THE PERCEPTION OF PRESENCE

The sense of presence, the subjective feeling of “being there” within a virtual environment, is a primary benchmark for VR quality. Historically treated as secondary to visuals, audio is now recognized as a primary determinant of immersion. A convincing VR experience requires audiovisual coherence: sound and imagery conveying compatible, reinforcing information. Spatially coherent, temporally synchronized, and semantically compatible audio significantly increases reported presence. This synergy operates through three key mechanisms:

- **Spatial Situating:** 3D audio cues place the user contextually, providing distance and direction data.
- **Environmental Context:** Ambient soundscapes and reverberation signal a space’s size, acoustic properties, and material composition.
- **Interactive Feedback:** Sound confirms user actions, grounding digital agency.

Equally, minor audiovisual incoherence, whether spatial, temporal, or semantic, forces the brain to recognize the artificiality of the environment, shattering the illusion. Strategic choices in spatial positioning, spectral characteristics, and dynamic processing are as critical as visual fidelity in establishing a convincing perception of presence.

Multisensory Stimulation and Multimodal Design

Combining audio with visual, haptic, and vestibular stimulation yields synergistic effects that enhance presence far beyond single-modality experiences. Audio-tactile systems create convincing feedback loops where audio conveys broad event data, while haptics provide fine tactile details about physical interactions, forces, and material compliance. This multimodal integration is invaluable for virtual object manipulation, musical performance, and fine motor control, where precise force data improves task performance.

Advanced VR educational tools often utilize this synergy, combining audiovisual instruction with haptic consequences to foster embodied learning. However, to ensure acoustic and tactile cues reinforce each other rather than clash, their temporal characteristics must be rigorously synchronized.

Mulsemmedia Systems and Sensory Integration

While complete mulsemmedia systems remain relatively rare, ongoing research is rapidly expanding their feasibility. Coordinating beyond sight and sound demands meticulous synchronization to ensure users receive consistent environmental data. For instance, a virtual culinary experience might synchronize visual food presentation with auditory preparation cues, haptic textures, and olfactory aromas. Carefully coordinating these diverse inputs produces engaging experiences where multisensory integration creates a perceptual reality exceeding the sum of its individual parts.

ACCESSIBILITY IN AUDIO DESIGN

Accessible audio design guarantees VR usability for individuals with diverse sensory, cognitive, and motor abilities. For users with hearing loss, strategies include visual alternatives to auditory cues, haptic feedback translating sound into tactile sensations, or frequency adjustments tailored to residual hearing. Conversely, high-quality spatial audio is vital for visually impaired users to navigate and establish environmental context. Designing accessible VR requires flexible systems adaptable to individual preferences. Inclusive practices, such as providing audio descriptions for visual content, captions for dialogue, and robust control over audio levels, ensure sound enhances the experience without creating barriers. Crucially, accessibility must be integrated proactively during initial design phases rather than retrofitted as a reactive correction, ensuring these features organically enhance the overall user experience.

Inclusive Design Principles for Audio in VR

Inclusive design principles guide the creation of VR audio experiences that work effectively for diverse user populations with varying skills, preferences, and prior experiences. Rather than designing for a supposed “average” user and retrofitting accommodations later, inclusive approaches begin by considering diverse capabilities, leading to systems that inherently work well for a broader scope of people. In audio design, this means carefully considering how individuals with different hearing abilities, language backgrounds, musical training, and cultural contexts will perceive and engage with the content.

Adaptable systems that enable user customization such as adjusting audio levels, frequency emphasis, and spatial audio intensity strongly support inclusive experiences. Providing multiple modalities for crucial information, like combining audio feedback with visual and haptic alternatives, ensures that users with different sensory abilities can access and engage with VR environments effectively (Figure 4.6). Inclusive audio design is both a practical and ethical imperative, improving usability and enjoyment for everyone.

PRACTICAL IMPLEMENTATION AND DESIGN GUIDELINES

Sound design for VR applications follows established practices developed through decades of research in interactive audio. Effective VR audio design begins with careful planning during early project stages, establishing goals that align with the overall creative vision and user experience objectives. Sound

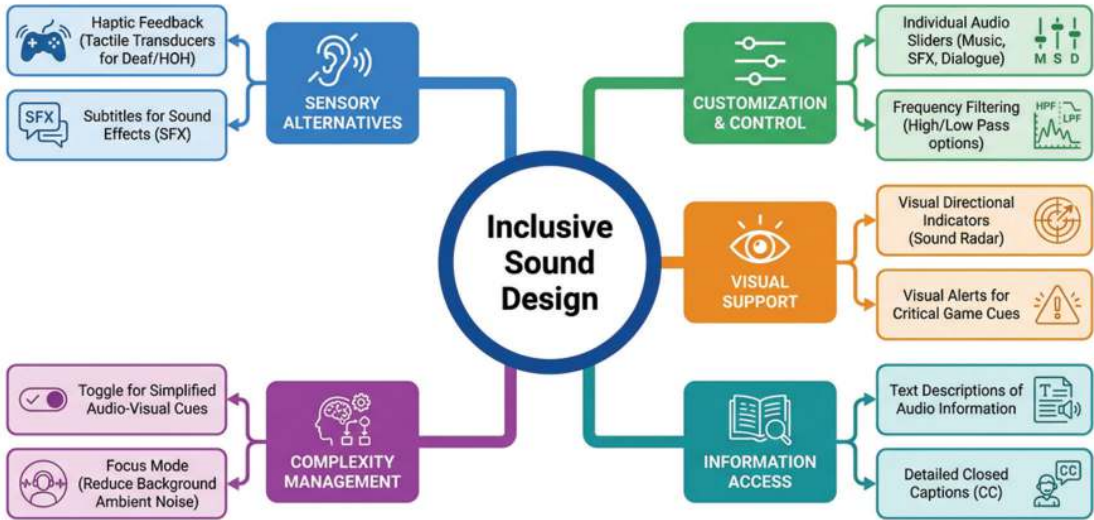


FIGURE 4.6 Inclusive sound design features and accessibility checklist.

designers must work closely with visual designers, interaction designers, and developers throughout the process to ensure audio choices support visual content rather than working at cross-purposes. The selection and design of sound effects require careful attention to both acoustic authenticity and perceptual distinctiveness, ensuring sounds accurately represent their sources while remaining easily identifiable within a complex soundscape. Spatial audio design must consider how sounds will be positioned and move within the virtual environment, reinforcing visual information about source locations. Additionally, music and ambient soundscapes should establish the emotional tone and environmental context, with careful attention to how different acoustic elements work together to create a cohesive auditory environment.

Quality of Experience Assessment

Assessing and maintaining a high-quality audio experience requires both technical metrics and subjective evaluation methods. Technical metrics including signal-to-noise ratio, frequency response accuracy, and latency measurement provide quantitative data on audio system performance. However, these metrics alone are insufficient for a comprehensive assessment. They must be complemented by subjective evaluation methods, such as user listening tests, questionnaires, and interviews, to capture user perceptions and satisfaction.

Quality of Experience (QoE) assessment models developed for interactive media provide frameworks for systematically evaluating this satisfaction in VR contexts. These models recognize that technical metrics do not fully dictate the user's perception of quality. User studies exploring how different audio quality levels, spatial accuracy, and interaction feedback characteristics influence perceived experience quality provide essential evidence to guide design decisions and establish quality standards.

User Testing and Physiological Measurement

Effective VR audio design demands rigorous user testing to validate that design choices achieve their intended effects and create engaging experiences. This testing should evaluate multiple aspects, including spatial audio accuracy, feedback clarity, the emotional impact of music and sound effects, and the overall contribution to presence and immersion.

Testing approaches include observational studies, quantitative questionnaires, and physiological measurements. Physiological metrics such as heart rate, skin conductance, respiration, and brain activity provide objective indicators of emotional arousal and engagement, complementing subjective reports. These measures can reveal emotional responses that users might not consciously recognize or be willing to articulate, offering hard evidence for the impact of audio choices. Combined with observational data and biometric sensors (Warp et al., 2022), physiological measurements provide a comprehensive assessment of how users respond to VR audio and which elements prove most effective.

Addressing Cybersickness and Side Effects

Cybersickness, characterized by nausea, dizziness, and disorientation in VR, represents a significant usability challenge that can limit application adoption. While visual factors like motion mismatch are extensively studied, audio factors including spatial audio cues, acoustic motion cues, and vestibular information conveyed through sound also influence susceptibility.

Careful audio design can reduce cybersickness by providing accurate spatial cues that perfectly match visual information about self-motion and environmental layout, thereby minimizing the perceptual conflicts that trigger symptoms. Strategies include using high-quality spatial audio, avoiding cues that create false perceptions of motion, and ensuring audio feedback supports visual motion data. Additionally, some research suggests strategically designed ambient audio and music might help users adapt to vestibular challenges, potentially reducing the incidence of cybersickness.

Audio Delivery and Technical Optimization

Delivering high-quality audio to VR applications requires the careful optimization of encoding, streaming, and processing to maintain fidelity while meeting strict latency and computational requirements. Audio streaming systems must deliver content with minimal buffering delay, often requiring specialized protocols that differ from traditional video streaming. The selection of audio codecs involves trade-offs between compression efficiency and quality; lossless codecs preserve full fidelity but require larger file sizes and bandwidth, while lossy codecs achieve substantial compression at the cost of some perceptible degradation.

Computational efficiency requires optimizing signal processing algorithms, utilizing hardware acceleration, and prioritizing essential functions within limited computational budgets. Real-time spatial audio processing is computationally heavy, particularly for higher-order spatial representations or multiple simultaneous sound sources. This demands optimization techniques like efficient convolution algorithms, GPU acceleration, and multicore parallelization. Understanding these technical trade-offs allows engineers to make informed decisions for specific VR applications.

FUTURE TRENDS

Emerging trends such as integrating AI-generated soundscapes and combining audio with haptic feedback are shaping the next generation of VR experiences. The horizon of VR sound design is shifting from static, pre-recorded assets toward dynamic, intelligent systems that evolve in real time alongside the user. As hardware expands, the objective is to move beyond mere imitation toward a fully responsive sonic ecosystem. A transformative trend is applying Artificial Intelligence (AI) and Machine Learning (ML) in sound design. Rather than relying on fixed audio libraries, future VR environments will utilize AI to generate procedural soundscapes adapting to a user's specific actions, physiological state, or emotional arc. For instance, a forest's ambient audio might shift in complexity and tone based on a player's heart rate or previous interactions, ensuring no two sonic experiences are identical.

The Emerging Risk of AI, Deepfakes, and Ethical Imperatives

A significant contemporary ethical challenge is the rise of Audio Deepfakes. AI solutions now generate highly realistic voice-overs using real people's voices without explicit permission. In VR, where social interaction and identity are central, the unauthorized use of vocal likeness poses severe risks to privacy and authenticity. Designers are increasingly tasked with establishing ethical boundaries for AI-generated content and ensuring synthetic voices are disclosed to users. As VR audio environments achieve unprecedented immersion, sound designers must navigate a complex intersection of legal mandates and ethical responsibilities. The transition to spatialized, interactive audio necessitates a rigorous approach to the Intellectual Property (IP) landscape: integrating musical assets and acoustic samples within a virtual space is governed by strict IP protocols. Practitioners must maintain a comprehensive understanding of licensing frameworks, from traditional Copyright to Creative Commons agreements. Failing to secure appropriate permissions can lead to significant legal liabilities and threaten project viability. To mitigate commercial deployment risks, designers frequently leverage curated sound banks and "royalty-free" repositories. Utilizing materials specifically licensed for commercial use ensures a legally robust workflow. This strategy maintains VR application scalability while respecting the economic rights of original creators.

Multimodal Integration and Innovations in Spatial Audio

The future of immersion lies in converging different sensory modalities. A key trend integrates haptic feedback with immersive sound technologies. By synchronizing spatial audio cues with tactile vibrations, developers simulate the physical "feel" of sound such as the low-frequency rumble of an explosion or the delicate tap of virtual rain. This synergy deepens physical presence by engaging the somatosensory system alongside the auditory pathway. Concurrently, spatial audio innovations focus on increasing the granularity of environmental acoustics. Future systems will likely feature sophisticated real-time ray-tracing for audio, allowing sound waves to bounce, diffract, and occlude with physical accuracy within complex virtual architectures. Sound is evolving through voice interaction and advanced speech recognition. Allowing users to utilize vocal inflections to control the environment transforms the voice into a primary interface tool rather than merely a communication medium. As these technologies mature, sound will transition from a visual "add-on" to the engine driving the user's narrative and emotional journey in next-generation VR (Potter et al., 2022).

Advanced Methodologies: Procedural and Adaptive Sonic Architectures

As VR matures, sound design shifts from static, sample-based triggers toward dynamic, systemic environments. This evolution favors real-time computational synthesis and reactive soundscapes mirroring the fluidity of the physical world. The frontier of immersive audio lies in procedural generation, where sounds are mathematically synthesized in real time rather than triggered from fixed assets. This allows for infinite variation, preventing "ear fatigue" from repetitive samples. Utilizing algorithms to generate audio based on physical parameters like a virtual object's velocity or a digital surface's tension creates an auditory experience uniquely responsive to every interaction. Distinct from procedural generation is adaptive audio, where a soundscape's complexity and tone dynamically shift based on the user's psychological state or location. Through contextual modulation, a musical score may increase in dissonance or tempo during high-stakes encounters (often managed through middleware like FMOD or Wwise). State-driven transitions ensure audio remains a continuous, living component of the environment, mirroring the "game state" to deepen emotional immersion. To achieve true perceptual realism, systems must utilize real-time

acoustic simulation, instantaneously calculating how sound waves interact with virtual geometry. This includes material-dependent reflection (sound behaving differently bouncing off virtual marble vs. velvet) and dynamic reverb (calculating changes in echoes and decay as a user moves from a cramped corridor to a vast cathedral) to match the visual scale and ground the user spatially. Finally, integrating voice interaction marks a pivotal shift toward naturalistic interfaces. Implementing advanced speech recognition and natural language processing (NLP) allows users to converse verbally with non-player characters (NPCs) or manipulate the environment through voice commands. This reduces traditional controller friction and fosters a more intuitive, presence-driven experience.

THE SONIC FUTURE OF VR

The evolution of sound in VR marks a paradigm shift from passive accompaniment to an active, architectural necessity. As explored, the efficacy of an immersive experience is fundamentally rooted in audiovisual coherence, the seamless synchronization of sight and sound that anchors a user's psychological and physiological sense of presence. Moving forward, the sound designer's role must continue transcending traditional boundaries, merging the skills of an acoustician with those of a world-builder. The transition toward procedural and adaptive systems signifies a move from static audio to "living" soundscapes that intelligently respond to user agency and physiological states. However, this sophistication brings significant technical and ethical responsibilities. Developers must prioritize minimizing latency to prevent the cognitive dissonance and cybersickness caused by misaligned audiovisual cues. Simultaneously, the rise of AI-generated audio and deepfakes demands a robust ethical framework to protect vocal likenesses and ensure authentic social interactions within virtual spaces. The pursuit of immersion must remain inclusive. By adopting universal design principles such as audio-haptic feedback loops and visual alternatives for auditory information, creators ensure VR remains accessible for individuals with diverse sensory abilities. When acoustic properties are mastered alongside cutting-edge spatialization technologies, the resulting environment transcends mere simulation. It becomes a tangible, multisensory reality and a portal into an infinite number of universes waiting to be explored.

REFERENCES

- Corrêa De Almeida, G., Costa de Souza, V., Da Silveira Júnior, L. G., & Veronez, M. R. (2023). Spatial audio in virtual reality: A systematic review. In *Proceedings of the 25th Symposium on Virtual and Augmented Reality*, Rio de Janeiro, Brazil (pp. 264–268).
- Cortés, C., Viola, I., Gutiérrez, J., Jansen, J., Subramanyam, S., Alexiou, E., Pérez, P., García, N., & César, P. (2024). Delay threshold for social interaction in volumetric extended reality communication. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(7), 1–22.
- Covaci, A., Saleme, E. B., Mesfin, G., Comsa, I. S., Trestian, R., Santos, C. A., & Ghinea, G. (2022). Multisensory 360 videos under varying resolution levels enhance presence. *IEEE Transactions on Visualization and Computer Graphics*, 29(4), 2093–2101.
- Kern, A. C., & Ellermeier, W. (2020). Audio in VR: Effects of a soundscape and movement-triggered step sounds on presence. *Frontiers in Robotics and AI*, 7, 20.
- Mavridou, I., Seiss, E., Ugazio, G., Harpster, M., Brown, P., Cox, S., Panchevski, F., Erie, C., Lopez, Jr. D., Copt, R., Nduka, C., Hughes, J., Butera III, J., & Weiss, D. N. (2025). Did you hear that?: Software-based spatial audio enhancements increase self-reported and physiological indices on auditory presence and affect in virtual reality. *Frontiers in Virtual Reality*, 6, 1629908.
- Potter, T., Cvetković, Z., & De Sena, E. (2022). On the relative importance of visual and spatial audio rendering on VR immersion. *Frontiers in Signal Processing*, 2, 904866.

- Richard, G., Smaragdis, P., Gannot, S., Naylor, P. A., Makino, S., Kellermann, W., & Sugiyama, A. (2023). Audio signal processing in the 21st century: The important outcomes of the past 25 years. *IEEE Signal Processing Magazine*, 40(5), 12–26.
- Shivappa, S., Morrell, M., Sen, D., Peters, N., & Salehin, S. M. (2016). Efficient, compelling, and immersive VR audio experience using scene based audio/higher order ambisonics. In *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*, Los Angeles, CA, USA. Audio Engineering Society.
- Warp, R., Zhu, M., Kiprijanovska, I., Wiesler, J., Stafford, S., & Mavridou, I. (2022). Validating the effects of immersion and spatial audio using novel continuous biometric sensor measures for virtual reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Singapore (pp. 262–265). IEEE.
- Zhang, H., Wang, J., Li, Z., & Li, J. (2023). Design and implementation of two immersive audio and video communication systems based on virtual reality. *Electronics*, 12(5), 1134.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART II

Interaction, Locomotion, and UI



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Interaction Modes

5

Iñigo Cuiñas

ABSTRACT

Virtual reality (VR) systems rely on diverse forms of user interaction that collectively determine the quality, realism, and effectiveness of immersive experiences. This chapter examines user interaction from three complementary perspectives: physical, visual, and psychological. It first reviews physical interaction modalities (including controllers, gesture recognition, and full-body motion tracking), highlighting their strengths, limitations, and influence on usability and immersion. Visual interaction is then explored through head-mounted and non-head-mounted display technologies, together with emerging methods such as eye tracking and multimodal interfaces. This chapter also analyzes key psychological mechanisms underpinning VR experiences, including presence, embodiment, emotional engagement, and the evolution of user behavior from first exposure to repeated interactions. By integrating these dimensions, this chapter outlines how human factors and technological design converge to shape VR experiences, while also addressing challenges such as accessibility, fatigue, motion sickness, and long-term psychological effects. The discussion concludes by identifying future opportunities for more natural, inclusive, and adaptive interaction paradigms that can support the continued growth of VR across education, healthcare, entertainment, and professional training.

THE CONCEPT OF USER INTERACTION

Virtual reality (VR) offers immersive experiences across various fields, including entertainment, education, healthcare, and training. As VR technology advances, understanding user interaction becomes essential for enhancing usability, improving experiences, and encouraging broader adoption. User interaction in VR refers to how people engage with VR hardware and software, the design principles shaping these experiences, and the challenges and opportunities involved in creating intuitive, effective environments.

Overview of VR Interaction

VR creates a simulated, three-dimensional environment that users can explore and interact with in real time. Unlike traditional interfaces, such as desktop computers or smartphones, VR aims to replicate a natural, immersive experience that blurs the line between the physical and digital worlds. Users typically engage with VR through specialized hardware like head-mounted displays (HMDs), motion-tracking controllers, and sometimes full-body motion sensors.

The goal of VR interaction is to enable users to manipulate virtual objects and environments in a manner that feels intuitive and responsive. Whether through gestural input, voice commands, or gaze tracking, the primary aim is to create a seamless experience that minimizes the cognitive load of interacting with the virtual world, allowing users to focus on their tasks or experiences.

Modes of Interaction in VR

Interaction modes are designed to reproduce natural actions as closely as possible. These modes determine how users select, manipulate, and navigate virtual elements, and they collectively shape the sense of immersion and usability of VR systems.

- **Controller-Based Interaction:** Handheld controllers remain the most common input method in VR. Equipped with buttons, triggers, and motion sensors, they allow users to point, grab, or manipulate virtual objects with a high degree of precision. Their familiarity, especially among gamers, and their integrated haptic feedback make them reliable and versatile. However, relying on physical devices leads to less natural forms of interaction and can increase the learning curve for inexperienced users.
- **Hand Tracking and Gesture Recognition:** Advances in computer vision and embedded sensors allow many modern VR systems to interpret hand and finger movements without requiring physical controllers. This modality enhances naturalism and reduces barriers between the user and the virtual environment, as actions such as pointing, swiping, or grasping are performed directly with the hands. While highly intuitive, hand tracking may suffer from reduced precision, sensitivity to lighting conditions, or fatigue during prolonged sessions.
- **Gaze-Based Interaction:** It uses the direction of the user's visual focus, either tracked through head orientation or eye-tracking, to select items or trigger actions. This method is efficient for simple tasks such as menu navigation or object highlighting, and it is particularly valuable for users with limited mobility or in contexts where hands must remain free. Its main limitation lies in the reduced expressiveness compared to manual input.
- **Voice Commands:** Voice recognition enables hands-free interaction based on natural language. This approach is useful for issuing commands, navigating menus, or controlling system functions. While convenient, its effectiveness depends on the accuracy of speech recognition, ambient noise levels, and the need to define clear command vocabularies.
- **Full-Body Tracking:** Some VR systems integrate sensors or camera-based tracking to reproduce the user's body movements in the virtual environment. This enhances presence and embodiment, enabling actions such as walking, crouching, or dancing. Although highly immersive, full-body tracking requires adequate physical space and can be limited by sensor occlusions or calibration complexity.

In practice, VR applications often combine these interaction modes to balance naturalism, precision, accessibility, and technological constraints. The choice of modality depends on the goals of the experience, the characteristics of the users, and the available hardware, making interaction design a central component of effective VR systems.

Design Principles for Effective VR Interaction

Creating effective and intuitive user interactions in VR requires careful attention to design principles that maximize immersion and minimize discomfort. These principles include those described in the following paragraphs.

- **Natural Interaction:** The more closely VR interactions mimic real-world actions, the more intuitive they become for users. Natural interaction reduces the learning curve, helping users feel more comfortable and confident when exploring virtual environments. It may involve actions such as grabbing, throwing, or pushing virtual objects through hand movements or body gestures.
- **Minimizing Latency:** VR interactions must feel responsive and immediate to avoid breaking immersion. Latency (i.e. delays between user actions and system responses) can disrupt the experience and even cause motion sickness. Designers must ensure that VR systems respond to user inputs in real time, with minimal lag between action and reaction.
- **Clear Feedback:** Users need clear feedback for their actions to feel confident in their interactions. This can be visual (objects highlighting when interacted with), auditory (sounds triggered by actions), or haptic (vibration feedback in controllers). Feedback confirms that the user's actions are recognized by the system and ensures that users remain engaged and immersed.
- **Ergonomics and Comfort:** Because VR experiences can be physically demanding, particularly those involving movement or prolonged controller use, ergonomics becomes essential. Hardware should be lightweight and comfortable for extended sessions, while user interfaces must minimize physical strain by accounting for factors such as range of motion and fatigue.
- **Immersive Environments:** The effectiveness of VR interactions also depends on how realistic and engaging the virtual environment is. High-quality graphics, detailed textures, and spatial audio strengthen the sense of presence, making users feel genuinely immersed. As immersion increases, interactions tend to feel more natural.

Challenges in VR User Interaction

Despite the advancements in VR technology, several challenges remain in creating fully intuitive and immersive experiences.

- **Motion Sickness:** One of the most common issues faced by users in VR is motion sickness, caused by a disconnect between the visual and vestibular (balance) systems. When users move in a virtual space while remaining stationary in the physical world, this sensory mismatch can lead to dizziness, nausea, and discomfort.
- **Learning Curve:** Although VR interactions aim to be intuitive, some users, especially those less familiar with technology, may struggle to adapt to the controls, particularly in complex environments. Keeping interactions simple, clear, and gradually introduced is essential for overcoming this challenge.
- **Limited Haptic Feedback:** While current VR systems provide basic haptic feedback through simple vibrations, truly immersive tactile sensations, such as perceiving the texture of virtual objects, remain in early development. This limited level of haptics reduces the realism of interactions, particularly when users manipulate objects in virtual environments.
- **Accessibility Issues:** VR is not always accessible to users with physical disabilities, vision impairments, or other limitations. While voice control and gaze tracking provide alternatives to traditional controllers, much remains to be done to achieve full inclusivity. Designing environments with accessibility in mind, such as offering adjustable text sizes, alternative input methods, and customizable control schemes, can help bridge this gap.

Opportunities for Future Development

As VR technology continues to evolve, there are numerous opportunities to enhance user interactions, as indicated (not exhaustively) in this section.

- **Improved Haptics:** Advances in haptic feedback technology could provide users with a more tactile experience, allowing them to feel the shape, weight, and texture of virtual objects. This would significantly increase the realism of VR interactions and expand the range of possible applications.
- **Artificial Intelligence-Driven Interactions:** Artificial intelligence can enhance VR interactions by enabling more responsive and adaptive environments. It supports the creation of smarter virtual characters, dynamic worlds that react to user behavior, and personalized interactions that adjust to individual preferences and actions.
- **Augmented Accessibility:** As VR becomes more mainstream, there is a growing focus on making it accessible to all users. This includes developing interfaces that cater to individuals with different abilities, incorporating features like eye tracking, and voice commands that provide a wide range of accessibility needs.

Evolution of User Interaction

User interaction within VR is a dynamic and evolving field that plays a central role in the success of VR systems. From traditional controller-based inputs to emerging techniques such as hand tracking and voice commands, the overarching aim is to deliver natural, immersive, and accessible experiences. Although challenges such as motion sickness and accessibility persist, ongoing advances in hardware, software, and interaction design continue to expand what VR can offer. As the technology progresses, so too will the ways in which users engage with virtual worlds, enabling more intuitive and immersive experiences across a wide range of applications.

HOW DOES A USER INTERACT WITH VIRTUAL OBJECTS AND ENVIRONMENTS?

VR has captured the imagination of users across diverse fields, from gaming to education and professional training. User interaction plays a key role in shaping these experiences. The ways users engage with VR evolve over time, influenced by their prior experience with technology, personal expectations, and how intuitive the tools feel. This section explores user interaction with VR, focusing on the first experience, the impact of previous exposure to virtual environments, and how repeated interactions shape the user's ability to engage effectively.

First Approach to VR

The first approach to VR is often filled with excitement, curiosity, and sometimes trepidation. Users experiencing VR for the first time may come from various backgrounds, but they generally share one characteristic: a lack of direct exposure to the full immersive potential of VR. For many, this first encounter can be disorienting or overwhelming, as they must adapt to navigating a three-dimensional, artificial world while still remaining physically grounded in the real world.

Key aspects of the initial VR experience involve familiarization with the hardware, such as wearing a HMD and learning how to use the input devices, whether they are motion controllers, hand tracking systems, or other methods of interaction. For first-timers, this stage can be challenging, especially if the interactions feel unintuitive or awkward. The process of learning to navigate and interact with the virtual environment (whether it involves walking, grabbing objects, or simply looking around) can be mentally and physically demanding, particularly if there is a steep learning curve.

Users unfamiliar with VR may initially feel discomfort, as their brains adjust to the experience. For instance, the sensation of movement in a virtual world while remaining stationary in the physical one can trigger motion sickness, a common issue for newcomers. Additionally, adjusting to the perceptual differences between VR and reality, such as the scale of objects or depth perception, can take time.

Nevertheless, the senses of presence and immersion often outweigh these initial challenges, leaving most users excited by the potential of the technology.

Interactions After Previous Experiences with Virtual Objects and Environments

User interaction in VR is strongly shaped by prior experience with virtual environments, particularly those gained through video games and other interactive media. Gamers, especially those familiar with first-person or simulation titles, tend to adapt more easily to VR because they already understand spatial relationships, controller use, and coordinated movement in virtual worlds. For them, VR often feels like an extension of skills they already possess: manipulating objects, navigating digital spaces, or controlling perspectives are familiar tasks, enabling them to explore environments with confidence and efficiency. They also tend to be more tolerant of the interface learning curve.

By contrast, non-gamers may face greater difficulties during early VR experiences. Lacking a foundation in virtual interaction, they often struggle with VR-specific controls and with orienting themselves in immersive environments. Even the sensation of immersion can feel new and disorienting. Without prior muscle memory for navigation and object manipulation, these users may experience frustration or confusion as they adapt to VR.

Second and Following Interactions with the Same Virtual Object and Environment

As users repeat interactions with the same virtual object or environment, their experience changes markedly. Growing familiarity with the interface and the scene reduces the cognitive effort required to navigate, enabling them to focus more on the task or experience itself. By the second or third exposure, users typically become more fluent in essential mechanics: movement, object manipulation, or task completion. Repetition fosters a form of virtual muscle memory, where many actions become automatic. This increased fluency allows attention to shift from basic interaction tasks to higher-level goals, such as problem-solving or creative engagement within the environment.

- **Improved Efficiency and Comfort:** Once users understand the layout of the environment and how to interact with objects, they are able to move more confidently and achieve their goals more quickly. This leads to a more seamless experience, enhancing the sense of presence and immersion, as users no longer need to consciously think about how to perform basic actions. The reduction in discomfort is also common as users' brains adjust to the unique demands of VR over time.
- **Deeper Exploration and Mastery:** As users grow familiar with a virtual environment, they gain the confidence to explore more deeply and experiment with new possibilities. In later interactions, they may try different strategies, discover hidden areas, or manipulate objects in ways they did not attempt initially. Once the novelty fades, users begin to test the environment's full potential, pushing the limits of what the system allows.

Evolution of User Interaction

User interaction with VR tools evolves through distinct phases, beginning with an initial period of adaptation shaped by prior experience with virtual environments. Gamers often adjust more quickly due to their familiarity with interactive media, whereas non-gamers may need additional time to become comfortable with VR's unique interface. Nevertheless, repeated exposure benefits all users, leading to greater efficiency, increased comfort, and deeper engagement with the virtual world. As VR technology continues to advance, understanding how users interact with these systems, and how those interactions evolve over time, will be essential for designing more intuitive, immersive, and inclusive virtual experiences.

VISUAL INTERACTION

Visual interaction is fundamental in extended reality (XR), as it determines how users perceive and interpret virtual content. The choice of display technology strongly influences immersion, realism, and situational awareness. This section summarizes the main modalities of visual interaction, distinguishing between head-mounted devices, non-wearable systems, and emerging techniques designed to support more natural engagement.

Head-Mounted Displays

HMDs present virtual imagery directly in the user's field of view, enabling high levels of immersion. Two main approaches exist:

- *See-through displays* overlay virtual elements onto the real environment using transparent optical components. Devices like the HoloLens achieve this through waveguides, maintaining visibility of the surrounding world while adding aligned digital content. This modality is suitable for applications requiring continuous awareness of the physical context.
- *Pass-through displays* use external cameras to capture the real environment and re-render it inside the headset. This allows accurate occlusion, environmental reconstruction, and mixed-reality blending (as in Figure 5.1), although it may introduce slight latency. Systems such as the Meta Quest combine VR-only and mixed-reality modes through this technique.

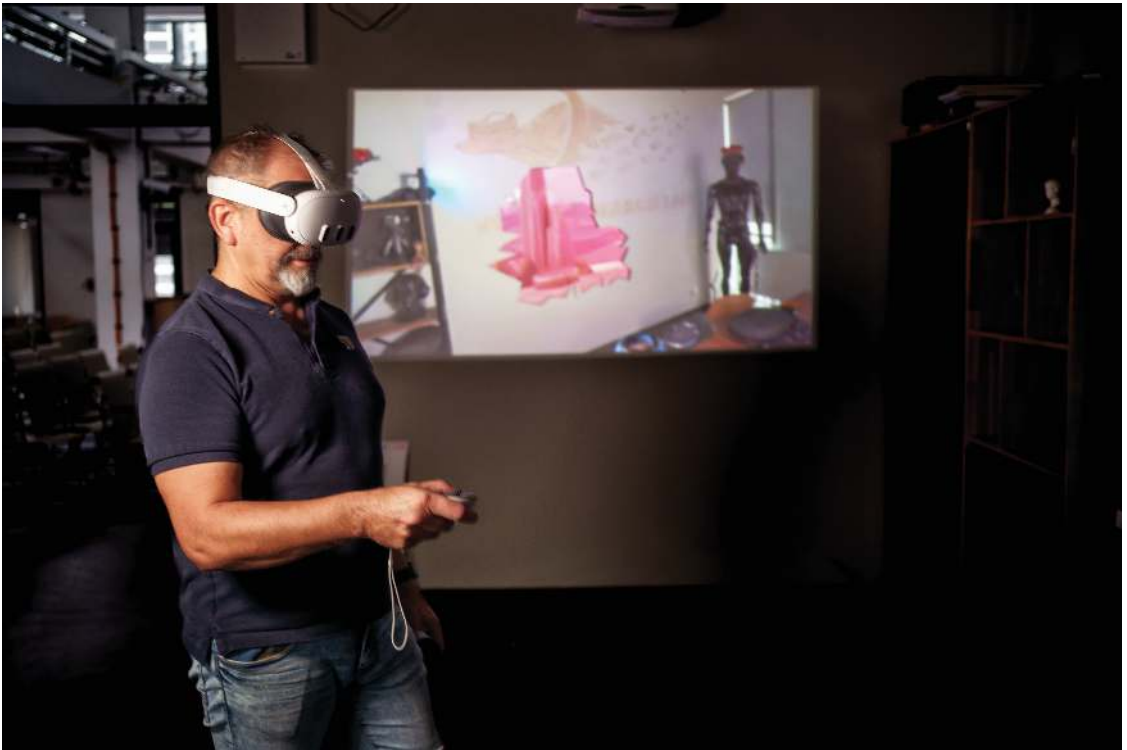


FIGURE 5.1 Pass-through displays in an XR environment (projection of user's view on the wall). (Credit: Krzysztof Pagacz.)



FIGURE 5.2 2021 Lights Fest in Lodz, Poland. (Credit: Iñigo Cuiñas.)

Non-Head-Mounted Displays

Non-wearable systems provide alternative XR experiences without isolating the user from the physical world.

Stereoscopic fixed screens simulate depth through 3D displays and glasses, enabling collaborative exploration of virtual content. While less immersive, these systems are effective in classrooms, labs, and shared visualization settings.

Projection-based augmented reality (AR) casts dynamic images onto real surfaces, transforming physical objects without the need for personal devices. This technique is commonly used in exhibitions and architectural demonstrations. An example is presented in Figure 5.2.

Handheld devices such as smartphones and tablets offer accessible AR experiences by overlaying digital content on camera images. Although less immersive, their portability supports widespread use. This approach is widely accessible and has been popularized by applications like Pokémon Go, which align digital content with the physical environment through device orientation and camera input.

Emerging Interaction Modalities

Eye tracking identifies the user's gaze point, enabling gaze-based selection and foveated rendering, which reduces processing load while enhancing responsiveness.

Hand gesture recognition uses camera-based tracking to identify hand movements, allowing object manipulation without controllers. This improves naturalism but may suffer from occlusion or lighting constraints.

Multimodal interaction combines gaze, gestures, voice commands, and controllers, providing flexible input options that adapt to user needs and enhance accessibility.

Summary of Visual Interaction

Visual interaction spans immersive HMDs, shared non-wearable systems, and emerging multimodal techniques. As XR technologies evolve, optimizing visual interaction remains essential for improving realism, accessibility, and user experience across diverse applications.

PHYSICAL INTERACTION METHODS

Physical interaction lies at the core of user engagement in virtual environments, shaping how individuals navigate, manipulate objects, and perform tasks in immersive systems. VR platforms rely on three main approaches: controllers, gesture-based interaction, and motion tracking. Each offers different levels of precision, naturalness, and usability.

Controllers: Direct and Precise Input

Controllers remain the most widely used physical interaction devices in VR. These handheld units incorporate buttons, joysticks, triggers, and inertial sensors to provide accurate and responsive input. Users can point, grab, or manipulate virtual objects through familiar physical actions, making controllers particularly suitable for applications requiring fine motor control, such as gaming, drawing, or operating virtual instruments, as it occurs in Figure 5.3.



FIGURE 5.3 Managing objects with controllers (projection in the wall shows what happens in the HMD). (Credit: Krzysztof Pagacz.)

Controllers offer high precision, providing well-defined inputs that support reliable object manipulation and navigation. Integrated haptic feedback enhances realism by simulating impacts, vibrations, and resistance, and its familiar design benefits experienced gamers by shortening the learning curve.

However, they also present limitations. Because they depend on physical buttons and do not always align with natural hand movements, controllers can reduce the sense of embodiment. Users unfamiliar with gaming interfaces may find them less intuitive, and prolonged use can lead to hand fatigue. They also rely on batteries or regular charging, which may interrupt training sessions or create logistical challenges in multi-user settings.

Gesture-Based Interaction: Natural and Intuitive Control

Gesture-based interaction aims to replicate real-world manual actions more closely by allowing users to control virtual elements through hand and finger movements. Modern VR systems employ optical tracking, depth sensors, or machine-learning models to detect hand poses and motions, enabling actions such as pointing, pinching, or swiping to interact with virtual environments. Moreover, gestures can be extended through the use of intermediary tools (such as mice, styluses, or data gloves), broadening the spectrum of interaction possibilities (Li et al., 2019).

Signal processing is used to convert the image into a collection of key points, and analyzing the coordinates of such reference points, it is possible to identify different hand gestures (see Figure 5.4). For example, Figure 5.5 shows how a user can grab, and then drop, virtual objects using his hands.

This modality enhances naturalism and reduces barriers between the user and the virtual world, eliminating the need for handheld devices. It is especially valuable in domains such as medical simulation, virtual classrooms, or design applications, where intuitive hand movement supports user engagement. Gesture-based systems can also interpret a wider range of expressive inputs, from simple commands to more complex manipulations.



FIGURE 5.4 Gesture recognition by HMD cameras. (Credit: Krzysztof Pagacz.)



FIGURE 5.5 Gesture recognition, using the hands to grab virtual objects. (Credit: Krzysztof Pagacz.)

Despite its advantages, gesture recognition faces challenges. Tracking accuracy depends on lighting conditions, camera quality, and the user's position within the sensing area. Fine-grained gestures may be difficult to detect reliably, and prolonged mid-air interaction can lead to physical fatigue (“gorilla arm”). Unintentional gestures can be misinterpreted, requiring robust filtering and well-designed interaction vocabularies.

Motion Tracking: Full-Body Embodiment

Motion tracking extends physical interaction by mapping users' full-body movements onto their virtual avatars. This produces a strong sense of presence and embodiment, allowing natural actions such as walking, crouching, or performing complex motor routines. Although this form of interaction is often perceived as the most intuitive and ecologically valid method of navigating virtual spaces, its effectiveness is highly dependent on the context and design goals of the application (Kamal et al., 2022). Two key tracking approaches are used: (i) outside-in tracking, which employs external cameras or sensors that detect markers on the headset or body; and (ii) inside-out tracking, which uses cameras embedded in the headset to infer position relative to the environment.

Inside-out systems dominate current consumer VR devices due to their portability and simplicity. They require minimal calibration and no external equipment, making them suitable for home and classroom settings. This is the case of the dancing activity in Figure 5.6, which shows the user and his views.

Full-body motion capture (through body-worn sensors; inertial movement units (IMUs), as accelerometers, gyroscopes, and also magnetometers; or optical suits) enables highly accurate reconstruction of limb movement. This capability benefits applications such as rehabilitation, fitness, sports analysis, or dance simulations, where whole-body movement is essential.

However, motion tracking introduces constraints. Accurate full-body capture typically requires a dedicated physical space, and tracking quality can degrade due to occlusion, limited camera coverage,



FIGURE 5.6 Motion tracking in a dancing VR application. (Credit: Krzysztof Pagacz.)

or rapid movements. Systems that rely on multiple sensors may involve complex setups or higher costs, reducing accessibility for general users.

Summary of Physical Interaction Modalities

Controllers provide precise, reliable input and are widely adopted across VR platforms, though they may reduce naturalism. Gesture-based interaction enhances intuitive control by using hands directly (Wang et al., 2020), but tracking limitations and physical fatigue remain challenges. Motion tracking offers the highest degree of embodiment and supports rich full-body interaction, yet requires adequate space and may increase hardware complexity.

In practice, VR applications often combine these methods to balance precision, immersion, and usability. Selecting the appropriate paradigm depends on the goals of the experience, the physical constraints of the environment, and the capabilities of the target users.

PSYCHOLOGICAL INTERACTION WITH VR

VR has reshaped how humans experience and interact with digital environments, enabling users to immerse themselves in simulated worlds that closely resemble real-life situations. Beyond its technical foundations, VR exerts a profound psychological influence by engaging the senses, shaping perception, and stimulating emotional responses. This psychological dimension is best understood through three key constructs: presence, embodiment, and emotional–cognitive engagement.

Presence: Perceptual and Emotional Immersion

Presence refers to the subjective feeling of “being inside” a virtual environment rather than merely observing it. It emerges when users perceive the virtual world as coherent, responsive, and credible, generating what Slater describes as place illusion and plausibility illusion, the combination of which produces realistic behavioral responses even when users remain aware of the artificial setting (Slater, 2009). Within this framework, some scholars have distinguished between different facets of presence (Barfield et al., 1993). Place illusion refers to the compelling sensation of being physically situated within the virtual environment. In parallel, plausibility illusion denotes the extent to which events and scenarios in VR are perceived as real or credible, even when users are fully aware of their artificial nature (Slater, 2009). When these two illusions co-occur (i.e. when the user feels present in the virtual space and simultaneously interprets the events as plausible) they can produce genuine, realistic reactions, like those that would occur in analogous real-world situations.

Several factors contribute to presence: sensory realism, spatial sound, system responsiveness, and the level of interactivity afforded by the VR application. High presence can produce strong emotional reactions, such as fear in height simulations or relaxation in calming environments, mirroring real-world affective responses. Attention is distributed between the physical and virtual worlds, with the balance depending on how effectively the virtual environment captures and sustains user engagement (Witmer et al., 1998). However, intense presence may also induce stress or discomfort if the content is emotionally charged or highly realistic. Designers must therefore balance immersion with psychological safety, especially in training or therapeutic contexts.

Embodiment: Experiencing a Virtual Body

Embodiment refers to the extent to which users experience a virtual body, avatar, or viewpoint as their own. This sensation emerges from the synchrony between visual feedback and physical movement: when a virtual limb responds in alignment with real-world actions, the brain incorporates the avatar into its self-representation.

Embodiment can significantly influence behavior and self-perception, as illustrated by the Proteus effect, in which users adapt their conduct to match the characteristics of their avatars. A taller avatar may provoke more confident behavior, whereas a smaller or weaker one may prompt more submissive actions. This phenomenon highlights VR’s potential to shape attitudes and behaviors both within virtual settings and beyond, with implications for training, therapy, perspective-taking, and motor-skill rehabilitation.

While embodiment enhances engagement, mismatches between the virtual and physical bodies may cause disorientation or emotional discomfort, particularly when users spend extended periods inhabiting avatars that differ markedly from their real-world identity.

Emotional and Cognitive Engagement

While VR offers significant potential for psychological engagement, it also raises concerns about the long-term effects of immersive technology on the mind. Prolonged exposure can generate both positive and negative outcomes, depending on the nature of the content and the amount of time spent in virtual environments.

VR has proven valuable in therapeutic and educational settings. Beyond its applications in rehabilitation and clinical treatment, it can foster positive psychological states such as increased creativity, empathy, and stress reduction. For example, VR-based meditation transports users to calming environments that promote relaxation and mindfulness.

However, extended use also presents risks. One major concern is cybersickness (a form of motion sickness arising from the mismatch between visual motion cues and the body's physical state) manifesting as dizziness, nausea, and disorientation. For some users, frequent exposure may even produce lingering sensations of imbalance or confusion.

There are also worries about the blurring of real and virtual experiences, particularly when users spend long periods in highly immersive simulations. Over-reliance on VR can make it harder to distinguish between virtual and physical events, potentially contributing to dissociation or a weakened sense of reality.

Furthermore, exposure to violent or disturbing virtual content may desensitize individuals to similar real-world situations. Prolonged engagement with violent simulations, for instance, may dampen emotional responses to real-world violence, reducing empathy and altering affective reactions.

Summary of Psychological Effects

Psychological interaction is fundamental to VR's transformative potential. Presence shapes how users accept virtual environments as real; embodiment influences behavior and self-perception; and emotional–cognitive engagement determines the depth of learning, empathy, and motivation within immersive experiences. While VR offers substantial psychological benefits, designers must also mitigate potential adverse effects through thoughtful interaction design, attention to user comfort, and careful content selection.

CONCLUDING REMARKS

User interaction in VR is a critical domain that determines the quality, accessibility, and overall impact of immersive experiences across fields ranging from education and healthcare to entertainment and professional training. This chapter has examined the multifaceted nature of VR interaction and the technological and human factors that underpin effective design.

From a physical perspective, interaction methods such as controllers, gestures, and full-body motion tracking provide distinct affordances. Controllers remain the most common due to their precision and tactile feedback, yet they also present usability and environmental challenges, including dependency on batteries. Gesture-based interaction and motion tracking support more naturalistic engagement and can strengthen embodiment and presence, though they bring difficulties related to accuracy, fatigue, and technical complexity.

Visual interaction plays an equally vital role by shaping how users perceive and interpret virtual elements. These modalities directly affect realism and immersion. Emerging approaches, including eye tracking and multimodal interfaces, promise more adaptive and intuitive interactions aligned with users' natural perceptual and behavioral patterns.

The psychological dimension is equally significant. Presence, embodiment, and emotional–cognitive engagement shape how users experience and respond to virtual environments. Repeated exposure increases familiarity, reduces cognitive load, and encourages deeper exploration. These mechanisms create opportunities for therapy, education, and empathy-building, but also raise ethical and mental health concerns, particularly when users are exposed to intense or prolonged experiences.

Looking forward, VR design must balance technical sophistication with strong user-centered principles. Accessibility, inclusivity, and comfort should remain central priorities, ensuring that VR systems are not only immersive and engaging but also safe, intuitive, and sustainable for long-term use. As VR continues to expand across industries, a nuanced understanding of user interaction will be essential to realize its full transformative potential. It should be grounded in both engineering practice and psychological insight.

REFERENCES

- Barfield, W., & Weghorst, S. (1993). The sense of presence within virtual environments: A conceptual framework. In *Proceedings of the International Conference on Human–Computer Interaction*, Orlando, Florida, USA (pp. 699–704).
- Kamal, A., & Andujar, C. (2022). Designing, testing and adapting navigation techniques for the immersive web. *Computers & Graphics*, 106, 66–76.
- Li, Y., Huang, J., Feng, T., Wang, H., & Dai, G. (2019). Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware*, 1(1), 9–21.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549–3557.
- Wang, D., Wang, X., Zheng, Q., Tao, B., & Zheng, G. (2020). How interaction paradigms affect user experience and perceived interactivity in virtual reality environments. In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, augmented and mixed reality: Design and interaction* (Lecture Notes in Computer Science, Vol. 12190, pp. 22–36). Springer.
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240.

Locomotion and Comfort

6

Grzegorz Zwoliński

ABSTRACT

This chapter discusses locomotion as one of the most critical design challenges in virtual reality, where movement directly affects comfort, spatial orientation, task efficiency, and presence. Because virtual movement often generates sensory conflict between visual and vestibular systems, locomotion design must be treated as a core user experience decision rather than a purely technical feature.

This chapter presents a taxonomy of locomotion methods, including natural locomotion, teleportation, smooth movement, indirect techniques, system-based solutions, and hardware-assisted approaches. It explains how locomotion should be selected according to task type, user characteristics, and usage context.

Special attention is given to comfort mitigation mechanisms such as vignetting, tunneling, snap turn, horizon stabilization, and adaptive speed control. This chapter also examines spatial reference frames, reorientation strategies, world scale, and different usage modes (seated, standing, and room-scale), showing how these factors influence movement perception.

The final section introduces comfort budget as a design framework in which comfort is treated as a limited resource consumed cumulatively by locomotion, animation, user interface motion, visual effects, audio events, and latency.

INTRODUCTION

Locomotion in virtual reality (VR) is a central design challenge because movement directly affects comfort, spatial orientation, task efficiency, and presence. Unlike conventional interfaces, movement in VR is experienced as part of embodied interaction and must therefore be treated as a core user experience (UX) decision.

In VR, users often perceive motion visually without corresponding physical acceleration, creating sensory conflict between visual and vestibular systems—one of the main causes of cybersickness (Reason & Brand, 1975; La Viola, 2000; Keshavarz et al., 2015).

Different locomotion methods create different trade-offs: teleportation minimizes discomfort but weakens spatial continuity, whereas smooth locomotion supports exploration at the cost of greater discomfort risk (Jerald, 2015).

For this reason, locomotion should be selected according to task type, user profile, session duration, and interaction logic.

LOCOMOTION TAXONOMY IN VR

One of the fundamental design challenges in immersive environments is selecting a locomotion model that simultaneously aligns with task logic, remains physiologically tolerable for the user, and can be mastered without excessive cognitive load. In practice, no single movement method is suitable for all VR experiences, because each produces a different balance between comfort, precision, immersion, and ease of learning.

The literature most commonly distinguishes locomotion into natural, teleportation-based, continuous, indirect, and system-based or special methods (La Viola et al., 2017; Jerald, 2015). This classification is not determined solely by movement mechanics, but primarily by the relationship between the user's physical motion and virtual translation.

Physical Locomotion (Natural Locomotion)

Physical locomotion is the most intuitive form of movement in VR because it relies on users' natural movement behaviors. In this model, actual body movement directly translates into positional change within virtual space.

The most basic implementation is room-scale walking, where users move within a physically available tracked area. This solution provides very high sensory consistency, because visual motion, proprioception, and vestibular signals remain aligned.

The primary advantage of this method is its exceptionally high level of immersion and natural exploration. Users do not need to learn abstract control schemes; instead, they rely on behaviors familiar from everyday experience. The main limitation, however, is the relatively small physical area available in most domestic or laboratory settings.

Natural locomotion can be extended through additional embodied movements such as leaning, where the user shifts the body to look behind an object or activate a spatial position; crouching, which supports vertical exploration by lowering the body; and step-based movement, in which locomotion is activated through stepping in place or performing half-steps.

These methods increase the sense of embodied presence but also require a well-designed spatial safety system.

A key component of this category is the use of physical boundaries, typically implemented through guardian or chaperone systems that warn users when approaching real-world obstacles. These boundaries are not only safety features but also influence world-scale perception and exploration strategies (Steinicke et al., 2009).

Teleportation-Based Locomotion

Teleportation remains one of the most widely used locomotion methods in VR applications because of its high comfort level and low degree of induced sensory conflict.

The most basic form is point-and-teleport, in which the user selects a destination with a controller and is instantly transferred to that location.

A more advanced version is arc teleport, where movement trajectory is displayed as a curved arc, making it easier to assess surface accessibility and spatial relationships across different elevations.

In contemporary teleportation systems, destination preview plays an important role by visualizing the target location before movement is executed, allowing users to evaluate distance, surface stability, and interaction accessibility after relocation.

A particularly important element is orientation preview, which allows users to define their post-teleport viewing direction before movement is completed. This reduces spatial disorientation and minimizes the need for corrective repositioning after arrival.

Teleportation significantly reduces cybersickness symptoms, but its design cost appears in reduced spatial continuity. Users do not experience the movement process itself, which may weaken mental map formation and reduce distance perception.

Continuous Locomotion (Smooth Locomotion)

Continuous locomotion involves smooth movement through virtual space using an input device, most commonly a joystick or controller thumbstick.

The most common implementation is joystick movement, where movement direction depends on either head orientation or controller orientation. This reference choice has major implications for comfort: head-relative control is usually learned more quickly by novice users.

Designing smooth locomotion requires careful attention to movement parameters, particularly speed curves, which determine how velocity increases and decreases, and acceleration profiles, which define how sudden movement transitions are softened.

Abrupt acceleration increases discomfort risk, which is why smooth ramping is often used.

Smooth locomotion offers high exploration continuity and strong trajectory control, but it remains one of the locomotion methods most strongly associated with cybersickness, especially among inexperienced users (Boletsis, 2017).

Indirect Locomotion

Indirect locomotion includes techniques in which movement in virtual space is generated through motor actions other than direct walking.

One of the most widespread examples is arm-swing locomotion, where rhythmic arm movement produces forward translation. This method increases kinesthetic congruence and often reduces discomfort compared with conventional smooth locomotion.

A second category is climbing-based locomotion, commonly used in exploration and training experiences. Here, users move by grasping objects and pulling their bodies through the environment.

A specific variation is pulling the world, also known as dragging the environment, where users move the environment relative to themselves rather than moving their own virtual position.

Indirect methods often enhance immersion because they engage the body more actively, but they may also increase physical fatigue during longer sessions.

System-Based and Special Locomotion

This category includes advanced locomotion solutions often requiring additional algorithms or specific environmental structures.

The most well-known example is redirected walking, in which the system subtly modifies virtual geometry so that users physically move within a limited space while perceiving exploration of a much larger environment (Steinicke et al., 2009).

Other solutions include moving platforms, where users are transported by moving objects without actively controlling locomotion.

Common implementations also include vehicle-based locomotion, in which movement occurs through vehicles, elevators, or capsules, and rail movement, where users travel along predefined trajectories.

Rail movement is particularly effective in narrative, museum, and educational experiences, where designers seek to control perceptual pacing and content exposure order.

Hardware-Assisted Locomotion

Hardware-assisted locomotion includes omnidirectional treadmills, active walking platforms, stabilization harnesses, and foot-based interfaces (Figures 6.1 and 6.2).

Omnidirectional treadmills allow multidirectional stepping while keeping users physically stationary (e.g., KAT Walk C2—see Figure 6.1, Virtuix Omni—see Figure 6.2), although gait remains less natural because movement often relies on sliding rather than heel-to-toe walking (Nilsson et al., 2018).

Active platforms provide higher biomechanical fidelity but require substantial space and cost (e.g., Infinadeck).

Additional solutions include stabilization systems supporting walking in place, running, jumping, and leaning, as well as foot-based interfaces such as pressure sensors or balance platforms.

Despite their immersive potential, such systems remain limited mainly to industrial training, rehabilitation, military simulation, and research because of cost, setup complexity, and limited universality.

Table 6.1 presents comparative characteristics of locomotion methods under typical VR conditions; actual performance depends on task type, user profile, and movement settings.

HOW TO MATCH LOCOMOTION TO TASK TYPE AND USER GROUP

In design practice, the choice of locomotion method should not be treated as a universal decision nor made solely on the basis of available technical solutions. The same movement technique may effectively support one type of task while significantly reducing the quality of experience in another usage scenario.



FIGURE 6.1 KAT Walk omnidirectional treadmill supporting hardware-assisted locomotion in virtual reality. (Photo: Tim Sheerman-Chase, CC BY 2.0, Wikimedia Commons [https://commons.wikimedia.org/wiki/File:Treadmill_Omni.jpg].)



FIGURE 6.2 Virtuix Omni omnidirectional treadmill in use during public demonstration. (Photo: tomemrich, CC BY 2.0, Wikimedia Commons [https://commons.wikimedia.org/wiki/File:Treadmill_Omni.jpg].)

TABLE 6.1 Comparison of locomotion approaches in VR design

LOCOMOTION TYPE	USER COMFORT	CONTROL PRECISION	LEVEL OF IMMERSION	EASE OF LEARNING	DESIGN CONSTRAINTS
Physical (natural locomotion)	Very high	High	Very high	Very easy	Limited physical space
Teleportation-based	Very high	Medium	Medium to high	Easy	Reduced spatial continuity
Continuous (smooth locomotion)	Medium to low	High	High	Medium	Cybersickness risk
Indirect locomotion	Medium	Medium	High	Medium	Increased physical fatigue
System-based/special locomotion	Implementation-dependent	High	Very high	Model-dependent	High design complexity
Hardware-assisted locomotion	High to very high	High	Very high	Medium	Cost, hardware availability

As emphasized in *The VR Book*, movement decisions in VR should be made at the level of experience architecture, because locomotion simultaneously affects comfort, spatial orientation, task efficiency, and sense of presence (Jerald, 2015).

Locomotion selection can be analyzed across three primary dimensions: task characteristics, user profile, and usage context. This approach is consistent with the principles of human-centered VR design,

in which locomotion serves as a mechanism for adapting the system to the user's perceptual capabilities and the intended purpose of the experience.

Task as the Primary Determinant

Task type remains the primary determinant of locomotion choice because different activities require different levels of movement continuity, positional precision, and perceptual stability. Spatial exploration benefits from smooth locomotion, which supports mental map formation, although discomfort risk increases (Boletsis, 2017). Procedural training usually favors teleportation or limited physical locomotion because positional predictability is more important than movement continuity (Jerald, 2015). Narrative applications often use rail movement or guided teleportation to control pacing, while educational environments benefit from simple locomotion that minimizes cognitive load. In rehabilitation, natural walking and hardware-assisted movement may directly support therapeutic goals (Nilsson et al., 2018).

User Profile as a Determinant

User profile strongly influences locomotion tolerance. Beginners usually prefer teleportation, snap turn, and low movement speed because early VR exposure is especially sensitive to sensory conflict (La Viola, 2000). Experienced users more often accept smooth locomotion because it offers greater trajectory control (Boletsis, 2017). Children require simple movement models, whereas older adults and users with mobility limitations benefit from seated mode, adjustable height, and remappable controls.

Usage Context

The same locomotion model may produce very different outcomes depending on session context. In short sessions, users usually tolerate more intensive movement because cumulative sensory effects remain limited. In longer sessions, however, perceptual fatigue management becomes crucial, and solutions that minimize sensory conflict are generally preferred, consistent with observations related to visually induced motion sickness (Keshavarz et al., 2015). In domestic environments, designers should assume limited physical space, diverse UX levels, and the absence of technical supervision; consequently, teleportation, seated mode, and limited smooth locomotion remain the dominant solutions. In laboratory settings, more advanced locomotion methods become possible, including room-scale movement, redirected walking, and hardware-assisted systems, with redirected walking being particularly effective because it requires precise management of physical space (Steinicke et al., 2009). In public installations, where rapid learnability and resistance to user error are essential, teleportation, rail movement, and guided movement are most commonly used.

COMFORT MITIGATION MECHANISMS

Designing locomotion in VR does not end with selecting a movement method. In practice, it is equally important to implement mechanisms that reduce the sensory conflict generated during movement in virtual environments. Even a well-chosen locomotion method may still produce discomfort if movement is not appropriately softened, predictably communicated, or perceptually stabilized.

Research on cybersickness consistently shows that the primary source of discomfort remains the mismatch between visual information and vestibular signals (La Viola, 2000; Keshavarz et al., 2015). Comfort mitigation mechanisms do not eliminate this conflict entirely, but reduce its intensity by limiting the perceptual stimuli that are most demanding for the user.

In practical terms, this means that the designer is not only designing movement itself but also designing the way movement is moderated by the system.

Vignetting

Vignetting dynamically narrows the effective field of view during movement by reducing peripheral visual input, thereby lowering vection intensity and discomfort risk (Keshavarz et al., 2015). It is most effective during continuous user-initiated motion when activated gradually, although excessive narrowing may reduce immersion (see Figure 6.3).

Tunneling

Tunneling creates a stable central viewing corridor while reducing peripheral motion. It is particularly effective during fast or curved movement and is often better tolerated during longer sessions.

Snap Turn

Snap turn replaces continuous rotation with discrete angular steps, most often 30°, reducing optical flow and improving comfort, especially for novice users (Jerald, 2015).

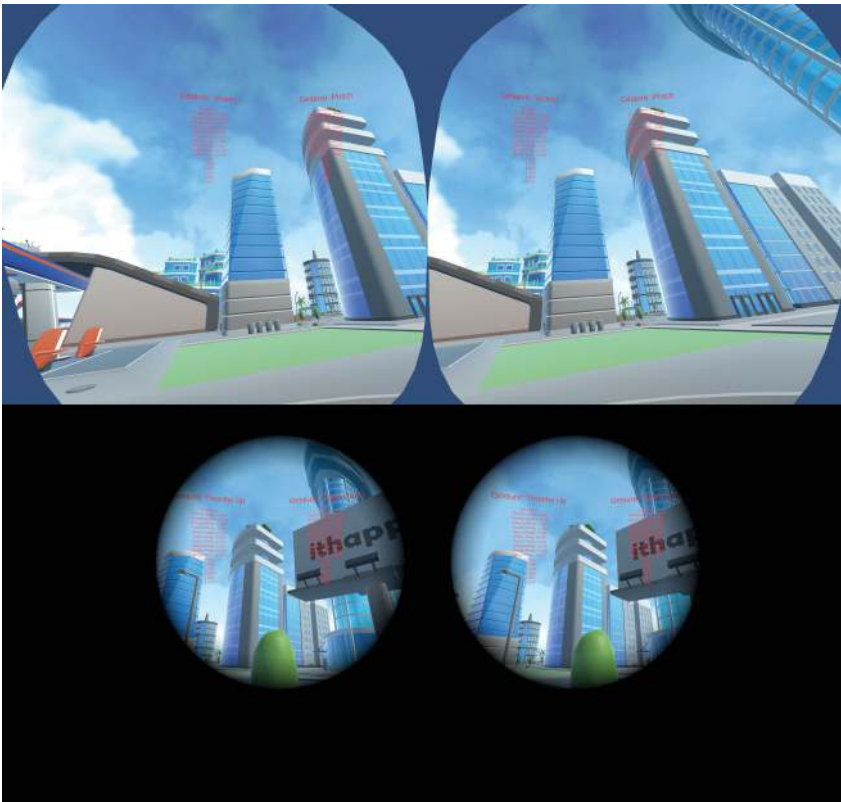


FIGURE 6.3 Illustration of the vignetting effect in immersive display rendering: the upper row presents images generated for the head-mounted display (HMD) without vignetting, whereas the lower row shows the same views with the vignetting effect applied.

Continuous Turn

Continuous turn offers greater continuity of spatial orientation but also generates a higher risk of discomfort. It may be used when the user has prior VR experience, rotation speed remains low, and additional comfort mitigation mechanisms are present. In this context, velocity limits are critical, because excessively rapid turning significantly increases cybersickness risk; therefore, limiting angular velocity and applying gradual acceleration profiles are strongly recommended. Continuous turn is best tolerated when users retain full control over movement initiation, which is consistent with research on user-initiated control (Stanney & Hash, 1998).

Horizon Stabilization

One of the most effective ways to reduce discomfort is to provide stable visual reference points during movement. Horizon stabilization consists of maintaining visually stable scene elements that support perceptual anchoring, most commonly through a clear horizon line, static architectural structures, or stable reference objects (see Figure 6.4). A special case is the cockpit effect, well known from simulation environments, where the presence of a cockpit frame, dashboard, or cabin significantly reduces discomfort by providing a constant point of reference even during intense movement (Reason & Brand, 1975).

Dynamic Speed Reduction

Dynamic speed reduction is one of the most subtle yet highly effective comfort mechanisms. It consists of adaptively slowing movement during perceptually demanding moments, such as approaching obstacles, turning, or changes in elevation. Comfort trigger systems may respond to sudden directional changes, high movement frequency, or prolonged continuous locomotion. This approach reduces the intensity of sensory conflict without introducing obvious perceptual intrusion.



FIGURE 6.4 Illustration of the cockpit effect in immersive environments: the presence of a stable cockpit frame provides a constant visual reference during motion, helping to reduce sensory conflict and improve user comfort.

Anti-Patterns: Common Design Mistakes

In practice, certain design decisions almost always increase discomfort risk. The most problematic cases include acceleration without warning, that is, sudden movement initiation without visual or temporal cueing; forced camera movement, where camera motion occurs independently of user intention; and diagonal drift, understood as simultaneous translation and rotation without clear control. Forced camera movement is a particularly strong trigger of cybersickness because users lose the sense of control over their own perspective (La Viola, 2000), while diagonal drift increases optical flow complexity, making perceptual adaptation more difficult.

REFERENCE FRAMES AND SPATIAL ORIENTATION

Three reference frames support orientation in VR: world-fixed (architecture, floor, walls), body-fixed (head and controller direction), and object-fixed (dominant task-related objects such as control panels). World-fixed references best support mental map formation, body-fixed references dominate local interaction, while object-fixed references are effective in procedural tasks but weaker during environmental transitions (Slater & Wilbur, 1997).

Reference Frames in VR

In VR environments, three main types of reference frames can be identified as supporting spatial orientation.

World-fixed reference frame

The most fundamental reference system is the world-fixed reference frame, based on stable environmental elements. Users orient themselves relative to architecture, walls, the floor, perspective lines, and fixed scene objects. In this model, spatial direction remains stable regardless of user movement, which best supports the construction of a mental map of the environment. In environments such as training warehouses or museums, a world-fixed reference frame becomes the primary basis for orientation because users can continuously rely on repeating structural landmarks.

Body-fixed reference frame

The second type is the body-fixed reference frame, in which orientation is anchored in the user's own body. Here, direction is determined by head orientation, controller position, hand movement, and current gaze direction. Body-fixed reference frames are particularly important during local interaction, but they may lead to loss of global orientation if the environment does not provide sufficient external reference points.

Object-fixed reference frame

The third type is the object-fixed reference frame, where orientation is organized around a specific object, such as a control panel, an operating table, or an industrial machine. In such cases, users structure their movement around a single dominant object. This model works well in procedural tasks but provides weaker support for orientation when users move between different environmental zones.

How Reference Frames Stabilize Comfort

Reference frames directly contribute to discomfort reduction because they stabilize movement interpretation. Among the most effective elements are horizon anchors, that is, persistent horizontal reference points such as

the horizon line, horizontal architectural elements, wall boundaries, and long floor surfaces. Their presence reduces perceptual uncertainty during movement and helps the brain interpret perspective changes more efficiently. A second important category consists of environmental cues, including color differentiation between zones, distinctive orientation objects, directional lighting, and the arrangement of furniture or equipment. In a well-designed environment, users do not need to consciously analyze space, because orientation emerges from repeated environmental relationships. Particularly important is floor visibility: a visible ground surface stabilizes spatial orientation by providing direct information about body position relative to the environment, whereas the absence of visible flooring increases the feeling of instability, especially during teleportation and smooth locomotion. Research on natural walking shows that users maintain orientation more effectively when they continuously perceive the relationship between their feet and the movement surface (Nilsson et al., 2018).

Reorientation Design

Every positional change requires a mechanism for restoring orientation.

After teleportation

Teleportation reduces sensory conflict but interrupts perceptual continuity of space. For this reason, immediately after teleportation the user should receive a clear viewing direction, a reference point, and a stable foreground element; without such cues, short-term disorientation often occurs.

After rotation

After snap turn or continuous turn, maintaining directional predictability becomes critical. If rotation occurs without a clear spatial context, users may lose orientation even after relatively small angular changes.

After scene transitions

Scene changes represent the most demanding moment of reorientation design. After a transition, users should immediately recognize where they are, which direction they are facing, and what the dominant spatial reference point is. For this reason, it is recommended to preserve at least one stable element between scenes, such as the entry axis, a dominant object, or a similar floor layout.

Direction Memory

Users in VR continuously construct directional memory. They do not memorize space as a purely geometric map, but rather as a sequence of relational cues concerning where they came from, where the entry point was, and where the target is located. If subsequent movements do not preserve logical directional relationships, spatial memory quickly deteriorates. For this reason, designers should avoid situations in which teleportation rotates the user without explicit intention, a scene resets viewing direction, or consecutive spaces lack a coherent spatial axis. Research on presence shows that loss of spatial predictability reduces the sense of presence almost as strongly as movement disturbances themselves (Slater & Wilbur, 1997).

WORLD SCALE AND THE PERCEPTION OF MOVEMENT

In VR environments, user movement is never interpreted solely through the displacement of a camera or avatar. Every movement is perceived in relation to the size of the surrounding environment, object

proportions, and distances between reference points. This means that even when locomotion parameters remain identical, users may experience very different levels of movement intensity depending on how world scale has been designed.

World scale influences the perception of presence, distance estimation, exploration strategies, and perceived movement speed. As emphasized in *The VR Book*, users in VR do not interpret space purely geometrically—they interpret it bodily, through the relationship between their own body and the surrounding environment (Jerald, 2015).

World Scale

In VR design, three main world-scaling strategies can be distinguished.

True scale

True scale refers to maintaining proportions consistent with physical reality. Objects appear at sizes users naturally expect: doors have dimensions corresponding to real-world doors, tables are positioned at natural heights, and distances between objects correspond to everyday experience. True scale most effectively supports natural spatial orientation because users automatically rely on familiar perceptual patterns. This approach is particularly important in procedural training, medical VR, and technical simulations.

Compressed scale

Compressed scale involves deliberately reducing the size of the world. Distances are shortened, space becomes denser, and objects are positioned closer to the user than they would be in physical reality. This solution is often used when designers aim to reduce the need for frequent locomotion, shorten exploration time, and increase interaction pace. Compressed scale works well in educational and narrative applications, but it requires careful control because excessive compression may distort distance estimation.

Exaggerated scale

Exaggerated scale consists of intentionally enlarging space or objects. This may involve very large interiors, enlarged educational objects, or monumental narrative environments (see Figure 6.5). Such solutions enhance dramatic effect and strengthen immersion, but they may also alter movement perception because users interpret the relationship between their own movement and world scale differently.

How Scale Influences Comfort

World scale affects comfort primarily by altering how movement is interpreted. In large environments, users are more likely to accept smooth locomotion because movement appears natural relative to environmental scale, although large spaces also increase the risk of orientation fatigue by requiring a greater number of reference points. In small environments, even minimal movement may feel intense because object relationships change rapidly, and excessive movement speed in a small room often produces greater perceptual overload than the same speed in a larger environment. This results from perceived velocity changes: users do not evaluate speed absolutely, but relative to moving environmental elements. In small spaces, more objects rapidly cross the visual field, optical flow becomes more intense, and vection increases more quickly, which directly raises the risk of visually induced motion sickness.



FIGURE 6.5 Illustration of exaggerated scale in immersive design, where architectural proportions are intentionally amplified to intensify spatial perception and emotional impact. (Background visualization: Łukasz Kłos, University of Ecology and Management in Warsaw, <https://nowawarszawa.pl/wiezowiec-z-kladka-nad-ulica/> with author's permission [original description in Polish: Wizualizacja Łukasz Kłos – Wyższa Szkoła Ekologii i Zarządzania w Warszawie . Udostępnianie wizualizacji i opisu projektu wyłącznie przy podaniu źródła – portalu NOWAWARSZAWA.pl – English translation: Visualization by Łukasz Kłos – University of Ecology and Management in Warsaw. The visualization and project description may only be shared with the source – NOWAWARSZAWA.pl].)

(Keshavarz et al., 2015). For this reason, small environments generally require lower velocity limits, stronger comfort mitigation mechanisms, and more frequent use of teleportation.

Reachability and Locomotion

One of the most practical design challenges concerns the relationship between arm reach and the need for locomotion. In VR, users very quickly interpret an object as either “reachable” or “unreachable” based primarily on visual relation rather than actual physical distance. If an object appears visually close but still requires additional movement, cognitive frustration emerges, especially when world scale deviates from true scale. At that point, the designer must decide whether to move the user or the object.

When an object is “too far away”

If an object lies beyond natural arm reach, three basic strategies are available: short user locomotion, pulling the object closer, or extending interaction range through ray interaction.

Should the user move or should the object move?

In procedural tasks, moving the user is usually preferable because it preserves spatial logic. In narrative or educational experiences, moving the object is often acceptable provided that it does not violate world credibility. From a design perspective, it is important to maintain consistency between perceived reach, expected reach, and actual interaction distance. Research in 3D interaction design shows that incorrect reachability assessment is one of the most frequent sources of micro-frustration in VR environments (La Viola et al., 2017).

USAGE MODES: SEATED, STANDING, AND ROOM-SCALE

One of the most frequently overlooked aspects of locomotion design in VR is that the same movement method may be perceived very differently depending on the user's body position. Movement performed while sitting, standing, or exploring in room-scale mode not only changes the control strategy, but also affects spatial perception, sensory tolerance, and interaction behavior. In practice, this means that selecting a usage mode is not a secondary decision. As emphasized in *The VR Book*, body posture determines how users interpret virtual environments because it influences the relationship between vision, balance, and expected movement range (Jerald, 2015). For this reason, locomotion systems should always be designed in relation to the intended mode of use.

Seated VR

Seated mode is one of the safest and most predictable forms of VR use.

Advantages

The primary advantage of seated VR is reduced physical load. In seated mode, body stability improves adaptation to virtual movement and reduces balance-related discomfort. A seated posture is particularly effective in simulators, procedural training, educational applications, and narrative experiences. An additional advantage is the natural comfort support provided by the cockpit effect, where the presence of a seat, control panel, or reference frame stabilizes movement perception (Reason & Brand, 1975).

Limitations

The main limitation of seated VR is reduced potential for natural spatial exploration. The user has a limited range of body rotation, leaning, and vertical posture change, which results in greater reliance on snap turn, joystick locomotion, and body rotation compensation.

User interface placement

In seated VR, the interface should remain within a narrow comfort zone. The most ergonomic placement is centrally in front of the user, slightly below eye level, and within limited interaction depth. Objects placed too low or too high force excessive neck movement.

Standing VR

Standing mode increases experiential naturalness but also introduces new ergonomic demands.

Fatigue

The most important limitation of standing VR is progressive fatigue. Even when movement remains minimal, longer sessions often lead to reduced concentration, shorter tolerance duration, and greater susceptibility to discomfort. Fatigue results not only from movement itself but also from maintaining postural readiness.

Comfort zone

In standing mode, the comfort zone becomes especially important, as it defines the area in which interaction is easiest and most sustainable. This zone should include the space directly in front of the user, the vertical range between chest level and head height, and limited lateral reach. Standing VR should also avoid frequent requirements for deep bending, prolonged arm extension, and unnecessary torso rotation. As shown in principles of 3D interaction design, overloading the comfort zone quickly reduces interaction quality (La Viola et al., 2017).

Room-Scale

Room-scale represents the most natural form of immersion because it relies on real physical movement.

Boundaries

The main limitation of room-scale remains the physical size of available space, which makes clear boundary definition essential.

Guardian systems

Contemporary VR systems use guardian or chaperone systems to warn users when they approach physical obstacles. A boundary system should appear early enough to prevent collision, remain visually clear, and disrupt immersion no more than necessary (see Figure 6.6).

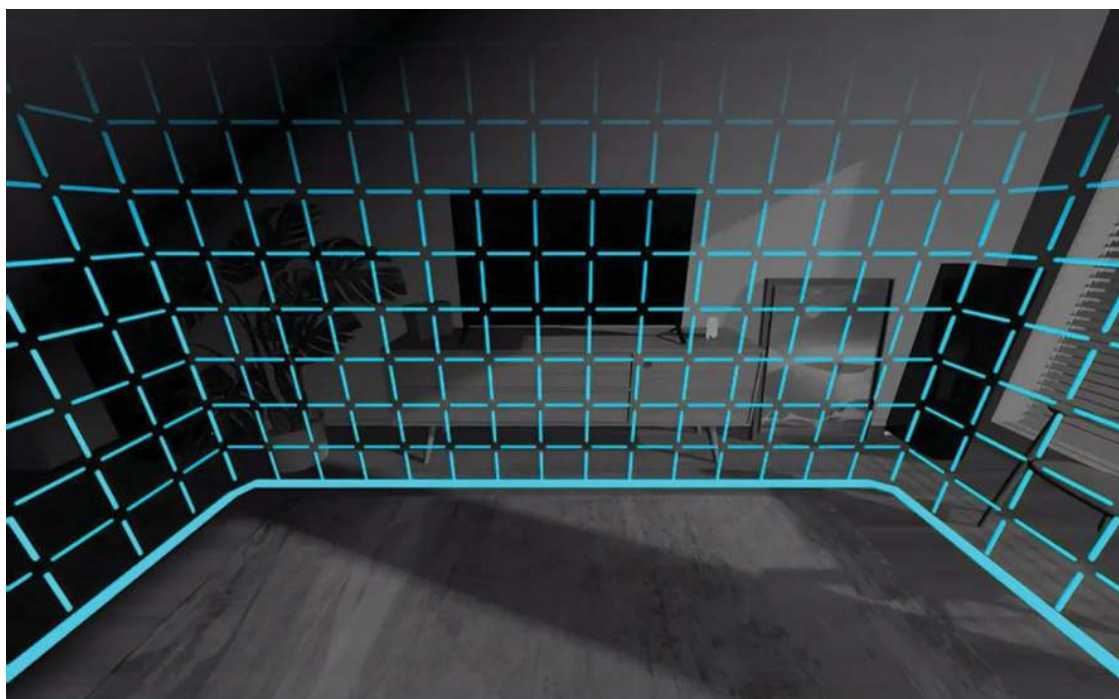


FIGURE 6.6 Example of the Meta Quest Guardian system, in which the real environment is revealed through passthrough visualization when the user approaches the defined safety boundary, helping to prevent collisions with physical surroundings.

Environmental planning

Room-scale also requires deliberate environmental planning. Interactive objects should be positioned so that users do not need to perform frequent 180° turns, backward movement is avoided when visibility is limited, and physical movement remains spatially logical. Research on natural walking shows that even limited physical exploration significantly strengthens the sense of presence (Nilsson et al., 2018; Slater et al., 1995).

Multi-Mode Design

Increasingly, VR systems are designed to operate across more than one usage mode, which means that a single locomotion system must support seated mode, standing mode, and room-scale use. This is not merely a matter of camera height.

One system, multiple UX logics

In practice, the same movement model often requires three different UX logics. For example, in teleportation, seated mode requires shorter targets and lower user interface (UI) placement, standing mode demands greater angular tolerance during selection, while room-scale must account for physical movement before teleportation. In smooth locomotion, seated mode typically requires lower speed, standing mode benefits from rotational stabilization, and room-scale demands integration with natural body movement. For this reason, multi-mode design should begin with a fundamental question: Does the system adapt to the user, or must the user adapt to the system? In well-designed VR, the answer should always favor the first option.

COMFORT BUDGET—DESIGNING COMFORT AS A PROJECT RESOURCE

In VR design practice, user comfort rarely depends on a single system component. Much more often, the final experience results from the accumulation of many small perceptual loads, each of which may seem acceptable in isolation, yet together exceed the user's tolerance threshold. For this reason, comfort should be treated not as a collection of isolated rules, but as the management of a limited design resource—a *comfort budget*.

By analogy with the concept of a performance budget in interactive design, *comfort budget* refers to the maximum level of sensory, cognitive, and motor load that a user can tolerate without a clear decline in experience quality.

Within this framework, every design element consumes part of the available budget.

There is therefore no neutral solution: every decision related to movement, animation, interface behavior, or world dynamics contributes to the overall comfort cost of the experience.

This perspective is consistent with findings from cybersickness research, which show that discomfort symptoms do not increase linearly, but rather accumulate over time (La Viola, 2000; Keshavarz et al., 2015).

Defining the Comfort Budget

Comfort budget may be defined as the sum of all loads generated by elements that influence movement perception and experiential stability. The main sources of comfort budget consumption include locomotion, animation, UI motion, visual effects (VFX), audio surprise, and interaction latency.

Locomotion

Locomotion is typically the largest single cost within the comfort budget. Smooth locomotion, continuous turning, and frequent elevation changes significantly increase sensory load.

Animation

Environmental animation also contributes to comfort cost, particularly when it involves movements of objects close to the user, dynamic scale changes, or background motion.

UI motion

Interface movement is often an underestimated source of perceptual load. If the interface follows the head with latency, shifts unpredictably, or appears abruptly, the user experiences additional visual conflict.

Visual effects

VFX, especially particle systems, lighting effects, or screen-based overlays, may increase perceptual load. The highest cost is typically generated by full-screen motion effects, blur, and dynamic flashes.

Audio surprise

Sudden audio stimuli do not directly generate motion, but they increase perceptual tension. When synchronized with movement, their comfort cost becomes higher.

Interaction latency

Latency between user movement and system response is one of the most critical comfort costs, because even minor instability in responsiveness reduces experiential predictability (Jerald, 2015).

Risk Budgeting

The comfort budget becomes a practical design tool when each project element can be evaluated in terms of both risk and protective mechanisms (Table 6.2).

Such a table allows designers to distribute risk consciously across the experience. If one element consumes a large portion of the comfort budget, other elements should be simplified accordingly. For example, when a project uses smooth locomotion in a large environment, it is advisable to limit abrupt animations, aggressive UI motion, and excessive VFX.

TABLE 6.2 Comfort budget risk distribution across selected VR design elements

<i>ELEMENT</i>	<i>RISK</i>	<i>MITIGATION</i>
Smooth locomotion	High sensory conflict	Vignette, velocity cap
Snap turn	Low	Angle adjustment
Dynamic UI	Medium	Motion damping
Intensive VFX	Medium/high	Frequency reduction
Audio surprise	Medium	Pre-cue
Interaction latency	Very high	Latency optimization

When a Design Exceeds the Comfort Budget

In most cases, the problem does not result from a single design error, but from the cumulative effect of multiple acceptable elements interacting over time.

Cumulative discomfort

Users may initially tolerate smooth locomotion, isolated animations, and occasional VFX. However, when these elements are combined over several minutes, discomfort gradually increases. This mechanism is particularly important because symptoms often appear with delay: users may function comfortably during the first minutes of exposure and then suddenly lose tolerance. For this reason, comfort testing should always include sessions longer than the first 2 or 3 minutes. Research on cybersickness shows that the first symptoms frequently emerge only after the accumulation of many small perceptual stressors (La Viola, 2000).

REFERENCES

- Boletsis, C. (2017). The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal Technologies and Interaction*, 1(4), 24.
- Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.
- Keshavarz, B., Riecke, B. E., Hettinger, L. J., & Campos, J. L. (2015). Vection and visually induced motion sickness: How are they related? *Frontiers in Psychology*, 6, 472.
- La Viola Jr, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, 32(1), 47–56.
- La Viola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D User interfaces: Theory and practice*. Addison-Wesley Professional.
- Nilsson, N. C., Serafin, S., Steinicke, F., & Nordahl, R. (2018). Natural walking in virtual reality: A review. *Computers in Entertainment (CIE)*, 16(2), 1–22.
- Reason, J. T., & Brand, J. J. (1975). *Motion sickness*. Academic Press.
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6), 603–616.
- Slater, M., Usoh, M., & Steed, A. (1995). Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3), 201–219.
- Stanney, K. M., & Hash, P. (1998). Locus of user-initiated control in virtual environments: Influences on cybersickness. *Presence*, 7(5), 447–459.
- Steinicke, F., Bruder, G., Jerald, J., Frenz, H., & Lappe, M. (2009). Estimation of detection thresholds for redirected walking techniques. *IEEE Transactions on Visualization and Computer Graphics*, 16(1), 17–27.

Animation and Motion in VR

7

Mário Vairinhos, Silvino Martins, Rui Raposo,
Pedro Reisinho, and Cátia Silva

ABSTRACT

This chapter explores the role of animation techniques and principles in the creation of immersive virtual reality (VR) content and environments. It provides an overview of issues such as motion perception, real-time animation methodologies, detailing the implementation of skeletal systems, inverse kinematics, and physics-based mechanics. This chapter also critically examines the adaptation of the classical expressive animation principles proposed by Johnston and Thomas to the spatial and interactive constraints of VR. It covers some of the technical challenges associated with the balancing of temporal demands of animation with issues like budgets and the low-latency responsivity required for optimal user experience. This chapter also states that merging kinetic realism, procedural integrity, and theatrical clarity is essential for establishing user presence, and that ensuring that VR environments function as responsive and cognitive sound spaces.

INTRODUCTION

Designing animation for virtual reality (VR) demands more than technical proficiency in rendering or artistic sensibility in motion design; it requires a grounded understanding of how the human perceptual system responds to artificially generated stimuli. This chapter addresses the interplay between perception and design through three complementary sections. The “Motion Perception in Immersive Environments” section establishes the perceptual foundations of motion perception in immersive environments. The “Animation Techniques for Real-Time Environments” section addresses the real-time animation techniques that operate within the strict performance budgets required by immersive displays. The “Expressive Principles of Animation” section examines how classical expressive animation principles are adapted to the specific context of VR.

MOTION PERCEPTION IN IMMERSIVE ENVIRONMENTS

Understanding motion perception in immersive environments implies establishing that interaction with a virtual environment demands from the human body a sensory adaptation for which it was not biologically prepared (La Valle, 2023). Under natural conditions, sensory organs receive stimuli from the physical environment and convert them into neural impulses that the brain integrates on the basis of a coherent and consistent structure, built from prior experience and from continuous interaction with the physical world (La Valle, 2023; Reason, 1978; Reason & Brand, 1975). Although the five senses (vision, audition, taste, smell, touch) are commonly described as the only sensory channels, motion perception unfolds into at least two additional systems: kinesthesia, which informs the brain about body movement and orientation, and the vestibular system, which provides information about body position relative to gravity and about acceleration through the detection of head movement and rotation (Gleitman et al., 2011). VR, aiming to be a sensory simulation that induces the perception of existing both physically and mentally within a virtual environment (Reisinho et al., 2024), disrupts the aforementioned biological mechanisms – which evolved over thousands of years – through a form of sensory hijacking: the sensory organs continue to function normally, but natural stimuli are intercepted and replaced by artificial signals that replicate, with varying fidelity, the energy patterns that would be produced had the stimulus originated in the real world (La Valle, 2023). When the virtual environment succeeds in simulating the senses, the brain interprets the artificial stimuli as plausible, leading to the sensation of presence – the psychological illusion resulting from the inability to discern that the experience is being mediated, giving the user the sensation of existing within the virtual environment (La Valle, 2023; Slater & Wilbur, 1997).

In order to grasp the fundamental biological reactions at the base of motion perception in immersive environments, it is important to begin by understanding the vestibular system. Located in the inner ear, the vestibular system comprises five organs: the utricle and saccule – the otolith organs, responding to linear acceleration, gravitational forces, and head tilt – and three semicircular ducts responsible for sensing angular acceleration, i.e., rotational head movements (Bear et al., 2021; Khan & Chang, 2013). In both structures, hair cells are deflected by inertia – through displacement of a crystal-laden gelatinous membrane in the otolith organs and through endolymph pressure on the cupula in the semicircular ducts – thereby converting mechanical displacement into neural signals (Bear et al., 2021; Khan & Chang, 2013; La Valle, 2023). Vestibular cues alone, however, possess limited perceptual value for representing the body in space and depend, to a large extent, on integration with vision and kinesthesia (La Valle, 2023; Reason & Brand, 1975). Correctly representing motion in virtual environments requires, therefore, that the design of the immersive experience address all three perceptual systems with a structure of neural stimuli that is coherent and consistent with real-world patterns (La Valle, 2023; Reason, 1978; Reason & Brand, 1975); in most head-mounted display configurations, however, the kinesthetic and vestibular channels are not artificially stimulated and continue to receive natural neural impulses originating from the physical environment (La Valle, 2023; Weech et al., 2019), rendering vection, i.e., the perception of self-motion induced by optical flow in the absence of physical displacement (Hettinger et al., 1990; La Valle, 2023), the dominant carrier of motion information.

Under natural conditions, optical flow designates the patterns of perceived motion across the visual field that arise whenever relative displacement occurs between an observer and the environment. In the real world, full sensory congruence prevails: the body moves forward, producing kinesthetic and vestibular confirmation, while objects captured by central vision approach and those in peripheral vision recede in the opposite direction, producing an optical flow field entirely consistent with the non-visual modalities (La Valle, 2023). Motion perception in the example is therefore not the product of vision alone, but the result of a convergence in which optical flow, vestibular input, and kinesthetic feedback all confirm the same event: the body is moving through space. In interactive media, the system exploits the optical flow patterns to induce vection by presenting the same approaching and receding pattern in the human visual field, resulting in the perception of self-motion (Hettinger et al., 1990; La Valle, 2023). In VR, if a simulation were designed with the same locomotion example of walking along a street, a sensory fracture would be felt: the brain would receive optical flow patterns signaling self-motion through vection, while the kinesthetic and vestibular information would remain invariable and continue to emit neural impulses signaling that the body is at rest (La Valle, 2023; Reason & Brand, 1975). Hence, in common head-mounted

display configurations, the strength and pervasiveness of thevection illusion are what ultimately determine the magnitude of the perceptual mismatch between visual and non-visual modalities; consequently, when the mismatch exceeds the tolerance of the sensory system, it may give rise to cybersickness – a range of symptoms akin to motion sickness, including nausea, fatigue, headache, and ocular strain (Kennedy et al., 1993; La Viola, 2000). The design of an immersive environment should therefore incorporate motion strategies and animation techniques to reduce visual–vestibular discrepancy, e.g., postural alignment between the physical and virtual environments, avoidance of active locomotion, or, if locomotion is crucial for interaction, the application of a vignetting strategy (Jerald, 2015; La Valle, 2023).

ANIMATION TECHNIQUES FOR REAL-TIME ENVIRONMENTS

In the context of digital environments, animation is defined by the change of an object’s properties over time (Jung & Vitzthum, 2022). While these changes can involve a diverse range of attributes, including position, scale, color, or geometry, the primary challenge in interactive media lies in how these changes are managed. Unlike the pre-rendered sequences found in noninteractive media such as cinema, animation in real-time environments must be computed dynamically. This allows the system to adapt to user inputs and system events, ensuring that the virtual world remains responsive.

Although this interactive paradigm is not exclusive to immersive media, the design of animation for VR introduces specific constraints. A paradigmatic example is the sense of embodiment, often supported by tracked controllers or full-body avatars (Spanlang et al., 2014), which raises user expectations for motion that is not only visually accurate but also consistent with their own physical actions (Debarba et al., 2022). Furthermore, as with many other aspects of designing for VR, animation requires a careful balance between the pursuit of realism and plausibility (Slater et al., 2022) and the strict performance budgets required to maintain high, stable frame rates (typically 90Hz or higher) to ensure user comfort and prevent simulator sickness (Jerald, 2015).

Articulated Motion and Skeletal Systems

The technique for representing movement in real-time environments depends on the complexity of the object being animated. Animating complex articulated entities, such as virtual humans, is usually achieved through *skeletal animation*. This approach relies on an internal hierarchical structure of interconnected “bones” or “joints”, known as a rig (Figure 7.1a), which serves as an abstract control system for the visible 3D mesh (Parent, 2012).

Rather than manipulating thousands of individual vertices, the system applies transformations to this underlying skeleton. The visible surface geometry then deforms accordingly through *skinning algorithms* (Figure 7.1b), such as Linear Blend Skinning or Dual Quaternion Skinning, that determine the “weight” or influence each joint has over specific areas of the model (Akenine-Möller et al., 2018; Le & Hodgins, 2016). Nevertheless, these mathematical approximations often introduce characteristic artifacts, such as collapsing elbows, “candy-wrapper” twists, and unnatural bulging around the joints (Le & Hodgins, 2016). In VR, the impact of such deformities can be more noticeable due to the user’s ability to inspect virtual entities from any angle and at close range.

Additionally, the manipulation of skeletal structures is typically managed through two kinematic approaches: *Forward Kinematics (FK)* and *Inverse Kinematics (IK)*. In FK systems, motion is defined by specifying the rotation of each joint in a hierarchical sequence. While this offers precise control for authoring specific poses, it can be less intuitive for dynamic tasks, such as ensuring a hand reaches a specific coordinate in space. IK addresses this by allowing the system to compute the necessary joint transformations automatically based on a desired end-position for a “limb” (Akenine-Möller et al., 2018; Parent, 2012).

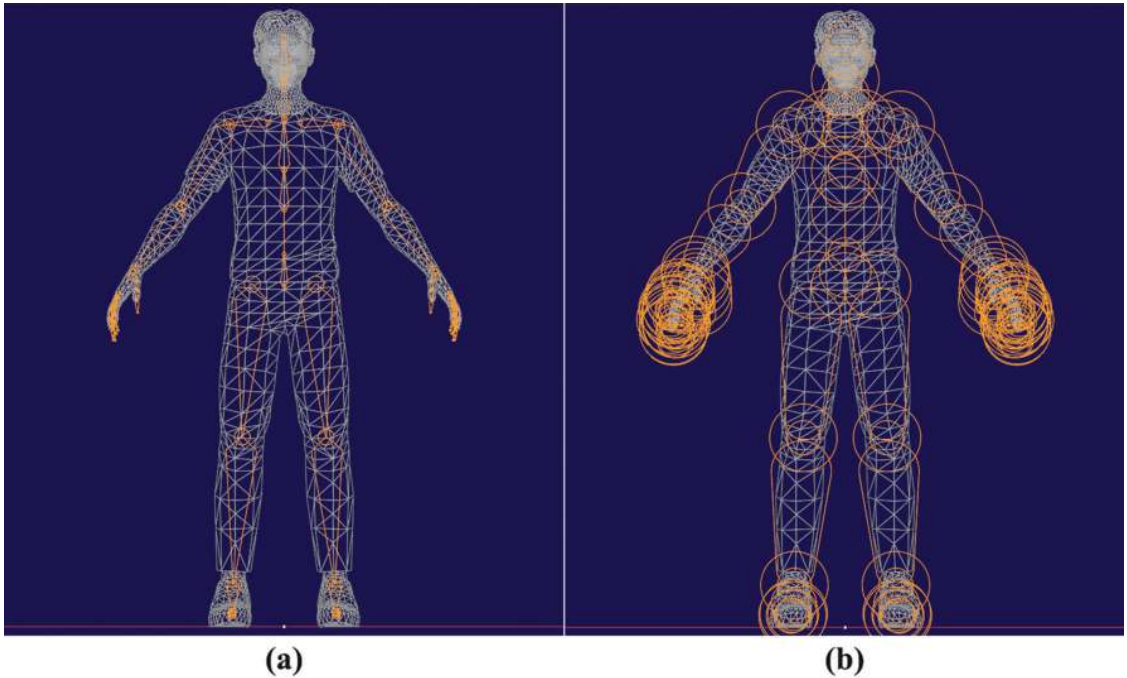


FIGURE 7.1 Character rigging and skinning visualization (blender). (a) A hierarchical skeletal rig defining the character's articulation points. (b) A representation of the deformation volumes, showing the spatial extent of each joint's influence over the character's mesh. (Model and rig by Ready Player Me (CC BY 4.0), as used in the Aurora VR prototype (Martins et al., 2026). Screenshots by the authors.)

In VR, IK is an indispensable tool for embodiment and self-representation. Since consumer VR systems often track only a sparse set of points, usually the head and hands, IK algorithms are frequently used to procedurally estimate the positions of the elbows, shoulders, and torso (Caserman et al., 2019; Jerald, 2015). Additionally, procedural IK allows characters to adapt dynamically to the environment, for instance, adjusting foot placement on irregular terrain or ensuring that a hand correctly grasps an object.

Motion Data: Keyframing, Capture, and Physics

The data that drives skeletal animation systems is typically derived from two main sources: *keyframe animation* and *motion capture (MoCap)*. In keyframe animation, specific values (e.g., joint rotations or positions) are manually defined at discrete time steps, while intermediate states are computed through interpolation (Jung & Vitzthum, 2022). This method is ideal for stylized or non-humanoid entities due to the high degree of artistic control it provides, especially when physical realism is secondary to expressive goals.

By contrast, MoCap records the movements of real actors to be mapped onto a virtual rig (Figure 7.2). Although this process was traditionally achieved through dedicated body-worn sensors, recent advancements have enabled motion acquisition via artificial intelligence (AI)-driven computer vision, which can estimate 3D skeletal poses directly from standard 2D video feeds (Zheng et al., 2024). The use of MoCap is often preferred for human-like characters as it enables the reproduction of fine and complex motion patterns that are challenging to replicate manually. Nevertheless, because human observers are exceptionally sensitive to inconsistencies in biological motion (Jerald, 2015), errors in timing, weight distribution, or joint acceleration can trigger the uncanny valley effect, where the proximity to human realism, if slightly “off”, evokes a sense of unease in the user (Mori et al., 2012).

Beyond authored animation clips, either created through keyframing or capture, *physics-based animation* is frequently used to introduce a layer of dynamic responsiveness. By representing objects as rigid

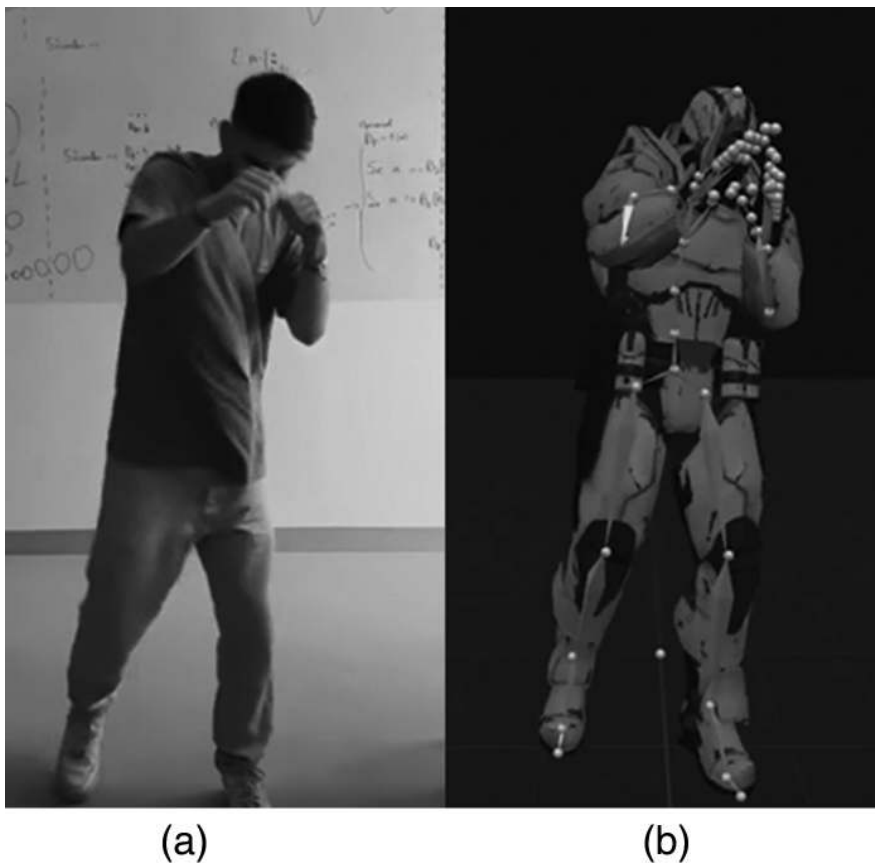


FIGURE 7.2 Example of an AI application translating human motion (a) into a functional virtual skeleton and character pose (b). (Image derived from student activities under the author’s supervision within the Master’s in Digital Games Development (University of Aveiro).)

bodies with properties such as mass, friction, and elasticity, physics engines can simulate motion governed by external forces and collisions. These simulations rely on simplified proxy geometries (collision shapes) to ensure these physical interactions can be computed within the narrow time window of a single frame (Jung & Vitzthum, 2022). This is essential for the direct manipulation VR affords, such as throwing, where the trajectory and impact of objects must be calculated in real time to match the user’s physical input.

Animation Systems and Optimization

In an interactive environment, animations usually must transition seamlessly based on user behavior or system states. To achieve this, *blending techniques* are employed to mathematically interpolate between different animation clips (e.g., a character transitioning from an “idle” pose to a “walking” cycle). This ensures continuity of motion and prevents abrupt switches between clips, which would otherwise disrupt the perceived coherence of the animation. The management of these transitions is typically handled by logic models that control how and when an entity should change its behavior. Finite state machines are the most common approach due to their conceptual simplicity and wide support in modern game engines (Jung & Vitzthum, 2022). However, as behaviors become more complex and less deterministic, more flexible alternatives include hierarchical finite state machines, decision trees, and behavior trees (Colledanchise & Ögren, 2019).

Finally, the implementation of these techniques is fundamentally determined by performance constraints, as VR systems must render two high-resolution views simultaneously. Consequently, the

computational cost of animation must be carefully managed through optimization strategies, such as *Animation Level of Detail (LOD)*, which reduces update frequencies or skeletal complexity for distant characters, and *culling techniques* that suspend calculations for animated elements outside the user's current field of view. Additionally, GPU-based skinning allows geometry deformation calculations to be offloaded from the CPU to the graphics card, a critical optimization for supporting more populated or complex scenes while maintaining the high frame rates required for user comfort (Akenine-Möller et al., 2018).

EXPRESSIVE PRINCIPLES OF ANIMATION

This section provides a concise overview of the core principles of expressive animation, originally developed and refined both technically and creatively by Disney Studios throughout much of the twentieth century. Although born within a single studio, these principles, illustrated in Figure 7.3, were widely adopted across the traditional animation industry. Due to their effectiveness and popularity, they were rapidly embraced by the global animation community, eventually becoming the definitive archetypal framework for the field.

Although the techniques and concepts underlying these principles were initially developed through informal practice, they were formally systematized by Frank Thomas and Ollie Johnston in their seminal 1981 work, *Disney Animation: The Illusion of Life*. This publication became a cornerstone reference for animation teams across diverse fields, extending its influence far beyond traditional 2D animation to become a foundational pillar for 3D digital animation, video games, and immersive environments. These principles were later discussed by Lasseter as to their application to do 3D animation (Lasseter, 1987) and reviewed by Thesen (2020) and further completed.

In summary, while the original work outlines 12 distinct principles, we shall focus on the most relevant to the contemporary context of real-time 3D animation within VR.

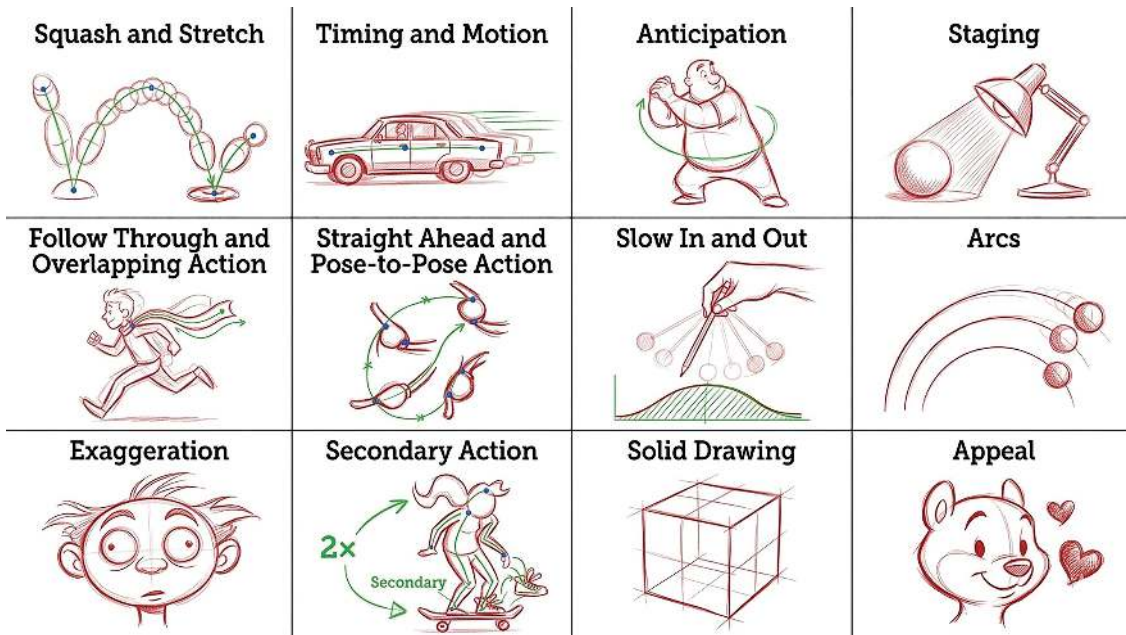


FIGURE 7.3 Illustration of the 12 principles for animation proposed by Frank Thomas and Ollie Johnston. (Image by the authors.)

Squash, Stretch, and Exaggeration

A character's face gains vitality and heightened expressivity through the application of Squash and Stretch, a technique achieved through non-uniform scaling operations. Thomas and Johnston (1981) demonstrated how this process modifies key facial features, such as the eyebrows and cheeks, making the character's emotional state more legible and communicative. In 3D animation and video games, this technique is frequently used to emphasize the direction and velocity of movement. A classic example is the "bouncing ball", which elongates as it falls and flattens upon impact with the ground. Similarly, in interactive ludic environments, an avatar's jump can be exaggerated using these same principles to enhance the feel of the action.

From a VR design perspective, however, the application of Squash and Stretch involves a critical trade-off regarding realism and verisimilitude. While stylized or thematic VR environments may benefit from such exaggeration, experiences leaning toward high-fidelity simulation may find this distortion immersion-breaking or counterproductive.

Furthermore, technical implementation presents its own challenges. Cooper (2021) notes that certain game engines used for immersive experiences do not natively support non-uniform scaling within their skeletal (bone) systems. Nevertheless, a similar (though perhaps less precise) expressive effect can still be achieved by manually scaling the rig's bones to simulate this classic distortion.

Staging

Staging is the animation principle centered on the clear and expressive communication of narrative events. Its primary goal is to ensure that the audience's attention is directed precisely where it is needed, eliminating ambiguity through visual language and composition. This is achieved by leveraging scenographic techniques such as strategic lighting, character positioning, and camera angles to frame the action effectively.

In the context of VR, staging presents a unique conceptual challenge: because the user often has total control over the camera (their own head movement), the traditional methods of framing an "unambiguous" event must be replaced by environmental cues. Instead of a fixed camera cut, the designer must use spatialized audio, light flickers, or character gazes to "stage" the scene within a 360-degree immersive space. Another frequently employed strategy, particularly within VR games, is the use of cutscenes (Cooper, 2021) or diegetic user interface (UI) elements, such as floating windows with written messages, to communicate specific narrative events. During these moments, user interactivity is often temporarily suspended. This "locking" of the player's agency ensures that critical information is conveyed without distraction, though it requires a careful balance to avoid breaking the sense of presence inherent to the immersive experience.

Anticipation, Follow-through, and Overlapping Action

The principles of anticipation, follow-through, and overlapping action constitute a fundamental triad for creating organic movement, rooted in the premise that a primary physical action rarely occurs in isolation. In fact, any expressive movement must be framed by preparatory and concluding phases that lend credibility to a character's weight and intent. Anticipation manifests as a preliminary movement, typically performed in the opposite direction of the main action, such as a character crouching and compressing their body before springing into a jump. This initial recoil not only gathers the necessary energy for the maneuver but also serves as a vital visual cue for the user, signaling the imminence of a kinetic or narrative event.

Once the primary action is initiated, the motion should not cease abruptly; instead, the principle of follow-through dictates that parts of the body and attached accessories continue to move due to inertia even after the main force has stopped, gradually returning to a state of rest. Closely linked to this concept is overlapping action, which specifies that a character should not be treated as a single, rigid mass. On the contrary, different components such as hair, clothing, or loose limbs possess their own dynamics, moving and settling at distinct speeds relative to the character's core body.

In the context of real-time 3D animation and VR, the rigorous implementation of these principles is vital to avoiding the robotic stiffness that often shatters immersion in digital environments. By ensuring that clothing and materials react with their own momentum and fluidity, the animation system reinforces a sense of physical presence and verisimilitude. This allows the user's brain to accept the avatar as a living entity integrated within the virtual space rather than a collection of static digital assets. Another significant challenge within the VR experience arises from the tension between these animation principles and the requirement for immediate system responsiveness to user input. This conflict, between the temporal demands of expressive animation and the necessity for low-latency control, represents a critical hurdle for maintaining immersion and will be explored in greater depth in the final section of this chapter.

Slow In & Slow Out and Arcs

These two animation principles, Slow In & Slow Out and Arcs, are intimately connected to the fact that real-world motion is governed by the laws of physics. Because of inertia, objects naturally tend to accelerate at the beginning of a movement and decelerate toward the end. When this behavior is ignored, the resulting animation appears rigid and unnatural. This principle applies to the majority of human bodily movements; for instance, when a character points a finger, the upward motion of the arm must accelerate from rest and decelerate before stabilizing in the air to avoid a mechanical appearance.

Similarly, most movements found in nature, whether performed by humans and animals or seen in the gentle swaying of a leaf on a tree, follow curved trajectories rather than linear paths. These Arcs are essential for visual fluidity, as movements following a strictly rectilinear outline again feel artificial and robotic. In the context of immersive environments, respecting these organic paths and timing curves is crucial for maintaining the user's suspension of disbelief, as the human eye is highly sensitive to the subtle physical cues that distinguish living motion from programmed execution.

Secondary Action

Secondary action, as the name implies, refers to the principle of incorporating animations that complement and support primary events or actions. Within the context of VR environments, these animations are essential for establishing credible world-building and reinforcing the user's perception of a living, breathing world. To prevent 3D environments from appearing static or sterile, they should be populated with subtle, short-duration looping animations, such as a character's idle movements or the gentle swaying of a curtain near a window. These secondary layers of motion serve to ground the virtual space, providing the peripheral details that convince the human brain of the environment's vitality.

CHALLENGES IN VR EXPRESSIVE ANIMATION

This final section offers a reflection on the core conceptual challenges involved in implementing expressive animation strategies within VR. As previously discussed, interactive immersive environments, driven by real-time 3D rendering, achieve significant depth through the application of classical animation

principles. These principles were extensively developed and refined throughout the latter half of the twentieth century in traditional media, and they continue to serve as a vital foundation in the modern era of digital 3D animation, visual effects, and video games.

However, VR as a medium possesses a unique set of properties and characteristics that challenge the direct application of principles originally conceived for other forms of media. A primary challenge arises from the inherent tension between the temporal precision required for expressive animation and the rigorous system responsiveness demanded by VR. Because the VR experience is interactive by nature, any delay introduced for the sake of emotional or artistic expression must be carefully balanced against the immediate feedback expected by the user's input. This is a fundamental point, widely supported by various theories within the fields of Game Studies, Interactive Environments, and Human–Computer Interaction (HCI). On the one hand, a core tenet of Flow Theory is the necessity of immediate feedback for user actions, particularly during skill-based and expert tasks. This instantaneous response is a critical factor in maintaining the user within the “flow channel”, a state of peak mental engagement characterized by a total immersion in the task, often referred to as an optimal experience (Csikszentmihalyi, 2008).

The transition from a traditional 2D or 3D type of animated content to VR demands that these principles foster motion and behavior believability within these simulated environments. Throughout the following pages we will detail how physical laws, procedural mechanics, theatrical clarity, and artistic appeal may be handled when applying the animation principles to animated content in VR.

Responsivity

In interactive and immersive experiences, perceptual fusion plays a vital role in maintaining the user's sense of presence. As an involuntary cognitive process, it allows the brain to synthesize various staged sensory stimuli, such as real-time 3D visuals, spatialized audio, and haptic feedback, into a single, unified sensation of reality. When these stimuli are synchronized with the user's expectations and actions, the boundaries between the artificial environment and the user's perception dissolve, resulting in a cohesive and believable lived experience.

According to Swink (2008), this phenomenon plays a vital role not only in the construction of interactive processes but also in the way animation is associated with these events. When a digital environment or an animation responds to a user's action in less than 100 ms, it is perceived as a direct consequence of that action, establishing a clear cause-and-effect relationship. This temporal window is critical because it allows the brain to fuse the user's motor input with the system's visual output, creating a seamless sense of agency and responsiveness that is fundamental to the immersive experience.

Consequently, we observe a persistent tension between the application of certain expressive animation principles and the necessity of maintaining a sense of user control and responsiveness. As Cooper (2021) notes, failing to balance these factors can severely compromise the ludic experience within a playful environment. While this issue is less pronounced regarding the avatar's direct body movements, since in VR, the user's physical body effectively becomes the avatar, bypassing the need for pre-rendered representative animations, it becomes highly visible when applying animation principles such as Anticipation. In practice, this conflict is most apparent when the duration of an anticipation phase or a follow-through action is excessively long. This typically occurs during environmental animations triggered by user interactions, such as using tools, operating items, or opening doors and drawers.

Kinetic Realism: Physical Laws in Virtual Spaces

The first challenge of creating any animated content is the need to make it believable or, at least, in tune with the narrative or storyline it is a part of. This is where the principles of Squash and Stretch, Slow In and Slow Out, Arcs, and Timing play an important role. In VR, these principles are worked on and applied not just for aesthetic purposes, but also as essential tools for reducing cognitive overload and preventing

discomfort during locomotion tasks by guaranteeing that virtual objects follow convincing settings and behaviors in the eye of the user. By basing virtual interactions on recognizable laws of physics, designers can foster a sense of immersion and realism that mirrors the physical dynamics users may deal with in real-life situations and reinforce values, properties, and significance attributed to characters, objects, and actions. Some of the authors' experience in projects such as VRChem (Cracow University of Technology, 2026) has taken this into consideration when designing and developing the application.

The Squash and Stretch principle acts as a vital feedback mechanism for object comportment and interaction by maintaining constant volume during distortion. Actions like when a user grasps or presses a virtual tool should preserve the illusion of the object's mass and solidity and contribute to improving the user's experiential understanding of the environment. This physical realism is, for instance, critical in VR training applications, where mechanical consistency guarantees that the virtual world presents itself and reacts in a constant manner throughout the experience. User ends up building a sense of trust and reliability as to what is being taught in the application.

The Slow In and Slow Out principle ensures that transitions, such as the movement of automated doors or elevators or running in a scene, feel smooth and organic rather than sudden and abrupt and contributes to the prevention of possible motion sickness symptoms. By respecting the laws of inertia, this principle contributes to an easier adoption of the technology by inexperienced users, allowing them to more easily adapt to the experience without the discomfort caused by unexpected and sharp movements. Similarly, the inclusion of Arcs aligns virtual movement with natural perception. Natural perception in VR refers to the users' capability to track and mimic real-world sensory experiences, enabling them to interact with digital environments by intuitively replicating behavior they would have in the real world. Whether walking through architectural spaces or interacting with a ball in a tag ball game, ensuring the environment remains close to reality frees up the users' attention and enables them to focus on their primary tasks rather than being distracted by the need to learn how the VR world they are in works. The principle associated with Timing can change the way the user perceives properties such as weight and the urgency or delayed nature of an action unfolding. In high-pressure scenarios, fast-paced timing for machines failing or emergency signals blinking very quickly creates a sense of pressure in which quick decision-making is needed, while slower timing for successful interactions rewards the user with a sense of relief and satisfaction for having accomplished something like a task or achieving a goal. Timing can also help reinforce properties perceived in relation to the physical properties of objects and characters observed or interacted with throughout the experience. Establishing these real-time feedback mechanisms is essential for laying out the rules for the users' behavioral changes and interpretation of the VR experience, bringing the user closer to its narrative and purpose (Kal & Samur, 2024).

Procedural Integrity: Structural Mechanics and Virtual Presence

Beyond aesthetics, designing VR experiences should also aim for seamless, user-friendly experiences that contribute to the sense of immersion and presence within the 360° environments. The principle of Solid Drawing therefore certifies that virtual assets possess consistent properties such as 3D volume, weight, and balance. This spatial consistency is vital for preventing errors in perception while users engage in actions like the use of teleport spots or interacting with objects and characters. Motion in VR may be further refined through the dual application of the Straight-Ahead Action and Pose-to-Pose principles. The first is ideal for rendering unpredictable, fluid elements, such as escaping steam or flickering emergency lights, and provides a realistic replication of high-stakes situations. In the case of the pose-to-pose principle defining specific keyframes for interactive tasks, such as inserting a sequence key, designers gain granular control over the timing and composition of the experience. As for the Follow-Through and Overlapping Action principle, it may help to avoid robotic nonorganic and fluid animations. In the case of experiences with, for instance, humanoid characters, this principle dictates that secondary elements, such as protective clothing or hair, continue moving while the character is moving and after the primary body stops. Integrating these animation nuances into the design of VR experiences is considered a useful addition that contributes to the sense of immersion and presence.



FIGURE 7.4 Example of the use of the staging principle in the VR game *Just Another Cooking Game* by Ruben Carvalho. (Courtesy of Carvalho, 2023.)

Theatrical Clarity: Directing Agency and Intentionality

When compared with noninteractive audiovisual content, directing a participant's attention in an immersive environment is quite a significant challenge. However, resorting to the principles of Staging, Anticipation, and Secondary Action may help to guide user agency, prevent cognitive overload (Wen et al., 2024), and facilitate efficient goal completion or skill acquisition. Anticipation serves as a preparatory motion informing the user of a forthcoming event. In VR experiences, a flickering alarm panel or a room that starts to tremble may act as a visual cue, allowing the user to prepare for a decision moment or a problem-solving task. The Staging principle involves the clear presentation of an idea and context to make its functional intent and significance unmistakable. Within VR games, staging may direct users toward specific tasks, such as restoring power or preparing for confrontation or an activity. This is applied in games such as *Just another Cooking game* (Carvalho, 2023) where users are informed as to what to do and what will happen afterward (Figure 7.4).

Proper staging minimizes the possibility of not understanding what should be done next by including animations in lighting, props, and interface elements. Theatrical Clarity in VR experiences may also rely on the application of the principle of Secondary Action, which can provide minor actions that support the understanding of the primary objective, thus adding soul and depth to the experience. Debris that is projected after an explosion or the flickering of lights adds layers to recognizable and engaging environments. These details mirror real-life environments, encouraging participants to maintain their state of flow and fostering a deeper application of previously acquired skills and knowledge as to how to react and behave.

Artistic Resonance: Appeal and the Search for the Adequate Aesthetic

Artistic resonance in VR refers to the profound emotional and sensory connection a user experiences when engaging with the VR application (Rani et al., 2024), where the digital environment makes the user to feel connected. By applying the principles of Exaggeration and Appeal in the design of the experience, creators try bridging between the experience and the user, potentially triggering a greater sense of immersion and presence.

Exaggeration as an animation principle applicable to the design of VR experiences does not imply a cartoonish over-the-top aesthetic but rather an effort to design a heightened clarity. Heightened feedback via exaggerated visual cues helps participants immediately grasp the outcomes of their decision-making.

Furthermore, things like slightly amplifying the facial expressions of avatars overcome the uncanny valley, ensuring that active listening and social cues are clearly perceived in the experience. The Appeal principle constitutes the quality that makes a virtual environment compelling and interesting to look at. By using clear shape language and balanced compositions, designers project an inclusive aesthetic that may help the user reach a potential optimal flow. Ultimately, appeal transforms a simulated environment into a space where users feel a sense of belonging, encouraging them to engage with the virtual scenarios in an involving and purposeful way (Rani et al., 2024).

REFERENCES

- Akenine-Möller, T., Haines, E., Hoffman, N., Pesce, A., Iwanicki, M., & Hillaire, S. (2018). *Real-time rendering* (4th ed.). A K Peters/CRC Press. <https://doi.org/10.1201/b22086>.
- Bear, M. F., Connors, B. W., & Paradiso, M. A. (2021). *Neurociências: desvendando o sistema nervoso* (4th ed.). Artmed.
- Carvalho, R. (2023). *Just another cooking game* (Version 1.04) [Video game]. <https://sidequestvr.com/app/23468/just-another-cooking-game>.
- Carvalho, R. J. M. (2025). Os fatores que influenciam a imersão e a presença em videojogos de realidade virtual [The factors that influence immersion and presence in virtual reality video games] (Doctoral dissertation, Universidade de Aveiro & Universidade do Porto) available online : <https://ria.ua.pt/handle/10773/47128>. With the author's permission.
- Caserman, P., Achenbach, P., & Gobel, S. (2019). Analysis of inverse kinematics solutions for full-body reconstruction in virtual reality. In *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)*, Kyoto, Japan (pp. 1–8). <https://doi.org/10.1109/SeGAH.2019.8882429>.
- Colledanchise, M., & Ögren, P. (2019). *Behavior trees in robotics and AI: An introduction*. CRC Press.
- Cooper, J. (2021). *Game anim: Video game animation explained* (2nd ed.). CRC Press. <https://doi.org/10.1201/9781003147893>.
- Cracow University of Technology. (2026). *Innovative integration of virtual reality in chemical engineering education*. Cracow University of Technology. <https://vrchem.pk.edu.pl/>.
- Csikszentmihalyi, M. (2008). *Flow: The psychology of optimal experience* (1st ed.). HarperCollins.
- Debarba, H. G., Chague, S., & Charbonnier, C. (2022). On the plausibility of virtual body animation features in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(4), 1880–1893. <https://doi.org/10.1109/TVCG.2020.3025175>.
- Gleitman, H., Gross, J. J., & Reisberg, D. (2011). *Psychology* (8th ed.). W.W. Norton.
- Hettinger, L. J., Berbaum, K. S., Kennedy, R. S., Dunlap, W. P., & Nolan, M. D. (1990). Vection and simulator sickness. *Military Psychology*, 2(3), 171–181. https://doi.org/10.1207/s15327876mp0203_4.
- Jerald, J. (2015). The VR book: Human-centered design for virtual reality. In M. T. Ozsu (Ed.), *The VR book* (1st ed.). Association for Computing Machinery. <https://doi.org/10.1145/2792790>.
- Jung, B., & Vitzthum, A. (2022). Virtual worlds. In R. Doerner, W. Broll, P. Grimm, & B. Jung (Eds.), *Virtual and augmented reality (VR/AR): Foundations and methods of extended realities (XR)* (pp. 71–106). Springer. https://doi.org/10.1007/978-3-030-79062-2_3.
- Kal, O., & Samur, Y. (2024). Educational virtual reality game design for film and animation. In N. Lee (Ed.), *Encyclopedia of computer graphics and games* (pp. 621–636). Springer. https://doi.org/10.1007/978-3-031-23161-2_168.
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3.
- Khan, S., & Chang, R. (2013). Anatomy of the vestibular system: A review. *NeuroRehabilitation*, 32(3), 437–443. <https://doi.org/10.3233/NRE-130866>.
- Lasseeter, J. (1987). Principles of traditional animation applied to 3D computer animation. *ACM SIGGRAPH Computer Graphics*, 21(4), 35–44. <https://doi.org/10.1145/37402.37407>.
- La Valle, S. M. (2023). *Virtual reality*. Cambridge University Press. <https://doi.org/10.1017/9781108182874>.
- La Viola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1), 47–56. <https://doi.org/10.1145/333329.333344>.

- Le, B. H., & Hodgins, J. K. (2016). Real-time skeletal skinning with optimized centers of rotation. *ACM Transactions on Graphics*, 35(4), 1–10. <https://doi.org/10.1145/2897824.2925959>.
- Martins, S., Vairinhos, M., Soares, S., Félix, A., Araújo, C., & Diniz, F. (2026). Evaluating usability and immersion in a virtual reality game for fear of flying: A pilot study with a non-phobic sample. In L. T. De Paolis, P. Arpaia, & M. Sacco (Eds.), *Extended reality* (Vol. 15737, pp. 213–225). Springer. https://doi.org/10.1007/978-3-031-97763-3_16.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>.
- Parent, R. (2012). *Computer animation: Algorithms and techniques* (3rd ed.). Morgan Kaufmann.
- Rani, N. A., Hashim, M. E. A., & Idris, M. Z. (2024). Literature survey: The potential of integrating immersive experience and aesthetic experience in virtual reality historical event. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 33(3), 112–123. <https://doi.org/10.37934/araset.33.3.112123>.
- Reason, J. T. (1978). Motion sickness adaptation: A neural mismatch model. *Journal of the Royal Society of Medicine*, 71(11), 819–829. <https://doi.org/10.1177/014107687807101109>.
- Reason, J. T., & Brand, J. J. (1975). *Motion sickness*. Academic Press.
- Reisinho, P., Silva, C., Ferreira, M., Raposo, R., Vairinhos, M., & Zagalo, N. (2024). An analysis of the sense of presence and cybersickness in virtual reality: The influence of content type, exposure time, and gender. In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, augmented and mixed reality* (Vol. 14707, pp. 263–282). Springer. https://doi.org/10.1007/978-3-031-61044-8_19.
- Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., & Oliva, R. (2022). A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in Virtual Reality*, 3, 914392. <https://doi.org/10.3389/frvir.2022.914392>.
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6), 603–616. <https://doi.org/10.1162/pres.1997.6.6.603>.
- Spanlang, B., Normand, J.-M., Borland, D., Kiltner, K., Giannopoulos, E., Pomés, A., González-Franco, M., Perez-Marcos, D., Arroyo-Palacios, J., Muncunill, X. N., & Slater, M. (2014). How to build an embodiment lab: Achieving body representation illusions in virtual reality. *Frontiers in Robotics and AI*, 1. <https://doi.org/10.3389/frobt.2014.00009>.
- Swink, S. (2008). *Game feel: A game designer's guide to virtual sensation*. CRC Press. <https://doi.org/10.1201/9781482267334>.
- Thesen, T. P. (2020). Reviewing and updating the 12 principles of animation. *Animation*, 15(3), 276–296. <https://doi.org/10.1177/1746847720969919>.
- Thomas, F., & Johnston, O. (1981). *Disney animation: The illusion of life*. Abbeville Press.
- Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: A review. *Frontiers in Psychology*, 10, 158. <https://doi.org/10.3389/fpsyg.2019.00158>.
- Wen, P., Lu, F., & Mohamad Ali, A. Z. (2024). Using attentional guidance methods in virtual reality laboratories reduces students' cognitive load and improves their academic performance. *Virtual Reality*, 28(2), 110. <https://doi.org/10.1007/s10055-024-01012-0>.
- Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz, N., & Shah, M. (2024). Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, 56(1), 1–37. <https://doi.org/10.1145/3603618>.

Agnieszka Dubiel

ABSTRACT

This chapter examines how user interfaces in virtual reality (VR) can be spatially anchored and narratively integrated to support presence, embodiment, and cognitive performance. Drawing on theories of spatial anchoring, interaction fidelity, and mental effort, it first outlines a conceptual structure for understanding how interface placement shapes perception and action. This chapter then defines interfaces in terms of their spatial and narrative properties and analyses their strengths and weaknesses. Based on this, it proposes a comparative perspective in which VR user interface design is treated as a context-dependent trade-off rather than a choice of a single “best” paradigm. The final sections discuss hybrid and adaptive configurations that combine diegetic and world-related elements and discuss how mechanisms that respond to biology and behaviour can help regulate the amount of information and mental effort. These analyses provide a set of design guidelines for structuring spatial interfaces in VR.

INTRODUCTION

Immersive technology has undergone a dramatic transformation since its inception. Initially, it served as an experimental research system, but it is now used in many areas of life, including education, health-care, training, and simulations. With the growing popularity, accessibility, and technical sophistication of immersive systems, experts in human–computer interaction have encountered the challenge of designing user interfaces (UIs) for virtual environments (Jerald, 2015; Mine, 1995; La Viola et al., 2017). In contrast to traditional two-dimensional graphical UIs, numerous factors must be considered when creating these interfaces (Bowman & McMahan, 2007; Jerald, 2015). Most importantly, virtual reality (VR) interfaces operate in immersive contexts, directly engaging the mechanisms of perception, spatial cognition, and embodiment. Furthermore, experts point out that two-dimensional screen metaphors cannot be directly transferred to augmented environments (Mine, 1995; La Viola et al., 2017; Jerald, 2015; Bowman & McMahan, 2007).

The most important concept in immersive environments is presence, which is commonly defined as the subjective experience of “being there” in a mediated context (Slater & Wilbur, 1997; Sanchez-Vives & Slater, 2005). This phenomenon is formed not only by graphic realism but also by other aspects such as interaction consistency, spatial congruence, and narrative integration (Slater & Wilbur, 1997). Closely related to presence is the concept of embodiment, which refers to the degree to which users perceive the virtual body as their own and integrate it into their body schema. It is also worth remembering that interface elements that disrupt body coherence or spatial stability can weaken both concepts, whereas spatially integrated systems can support them (Kilteni et al., 2012).

From a cognitive perspective, immersive environments demand unfamiliar processing. Research shows that immersive VR increases cognitive demand compared to less immersive formats, especially when users must process and integrate spatial information with task goals (Makransky et al., 2019). Mental effort theory helps us understand how excessively complex interfaces hinder learning and task performance (Sweller, 1988).

In this context, the spatial anchoring of interface elements begins to play a key role. Interfaces can be directly linked to the user’s point of view (head-locked), embedded in the virtual environment (world-locked), or embedded in the narrative logic of the virtual world (diegetic). Types of VR interfaces are often mentioned in both industry and academic discussions, yet a systematic comparison of their architectural attributes and cognitive implications remains limited (McMahan et al., 2012; Bowman & McMahan, 2007; La Viola et al., 2017).

Diegetic interfaces are elements of the UI embedded within the narrative context of the virtual world and designed to support immersion by aligning information with the world’s logic (La Viola et al., 2017). In contrast, world-synchronised interfaces remain static in the environment, delivering stability and clarity. Rather than a single optimal paradigm, VR interface design is therefore treated here as a context-dependent decision space determined by task requirements, cognitive constraints, and empirical goals.

CONCEPTUAL FOUNDATIONS

Before distinguishing between diegetic and non-diegetic interfaces, it is necessary to outline the conceptual constructs that frame spatial interface design in immersive environments. Interface placement in VR is not simply a graphical choice; it is tightly linked to concepts such as presence, embodiment, interaction fidelity, and cognitive demand (Slater & Wilbur, 1997; Bowman & McMahan, 2007). The key constructs referred to throughout this chapter are summarised in Table 8.1.

TABLE 8.1 Key theoretical constructs for VR interface design

CONSTRUCT	BRIEF DESCRIPTION
Presence	Subjective sense of “being there” in a virtual environment, based on coherent sensorimotor contingencies.
Plausibility illusion	The impression that events in the environment are really happening and follow the world’s internal logic.
Embodiment	Sense of body ownership, self-location, and agency with respect to an avatar or virtual body.
Interaction fidelity	Degree to which actions in VR resemble those in the physical world (temporally, spatially, kinaesthetically).
Spatial anchoring	Relational positioning of interface elements with respect to the user and the environment (head-, world-, diegetic anchoring).
Cognitive load	Number of cognitive resources required to perform a task with a given interface configuration.
Spatial cognition	Mechanisms of building cognitive maps and navigating in virtual environments.

Spatial Anchoring and Interaction Fidelity

Spatial anchoring refers to the relational positioning of interface elements with respect to the user and the virtual environment. Early work on virtual environments highlighted that three-dimensional UIs differ fundamentally from traditional 2D GUIs due to depth perception, embodied manipulation, and complex spatial reference frames (Mine, 1995). As a result, 3D UIs must align perceptual cues with interaction methods to maintain both usability and cognitive coherence (La Viola et al., 2017).

Bowman and McMahan (2007) proposed viewing immersion as a continuum defined by display and interaction fidelity, rather than a binary property. Interaction fidelity describes how closely actions in VR resemble those in the physical world, including temporal, spatial, and kinaesthetic aspects (Bowman & McMahan, 2007; Jerald, 2015). Interface anchoring strategies directly influence this dimension: world-locked and diegetic interfaces tend to preserve spatial stability, whereas head-locked overlays can disrupt depth signals and integration with the scene (Jerald, 2015). Empirical work by McMahan et al. (2012) showed that higher interaction fidelity can improve task performance but may also introduce additional complexity when interaction demands exceed users' perceptual capabilities (Figure 8.1).

Presence, Plausibility, and Embodiment

Presence is widely regarded as a defining feature of immersive systems and is commonly defined as the subjective sense of “being there” in a mediated environment (Slater & Wilbur, 1997; Sanchez-Vives & Slater, 2005). Slater (2009) improves this concept by distinguishing between place illusion (the feeling of being in a particular location) and plausibility illusion (the feeling that events in the environment are really occurring).

Sanchez-Vives and Slater (2005) emphasise the role of consistent sensorimotor contingencies: if UI elements violate depth relations, occlusion, or causality, they disrupt the expected link between action and consequence. Diegetic interfaces, by embedding information within the world's spatial and narrative structure, are more capable of supporting plausibility than “floating” overlays that sit outside that logic (La Viola et al., 2017).



FIGURE 8.1 Head-locked HUD in a VR rhythm game: score and status indicators stay fixed to the user's view (Beat Saber, Meta Quest). (Credit: Agnieszka Dubiel.)

Research on embodiment further clarifies how interface positioning shapes user experience. Kiltner et al. (2012) characterise embodiment as comprising body ownership, self-location, and agency. UI elements that correspond with the user's body schema, for example, wrist-mounted or hand-attached systems, can strengthen agency and reduce fragmentation between "body" and "tool," whereas separate, free-floating interfaces may create a split between action space and information space (La Viola et al., 2017). Studies on body ownership illusions show that multisensory coherence strongly enhances embodiment (Botvinick & Cohen, 1998; Slater et al., 2010; Petkova & Ehrsson, 2008), suggesting that spatially matching UI elements to body representations and providing congruent visual-tactile cues may improve the coherence of immersive interactions.

Mental Effort and Spatial Cognition in VR

Cognitive demand theory distinguishes between intrinsic load, related to the complexity of the material, and extraneous load, induced by the way information is presented (Sweller, 1988). In immersive environments, extraneous load can increase when users must manage conflicting spatial frames of reference or divide attention across multiple information layers (Sweller, 1988; Wickens, 2002).

Makransky et al. (2019) found that immersive VR can increase perceived presence while simultaneously raising mental effort and, under some conditions, reducing learning outcomes. Studies of spatial cognition in VR show that environmental indicators and consistent spatial references strongly influence navigation and the formation of cognitive maps (Darken & Sibert, 1996). Jerald (2015) synthesises these conclusions by emphasising that depth signals, motion parallax, and motor coordination all add to overall mental effort.

DIEGETIC INTERFACES

Diegetic interfaces are one of the most characteristic approaches to UIs in VR. Unlike classic heads-up display (HUD) overlays, which are treated as an additional layer "superimposed" on the image, a diegetic interface is embedded in the depicted world. What is more, this type of interface is subject to the same spatial and narrative rules as other objects in the environment.

Definition and Criteria

In this publication, diegetic interfaces are defined based on three conditions (Dickinson et al., 2021; Dörner et al., 2022):

1. The UI element exists as an object, surface, or phenomenon in the virtual environment.
2. Its presence is narratively justified – there is a coherent reason why this element exists in a specific form and location.
3. The element can be perceived by other entities in the world, not just by the user's camera.

This definition extends earlier descriptions of diegetic UI as controls and information that "appear as part of the simulated world itself, rather than on a separate menu system or screen overlay" (Dickinson et al., 2021; Table 8.2).

Typology of Diegetic Interfaces

Although diegetic interfaces can take many forms, three recurring patterns are especially relevant for VR: object-embedded interfaces, body-integrated interfaces, and environmental projections (La Viola et al.,

TABLE 8.2 Operational criteria used to define diegetic interfaces in VR

CRITERION	DESCRIPTION	DESIGN IMPLICATION
In-world existence	The UI element is realised as an object, surface, or phenomenon within the virtual environment.	Requires spatial anchoring and integration with 3D geometry.
Narrative justification	The element has a function or rationale that is consistent with the world's internal logic (narrative, functional, social).	UI form and behaviour must align with the world's affordances and fiction.
Shared perceptibility	The element can, at least in principle, be perceived by other entities in the world, not only by the user's camera.	Excludes purely camera-bound overlays; supports plausibility and social presence.

**FIGURE 8.2** Diegetic object-embedded interface: controls appear as panels and buttons on an in-world console operated by the user (First Steps, Meta Quest). (Credit: Agnieszka Dubiel.)

2017; Billinghurst et al., 2015). These types are not mutually exclusive but provide a useful vocabulary for analysing spatial anchoring and cognitive implications (Figure 8.2).

Object-embedded interfaces integrate UI elements into specific virtual objects, such as panels, devices, or tools. Information and controls are tightly coupled to the object of interest, supporting clear affordances and lessening the need for global overlays (Saalfeld et al., 2015; Dickinson et al., 2021). Body-integrated interfaces attach UI elements to the user's virtual body or to body-proximate artefacts, for example, wrist-mounted displays or hand-referenced menus (Lediaeva & La Viola, 2020; Reiter et al., 2022). These designs exploit body-centred reference frames and can strengthen agency and embodiment (Kiltner et al., 2012; Slater et al., 2010). Environmental projections distribute UI across the environment's surfaces and structures as projections, holographic displays, or spatial annotations (Billinghurst et al., 2015; Saalfeld et al., 2015). Such interfaces leverage large-scale spatial organisation and stable landmarks, supporting navigation and cognitive mapping in complex virtual worlds (Darken & Sibert, 1996; Table 8.3).

In practice, immersive systems usually combine these approaches. For instance, an interface integrated into the body may activate diegetic object panels or launch environmental projections (La Viola et al., 2017; Dickinson et al., 2021).

TABLE 8.3 Typology of diegetic interfaces, including main advantages and limitations

<i>INTERFACE TYPE</i>	<i>BRIEF DESCRIPTION</i>	<i>MAIN ADVANTAGES</i>	<i>MAIN LIMITATIONS</i>
Object-embedded	UI elements integrated into specific in-world objects (panels, devices, artefacts).	Strong coupling to task context, clear affordances.	Susceptible to occlusion, dependent on object position and visibility.
Body-integrated	UI elements placed on the user's (virtual) body or in near peripersonal space.	Supports embodiment, reduces visual search time for controls.	Risk of fatigue, occlusion of the scene, increased motor demands.
Environmental projections	UI distributed across environmental surfaces and structures (projections, annotations, volumetric displays).	Good spatial scalability, supports navigation and cognitive mapping.	Risk of visual clutter, reduced legibility in highly dynamic scenes.

Design Advantages of Diegetic Interfaces

The designer limits the number of elements that clearly signal the presence of an additional layer of the system by embedding information in the depicted world (Dickinson et al., 2021; Koehle et al., 2021). According to Slater (2009), consistency between stimuli and the environment's logic supports both the illusion of place and the illusion of credibility. A diegetic interface, if spatially and narratively consistent, can reinforce the user's belief that the events they observe are “really” happening in the VR world (Sanchez-Vives & Slater, 2005).

It is also worth mentioning integration with embodiment. Body-integrated interfaces leverage the fact that users develop a sense of “ownership” of their virtual body and agency over its movements (Kilteni et al., 2012; Slater et al., 2010). By placing the interaction on the user's hand, wrist, or forearm, the designer can refer to motor patterns and body-based frames of reference (Lediaev & La Viola, 2020; Reiter et al., 2022). An additional advantage is that a diegetic interface, which distributes information across objects, the body, and the environment, can reduce local density and enable better use of available spatial resources (Dickinson et al., 2021; Koehle et al., 2021).

Restrictions and Obstacles

Diegetic interfaces also have major limitations. One of them is the scalability of information. In the case of UI elements, it is difficult to present large volumes of data without risking overloading the environment or severely fragmenting attention, as they are associated with specific objects, body parts, or locations (Saalfeld et al., 2015; Dickinson et al., 2021). In tasks calling for simultaneous monitoring of multiple parameters, the designer must decide on a compromise between local information accuracy and global situational awareness. Furthermore, interfaces which rely strongly on subtle spatial signals, small typography, or precise relationships between elements may be harder for users with visual, motor, or cognitive impairments (Makransky et al., 2019; Makransky & Petersen, 2021).

Diegetic interfaces are also particularly sensitive to scene dynamics. For example, rapid user movements, changes in lighting, or moving objects may greatly reduce the readability of the environment and labels embedded in objects – especially when the user must simultaneously respond to events in the environment (Saalfeld et al., 2015; Makransky et al., 2019). As a result, the diegetic interface should not be treated as a universal solution; instead, it is a powerful design strategy that needs to be balanced with requirements for information quantity, accessibility, and scene dynamics (Slater, 2009; Sanchez-Vives & Slater, 2005).

WORLD-LOCKED INTERFACES

In contrast to diegetic interfaces, which are tightly integrated into the fictional logic of the virtual world, the following discussion turns to world-locked interfaces. These interfaces prioritise spatial stability and clarity of information. This section defines world-locked UI and describes its architectural properties, applications, and limitations.

Definition

World-locked interfaces are those in which UI elements are fixed in the environment's three-dimensional space and therefore do not move with the user's head (Jerald, 2015). These items remain connected to a specific location or structure inside the virtual world, maintaining a fixed position regardless of viewpoint changes (La Viola et al., 2017). Unlike head-locked interfaces, they do not function as a “stuck-on” layer in front of the eyes, but unlike diegetic interfaces, they do not require a strong narrative or social justification (Dickinson et al., 2021; Dörner et al., 2022). They can therefore be treated as an intermediate form: they are not a classic HUD (flying overlay), but they also do not fully merge with the presented world (Sharples et al., 2025).

Structural Characteristics

The most important feature of world-locked interfaces is their fixed position in the world space: the panel or widget remains in the same place relative to the scene, while the user can move freely around it (Darken & Sibert, 1996; Jerald, 2015). Such anchoring supports the creation of cognitive maps and orientation. The panels used can serve both as “landmarks” and as information carriers (Darken & Sibert, 1996). The second feature is spatial scalability. Stable embedding in the world allows information to be spread over larger surfaces (e.g., walls, boards, virtual screens) without irrevocably obscuring the central field of view (Saalfeld et al., 2015; La Viola et al., 2017). The third aspect worth noting is modularity. World-locked elements can be designed as modular panels and widgets that can be easily modified, duplicated, and moved between scenes. Spatial anchoring systems in development tools maintain the steadiness of these items even in large, multi-room VR or MR (mixed reality) environments (Billinghurst et al., 2015).

Applications

Below are several examples of how these interfaces can be used. Of particular note is education, where world-locked panels commonly function as virtual blackboards, projection screens, or “knowledge walls” that users can stop at to analyse material from different distances. Research on VR classrooms indicates that shared, world-embedded information surfaces promote attention synchronisation and facilitate the organisation of group activities (Renman, 2020).

In medicine and clinical training, world-locked interfaces appear as panels displaying patient data, treatment algorithms, or monitoring results (Makransky & Petersen, 2021; Dörner et al., 2022). The fixed positions of these items in the “virtual room” ensure predictability – the user quickly learns where to look for critical information, regardless of gaze direction.

In industrial simulators, on the other hand, world-locked UI is used to display system statuses, step-by-step procedures, or spatial maps assigned to specific work areas (Jerald, 2015; Billinghurst et al., 2015). This arrangement minimises interference with the egocentric perspective field of view while allowing easy access to contextual information near objects or workstations.

Limitations

Stable world-locked panels can lessen perceived immersion in certain configurations, especially when their graphic style differs significantly from the rest of the environment or when there are too many of them (Sharples et al., 2021). The lack of full narrative integration causes some users to see them as “system overlays” rather than natural elements of the world (Slater, 2009).

Additional challenges appear from conflicts of depth and attention. Panels placed too close to the user can interfere with depth signals, obscure objects, and make it difficult to track movement in the scene (Jerald, 2015; Saalfeld et al., 2015). Research on head-locked and world-locked presentation in immersive systems also indicates that different anchoring methods have different effects on performance and cognitive demand. In this comparison, world-locked is sometimes subjectively preferred, but it does not always result in the shortest task completion times (Ghasemi et al., 2021). The choice of this strategy should therefore be based on an analysis of particular tasks and user needs, rather than on the assumption of the “default” superiority of one option.

COMPARATIVE STUDY

Immersion–Usability Trade-Off

From a design perspective, diegetic and world-locked interfaces can be viewed as two points on a continuum between immersion and usability (Slater, 2009; Dickinson et al., 2021). Diegetic interfaces, thanks to their strong connection to the logic of the world and its causal structure, promote high immersion and credibility of the experience, but their usability is highly context-dependent (i.e., on scene density, user profile, and task type) (Sanchez-Vives & Slater, 2005; La Viola et al., 2017).

World-locked interfaces, on the other hand, offer more stable and predictable usability, as the panels are easy to find, fit within the space, and their functions are usually clear (Jerald, 2015). In addition, their impact on immersion is moderate. Namely, they may slightly “pull” the user out of the fiction, but in return, they deliver clarity and control over information (Sharples et al., 2025). In simple terms, moving from diegetic to world-locked interfaces increases clarity but reduces immersion.

Accessibility and Scalability

From a scalability perspective, world-locked UI usually has an advantage because it allows the creation of vast information spaces in which the user controls the level of detail by physically moving closer or further away (Saalfeld et al., 2015; Billinghurst et al., 2015). Diegetic interfaces are more “cramped.” This is because, in their case, information must remain consistent with the medium (object, body, or fragment of the environment), which makes it difficult to present large, dynamic datasets without the risk of clutter (Dickinson et al., 2021).

In terms of accessibility, however, world-locked interfaces are easier to adapt to multiple needs: designers can manipulate the size, contrast, height, and distance of panels using practices known from the design of accessible 2D interfaces (Makransky et al., 2019; Wickens, 2002). Diegetic interfaces, which are closely linked to the fiction of the world and the body schema, can be more difficult to adapt; for example, elements on the wrist are less user-friendly for people with mobility or visual impairments (Makransky & Petersen, 2021; Koehle et al., 2021). These differences are summarised in Table 8.4.

TABLE 8.4 Diegetic vs. world-locked UI: scalability and accessibility

ASPECT	DIEGETIC UI	WORLD-LOCKED UI
Scalability	Limited by the carrier (object, body, or environment); harder to present large data sets.	Easy spatial expansion of panels and “information walls.”
Accessibility	Strongly dependent on fiction and body schema; harder to adapt to diverse user needs.	Size, contrast, and spatial position of panels can be adjusted more flexibly.

Context Dependence

The choice between diegetic and world-locked solutions (or a combination of both) should be determined by the task context, cognitive requirements, and user profile (Sweller, 1988; Wickens, 2002). Narrative- and embodiment-oriented experiences benefit more from diegetic strategies. Meanwhile, training, educational, and industrial applications often require world-locked panels due to scalability and transparent control over the user’s mental effort (Makransky et al., 2019).

HYBRID AND ADAPTIVE INTERFACES

Hybrid Interfaces

Hybrid interfaces combine diegetic and world-locked elements into a single coherent configuration, leveraging the advantages of both approaches (Dickinson et al., 2021; Sharples et al., 2025). Typically, diegetic components are responsible for preserving immersion, credibility, and embodiment, while world-locked elements serve as stable carriers of critical information (Jerald, 2015; La Viola et al., 2017; Figure 8.3).



FIGURE 8.3 World-locked system panel in a VR home: the application library is anchored as a single large surface in virtual space (Meta Quest interface). (Credit: Agnieszka Dubiel.)

For example, environmental or tool information can be represented by diegetic objects, while key task parameters (time, safety, progress) are displayed on a world-locked panel near the user's work area (Billinghurst et al., 2015). Research on diegetic tools in VR simulations suggests that this technique reduces the overload of a single interface layer and improves subjective control over the task (Dickinson et al., 2021).

Adaptive Interfaces

Adaptive interfaces in VR use behavioural and biological signals to automatically adjust the way information is presented and the interface configuration (Makransky et al., 2019). For example, based on estimated cognitive demand, the system may temporarily dim less relevant diegetic elements or strengthen world-locked cues when the user becomes disoriented (Makransky et al., 2019; Makransky & Petersen, 2021).

The proposed adaptation framework includes, among other things, dynamic control of the amount of information, adjustment of panel distance and size, and contextual activation of “support layers” for users with higher cognitive demand or specific needs (Makransky et al., 2019; Wickens, 2002). Such solutions correspond to the concept of bio-responsive systems, in which the interface serves as an active mechanism for managing mental and sensory resources (Makransky & Petersen, 2021).

DESIGN GUIDELINES

Finally, this section summarises the analyses above into concise design guidelines for spatial interfaces in VR. In doing so, it clearly translates conceptual constructs and empirical research findings into recommendations for practitioners.

Spatial Placement

Research on VR visual ergonomics indicates that the comfortable zone for displaying information for close-up elements (head-locked, body-integrated) is located at (Jerald, 2015; Ghasemi et al., 2021):

- A distance of approx. 0.5–1.5 m.
- Within $\pm 30^\circ$ on the horizontal axis relative to the direction of gaze.

Larger-scale world-locked panels can be placed further away as fixed navigation targets, provided that they remain legible (La Viola et al., 2017). To avoid straining the user, it is important to avoid placing key elements at extreme distances or at large angles (Jerald, 2015).

Information Density

In UI design, it is worth applying the principle of progressive disclosure, i.e., displaying information in layers, revealing details only when necessary (La Viola et al., 2017). Diegetic interfaces are ideal for contextual activation, e.g., additional data appears when approaching an object. World-locked interfaces, on the other hand, should be limited to content relevant to the current task to avoid cluttering the scene (Saalfeld et al., 2015).

Compatibility with Embodiment

This means, among other things, avoiding unintuitive positions (e.g., menus that constantly “hang” outside the natural range of motion) and preserving consistency between body movement and UI behaviour (Lediaeva & La Viola, 2020). World-locked interfaces, on the other hand, should minimise the need to frequently look away from the task space. In this case, critical data should be placed closer to the user’s movement trajectory, rather than in distant, hard-to-see areas (Wickens, 2002).

Minimising Cognitive Fragmentation

Multiple resource theory suggests that carrying out tasks that draw on the same attentional resources leads to strong interference (Wickens, 2002). In the context of VR, this means avoiding the simultaneous use of multiple spatially separated visual-spatial channels. The designer should strive for spatial stability of key elements during a given task sequence and limit the number of active interface layers (Sweller, 1988; Makransky et al., 2019).

REFERENCES

- Billingshurst, M., Clark, A., & Lee, G. (2015). A survey of augmented reality. *Foundations and Trends® in Human-Computer Interaction*, 8(2–3), 73–272.
- Botvinick, M., & Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature*, 391(6669), 756–756.
- Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: How much immersion is enough? *Computer*, 40(7), 36–43.
- Darken, R. P., & Sibert, J. L. (1996, April). Wayfinding strategies and behaviors in large virtual worlds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, British Columbia, Canada (pp. 142–149) ACM.
- Dickinson, P., Cardwell, A., Parke, A., Gerling, K., & Murray, J. (2021, March). Diegetic tool management in a virtual reality training simulation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, Lisbon, Portugal (pp. 131–139). IEEE.
- Doerner, R., Broll, W., Jung, B., Grimm, P., Göbel, M., & Kruse, R. (2022). Introduction to virtual and augmented reality. In R. Dörner, W. Broll, P. Grimm, B. Jung, M. Göbel, & R. Kruse (Eds.), *Virtual and augmented reality (VR/AR) foundations and methods of extended realities (XR)* (pp. 1–37). Springer.
- Ghasemi, Y., Singh, A., Kim, M., Johnson, A., & Jeong, H. (2021, September). Effects of head-locked augmented reality on user’s performance and perceived workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Denver, CO, USA (Vol. 65, No. 1, pp. 1094–1098). SAGE Publications.
- Gonzalez-Franco, M., & Peck, T. C. (2018). Avatar embodiment: Towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5, 74.
- Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.
- Kiltner, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4), 373–387.
- Köhle, K., Hoppe, M., Schmidt, A., & Mäkelä, V. (2021, August). Diegetic and non-diegetic health interfaces in VR shooter games. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, & G. Desolda (Eds.), *IFIP conference on human-computer interaction* (pp. 3–11). Springer.
- La Viola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: Theory and practice*. Addison-Wesley Professional.
- Lediaeva, I., & La Viola, J. (2020). Evaluation of body-referenced graphical menus in virtual environments. In *Proceedings of Graphics Interface 2020*, Toronto, ON, Canada (pp. 245–252). Canadian Information Processing Society.
- Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, 33(3), 937–958.

- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236.
- McMahan, R. P., Bowman, D. A., Zielinski, D. J., & Brady, R. B. (2012). Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Transactions on Visualization and Computer Graphics*, 18(4), 626–633.
- Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(12), 1321–1329.
- Mine, M. R. (1995). *Virtual environment interaction techniques*. UNC Chapel Hill CS Dept.
- Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLOS ONE*, 3(12), e3832.
- Reiter, K., Pfeuffer, K., Esteves, A., Mittermeier, T., & Alt, F. (2022, June). Look & turn: One-handed and expressive menu interaction by gaze and arm turns in VR. In *2022 Symposium on Eye Tracking Research and Applications*, Seattle, WA, USA (pp. 1–7). ACM.
- Renman, F. (2020). *Multi-user classroom environment in virtual reality: Creating and implementing new features* (Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden).
- Saalfeld, P., Glaßer, S., & Preim, B. (2015). 3D user interfaces for interactive annotation of vascular structures. In *Mensch und Computer 2015—Proceedings*, Stuttgart, Germany (pp. 255–258). De Gruyter Oldenbourg.
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332–339.
- Sharples, J., Williams, B., & Reeves, S. (2025, September). Designing UI for a VR environment: Diegetic and non-diegetic approaches. In *CGVC25*, London (pp. 1–5). The Eurographics Association.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549–3557.
- Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLOS ONE*, 5(5), e10564.
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6), 603–616.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159–177.

PART III

Narrative, Social, Ethics, and Accessibility



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Narrative and Storytelling in Virtual Reality

9

María Cerdeira

ABSTRACT

Storytelling in virtual reality (VR) integrates traditional narrative principles with the medium's immersive and interactive potential. This chapter examines narrative design from foundational concepts to VR applications, beginning with key ideas such as narration, storytelling, and narrative structures that organize actions, characters, time, and space. It then explores how these elements transform in VR, where users shift from passive observers to active participants or prosumers who influence and co-create the story. This chapter also analyzes immersive narrative mechanisms, including focalization, ocularization, and auricularization, as well as the player's hybrid role as both actor and audience. Finally, it presents narrative models for VR, from linear to open and emergent structures, and details core narrative tools, including dialogues, choices, mechanics, visual design, and sound, which collectively enable coherent, meaningful, and engaging experiences. By integrating authorial intent and user interaction, this chapter provides a comprehensive framework for understanding and designing stories in VR.

NARRATIVE AND STORYTELLING: FROM TRADITION TO DIGITAL MEDIA

Since the dawn of humanity, storytelling has been one of the central pillars of different cultures around the world. Through storytelling, human beings have passed on knowledge, values, beliefs, and experiences, thereby shaping their individual and collective identity.

Narrative can be understood as a fundamental way of organizing human experience. In fact, Bruner (1991) defines it as the mechanism by which people organize, give meaning and significance to their experiences, transforming events into comprehensible and communicable stories. In this way, narrative not only fulfills a communicative function, but also a cognitive and cultural one.

Linked to this concept is narrative competence, understood as the ability to understand and produce narrative texts in which a sequence of actions unfolds in a specific time and space. These actions

are structured around a plot or storyline and feature specific characters. This competence involves both interpreting stories and the ability to construct them in a coherent and meaningful way (Orellana García et al., 2022).

With the rise of new technologies, new storytelling methods and formats have emerged, taking advantage of the possibilities offered by digital and virtual environments. Within this landscape, one of the most significant narrative approaches across media is Digital Storytelling (DST), or digital narrative (Hurtado-Mazeyra et al., 2023).

To fully understand this concept, it is first necessary to clarify what storytelling entails. In narrative discourse, it is not enough to transmit a message; it must be structured in a way that captures the audience's attention and creates an emotional connection. In this sense, storytelling functions as a key narrative technique.

Storytelling is the art of narrating stories through emotional language that helps the listener connect with the narrative, build trust, and facilitate the transmission of the message. Its application requires a narrative structure capable of generating conflict and intrigue, as well as expressive resources such as rhetoric and metaphor, which enrich the discourse and support message comprehension (Carpio Padilla, 2024).

DST, meanwhile, is a narrative tool that incorporates multimedia elements: audio, video, and digital images. This combination of expressive languages enhances the story and improves the recipient's experience, fostering more effective and meaningful communication (Smeda et al., 2014). In any of its forms (oral, written, or digital), all storytelling relies on an internal architecture that organizes its actions and gives meaning to the whole, known as narrative structure.

NARRATIVE STRUCTURE: PRESENTATION OF ACTIONS, MODES OF EXPRESSION, AND NARRATIVE ELEMENTS

The narrative structure acts as the organizing principle of the story, as it orders events, establishes relationships between them, and ensures their overall coherence. From this perspective, and following Fournier Marcos (2002), it is possible to distinguish between an external or general narrative structure and an internal or particular structure, complementary dimensions that intervene in the configuration of an effective story.

Types of Narrative Structures: External or General and Internal or Specific

The external or general narrative structure refers to the way in which the actions are presented within the story, that is, the visible organization of the story. Traditionally, this presentation is articulated through the classic division into three parts: exposition, climax, and denouement (Fournier Marcos, 2002).

The exposition introduces the theme and provides the necessary background to situate the reader or viewer in the context of the work. The rising action develops the main conflict and the events that lead to the climax, the moment of maximum narrative tension. Finally, the denouement shows the resolution of the conflict and the conclusion of the story.

Likewise, the external structure includes the forms of presentation of the actions, which may vary depending on the language used (whether simple, complex, everyday, or symbolic) and the types of text used, such as narration, description, or dialogue, among other expressive resources.

The internal or specific narrative structure focuses on the elements that make up the core content of the story, i.e., the where, when, and who of the story. According to Fournier Marcos (2002), these elements form the internal core of the narrative. They include:

- The plot, understood as the detailed development of the actions in the work, explains the causes, motives, and consequences of events. It can also be open or closed.
- The storyline, meanwhile, is a summary of the most important actions in the story and covers the main events in the narrative.
- Time, which refers to the duration and temporal organization of events.
- Space, understood as the physical or temporal environment in which actions take place.
- The narrator, who represents the point of view from which the story is told and who may be located inside or outside the story.
- Finally, characters are the protagonists of narrative events and can be classified according to their importance as main, secondary, environmental, or fleeting, as well as protagonists and antagonists depending on their role within the story.

The elements described above form the basis of the traditional narrative model and serve as a starting point for understanding its evolution in immersive contexts.

NARRATIVE ELEMENTS IN VIRTUAL REALITY

The incorporation of virtual reality (VR) into narrative processes transforms the traditional model in which the recipient is merely passive, giving the user an active role that shapes the story through their interaction with space, events, and narrative elements. From this perspective, Bucher (2017), drawing on an interview with Noah Nelson, outlines key principles of immersive narrative: VR does not invent a new narrative form but reinterprets it, and any significant change within the experience must capture attention and mark a new phase in immersion. The relationship between audience, actors, and space becomes central, requiring every narrative component to have a clear purpose.

The Term Agency and the Types of Interaction

One of the fundamental concepts in VR narrative is that of agency, closely linked to interactivity. This term refers to the user's ability to intervene, influence, or modify the virtual environment and the development of the narrative experience. Agency introduces a level of participation that alters the traditional narrative scheme and places the user as an active agent within the story.

Based on this degree of interaction, two main types can be distinguished:

- User-scene interaction refers to the level of control the user has over the narrative environment and their ability to influence the story. It depends on design decisions involving user experience (UX) and ergonomics, as well as the user's willingness and sense of presence. This interaction can range from manipulating objects to assuming a specific narrative role.
- User-technology interaction concerns the system's ability to transmit information through different senses and to let the user affect the virtual world. In VR, interaction and narrative potential grow as technology improves its capacity to process and represent information, enriching immersion.

This degree of agency is directly connected to the concept of ludonarrativism, applied mainly to video game design.

The Concept of Ludonarrativism and Agency

Ludonarrativism is an approach that emphasizes the harmonious integration between narrative (the story) and game mechanics (the gameplay). This concept analyzes how the player's actions and the narrative structure intertwine to create a coherent and meaningful experience.

One of the core principles of ludonarrativism is *ludonarrative harmony*, the alignment between game mechanics and narrative so both reinforce each other and create a coherent, immersive experience. When this alignment fails, *ludonarrative dissonance* emerges. As Grabarczyk and Kampmann (2022) explain, the term was originally introduced by Clint Hocking (2007) in his critique of *BioShock* to describe the conflict between its mechanics and narrative framing. Hocking argued that while the story promotes altruism, the gameplay rewards self-interested behavior, producing a tension between player actions and the values suggested by the narrative. This disconnect results in a rupture between mechanics and fiction, undermining the coherence of the experience.

The consistency between narrative and gameplay promotes greater player immersion, as players perceive that their decisions and actions have a real impact on the development of the story. In this way, interactivity takes on a meaningful character and reinforces the sense of agency and protagonism, fundamental aspects in interactive narratives and VR environments.

Likewise, this same capacity for intervention transforms traditional narrative elements, turning them into active components of the experience and giving them a central role in the construction of the VR story.

Ludonarrative Elements: Prosumer, Focalization, and Virtual Environment

In VR environments, narrative elements gain new relevance due to the medium's specific characteristics. Elements that were secondary in traditional audiovisual narratives, such as space, become central to the story, shifting from background to an active narrative component (Montoya, 2018).

Brarda (2016) defines narrative elements as expressive units that give audiovisual discourse meaning. These include sound, images, typography, and movement, which together shape the story within the narrative space, the place where visual, sound, and narrative components converge. From this perspective, video games function as ludofictional worlds (Planells, 2013), interactive narrative environments where game rules and story are closely intertwined.

The player's role becomes hybrid, as they not only receive the story but also act as both narrator and actor thanks to the interactive nature of the virtual environment (Suárez Mouriño, 2021). This active involvement makes the player a key agent in shaping the narrative, since their decisions influence its development. The design of space and mechanics reinforces this agency, consolidating the player's position as a co-author within ludonarrative environments.

By exploring virtual worlds, players can access a wealth of information related to current events, the history of the places they visit, and the relationships between their inhabitants. Therefore, the virtual environment plays a central role in the narrative process, functioning not only as a setting but also as an element that can assume the functions of a character within the story. Connel (2017) describes virtual worlds as narrative canvases where the details of the story are shaped through art and stage design.

Finally, sound is a key factor in capturing attention in VR and plays a crucial role in narrative construction (Bucher, 2017). Together with lighting, it guides the user toward specific narrative elements (Bala et al., 2016). Music shapes rhythm and emotional interpretation, while voice stands out for its communicative power (Rafols & Colomer, 2010). In VR, auditory perception conditions visual perception, turning image and sound into an aesthetic unit interpreted through the user's sensations and cultural context, thus reinforcing the narrative experience.

HOW TO DESIGN STORIES IN VR

Designing stories in VR requires a specific approach that takes into account the medium's unique characteristics, especially its immersive, interactive, and participatory capabilities. As mentioned earlier, unlike traditional audiovisual formats, VR places the user at the center of the narrative experience, making them an active agent within the fictional world. Therefore, narrative design must address both the author's creative decisions and the player's possibilities for action, as well as the technological, cultural, and expressive limitations of the virtual environment.

What Kind of Experience Do I Want to Design?

One of the first steps in designing a narrative experience in VR is to define the type of experience you want to create, as this decision determines the narrative structure, the degree of interactivity, and the role of the user. In general terms, VR experiences can be classified into multimedia content, applications, and video games.

- Multimedia content tends to focus on the visualization of environments or stories with a limited level of interaction, where the narrative is mostly predefined.
- VR applications can have educational, cultural, or informational objectives, combining narrative elements with specific tasks or guided tours.
- VR video games represent the highest degree of narrative and interactive complexity, as they integrate game mechanics, rule systems, and fictional worlds in which the user actively participates.

Choosing the type of experience involves reflecting on the degree of player agency, the level of freedom they will have to explore the world, and the way in which the story will unfold during the experience.

What Do I Want to Convey? Context, References, and Intertextuality

Once the type of experience has been defined, it is essential to determine what you want to tell and from what cultural and narrative framework the story is constructed. In this sense, the narrative does not arise in isolation but is deeply influenced by the sociocultural context in which it is produced.

The cultural components that influence narrative are known as facts and are classified as mindfacts, sociofacts, and artifacts (Fantini & Fantini, 1997):

- Mindfacts refer to the ideas, beliefs, values, or knowledge shared by a society, such as religion, myths, or ideologies.
- Sociofacts regulate social interaction through norms, customs, or institutions, such as gender roles or the education system.
- Artifacts, meanwhile, constitute the material manifestation of a culture, including tools, objects, architecture, or, in the case of VR, the application or video game itself as a cultural product.

These elements directly shape the narrative, the world it constructs, and the way the user interprets it. In this context, intertextuality, understood as the dialogue between a work and previous ones, becomes especially relevant. Designers must consider their references and the type of dialogue they establish with

them. When a work incorporates recognizable elements from earlier stories, such as *Pokémon Scarlet and Violet*, inspired by the Iberian Peninsula and its cultural features (Game Freak, 2022), this recognition enhances the player's immersion.

This intertextual dialogue, understood as the way a work references previous stories or cultural elements, creates a pact with the user based on shared expectations. The more culturally and narratively literate the player, the better they can interpret and enjoy the experience. In this way, the narrative connects to the user's identity and shapes expectations that may be fulfilled or subverted throughout the game.

User Perspective in VR Narrative

When constructing an audiovisual story, and particularly in VR environments, it is essential to decide how narrative information is presented and from what point of view it is experienced by the user. This issue is articulated through narratological concepts such as the theory of focalization, ocularization, and auricularization, which allow us to analyze the degree of narrative knowledge and the sensory perspective from which the story is constructed.

Genette (1989) systematized the theory of focalization, distinguishing three types: zero, external, and internal.

- Zero focalization appears when the narrator knows more than the characters, following the classic omniscient model.
- External focalization limits narration to observable actions, without access to characters' inner worlds, typical of objectivist approaches.
- Internal focalization filters the story through a character's perspective, granting access only to their thoughts. Genette (1989) further differentiates fixed, variable, and multiple internal focalization depending on whether the perspective stays with one character, shifts between several, or recounts the same event from different viewpoints.

In VR environments, internal focalization becomes central because the user typically occupies a subjective position within the narrative world, acting as a witness, protagonist, or guiding observer. Experiences such as *Anne Frank House VR* (Force Field & Vertigo Games, 2018) exemplify this by allowing users to explore the story from their own perspective. Others, like *Half-Life: Alyx* (Valve, 2020), place the user directly in the protagonist's role, integrating narrative and mechanics to heighten immersion. Games such as *Moss* (Polyarc, 2018) adopt a hybrid approach, where the user guides the protagonist while remaining as an external observer, blending internal and external focalization. These examples show how VR constructs focalization through spatial and interactive design, enabling engagement beyond linear storytelling.

Ocularization defines the visual point of view, distinguishing between zero ocularization, an external, objective view, and internal ocularization, where the perspective aligns with a character's gaze. VR predominantly uses internal ocularization, placing the user inside the narrative space and reinforcing presence. Auricularization similarly shapes access to sound, guiding attention and meaning within the virtual environment.

Together, the articulation between focusing, ocularization, and auricularization allows for the design of narrative experiences consistent with the expressive possibilities of the medium, promoting user immersion and active involvement in the construction of the story.

The player's role becomes hybrid, as they not only receive the story but also act as both narrator and actor thanks to the interactive nature of the virtual environment. This active involvement makes the player a key agent in shaping the narrative, since their decisions influence its development. The design of space and mechanics reinforces this agency, consolidating the player's position as a co-author within ludonarrative environments (Suárez Mourinho, 2021).

Narrative Models in DST Applicable to VR

To ensure coherence and structure in a VR experience, events should be designed in accordance with specific narrative models. Although interactive media expand storytelling possibilities, they still draw on the classical structure of beginning, middle, and end.

From a structural perspective, two main models can be distinguished: linear and non-linear. The linear model follows the traditional exposition–conflict–resolution scheme, concentrating emotional impact in the final outcome (McKee, 2002). In contrast, non-linear structures allow greater temporal flexibility and player participation, organizing sequences in ways that do not alter the overall meaning of the story (Shaw, 2019). As Landow (1992) notes, interactivity enables multiple narrative paths to emerge from the same narrative universe.

In the field of video games, narrative structures often draw on classical dramaturgy to organize conflict, development, and resolution. Based on this foundation, several models describe how interactive stories unfold, including strictly linear narratives, the “pearl necklace” model (where players advance through fixed narrative nodes), branching structures with different degrees of agency, and hybrid designs such as funnel or critical-path configurations that combine freedom with converging milestones.

Within non-linear approaches, multi-level storytelling (common in Massively Multiplayer Online games) offers fragmented access to lore, while the “amusement park” model organizes narrative spatially rather than chronologically. Open systems such as *Animal Crossing: New Horizons* (Nintendo EPD, 2020) allow players to construct personal stories through interaction with the world.

Beyond these structures lies emergent narrative, which does not rely on a fully scripted sequence but arises dynamically from the interaction between game systems and player actions (Gómez-García et al., 2021). This framework of possibilities is especially relevant in VR, where immersion and embodied interaction reinforce the player’s role as co-creator of the narrative experience.

Key Aspects for Designing a Narrative Experience in VR

Narrative design in VR must consider a series of fundamental aspects that allow for the creation of a coherent, immersive, and meaningful experience. Authors such as García Moreno (2021) highlight the importance of balancing the author’s narrative, the player’s experience, and the expressive potential of the virtual environment.

From this perspective, the first step is to define the author’s story: the narrative framework that establishes the fictional universe, its conflict, and overall structure. This includes deciding how the story is presented (through missions, environmental storytelling, cinematics, or text) and whether its progression is fixed or flexible. Even in highly interactive experiences, this authorial layer provides the coherence that guides the entire narrative.

It is also essential to consider the player’s story, understood as the actions, decisions, and interactions they perform in the virtual world. In VR, embodied interaction heightens presence and allows players to deviate from the scripted path, creating personal trajectories. Anticipating the consequences of these actions and how they reshape the world is key to maintaining agency and narrative consistency.

The relationship between these two dimensions directly affects the definition of the protagonist. The playable character may operate merely as an avatar, i.e., an interface for action, or as a fully developed dramatic subject with motivations, needs, and evolution throughout the narrative (Cuadrado, 2013). This decision shapes the player’s identification with the story and conditions the depth of narrative engagement.

Beyond character construction, narrative coherence in VR also depends on the configuration of the world as a ludonarrative space (Planells, 2013). The virtual environment communicates meaning through spatial design, environmental storytelling, and embedded lore (Connel, 2017; Nitsche, 2008). Establishing whether the world possesses a past, a historical framework, or a system of internal rules becomes especially relevant in open or exploratory structures.

Closely linked to world-building is the management of temporality. Narrative time may mirror real time or follow the internal logic of the fictional system, for example, through cyclical days, accelerated rhythms, or time progression tied to player actions (Pérez, 2010). Temporal design influences pacing, immersion, and the perception of consequence.

Equally, the design of the virtual environment itself plays a decisive role. Space is fundamental for generating presence (Lombard & Ditton, 1997) and must be convincing, coherent, and meaningful. Including a population in the virtual world is essential for generating familiarity and enhancing player immersion (Szymanczyk et al., 2011), as characters reinforce realism and reduce the perception of space as purely fictional (Černý et al., 2017). Furthermore, consistent design and animation of the world's inhabitants allow for more complex situations that simulate life in the virtual environment (Kershner, 2018).

In character design, it is important to consider the uncanny valley, which warns of possible rejection toward overly realistic anthropomorphic figures (Araujo & Lascrain, 2022; Maldonado Nolivos, 2022). Beyond realism, characters function as narrative tools shaped not only by dialogue or appearance but also by animation, since even their movements convey meaning. Color likewise operates as an expressive resource that creates symbolic contrasts and visually distinguishes narrative elements (Lambert, 2023).

Maintaining ludonarrative coherence is essential. Ludonarrative dissonance occurs when the actions allowed by the system contradict the narrative meaning the experience aims to convey (Hocking, 2007). In VR, where embodied interaction intensifies presence, these inconsistencies become even more noticeable. Aligning narrative, mechanics, and player agency ensures that the story is expressed through interaction, reinforcing immersion and emotional engagement.

Narrative Tools in VR

To structure narrative information in a VR experience, designers rely on various tools that modulate user interaction and ensure coherence. These resources shape not only how the story is perceived, but also immersion, agency, and the relationship between mechanics and narrative. Among the main ones are:

- Non-interactive dialogues, verbal or written messages the user cannot influence. They provide context, emotions, or instructions without altering the narrative flow, for example, a character explaining historical events in a VR scenario.
- Interactive dialogues, which let players choose responses that affect the conversation, the story, or relationships. They reinforce agency and are common in narrative and role-based VR experiences.
- Decisions, choices that influence the story or mechanics. Whether binary, multiple, or open-ended, they determine if the narrative is linear, branching, or emergent, personalizing the experience and shaping future interactions.
- Mechanics, the rules governing how players interact with the environment. Movement, object manipulation, or specific actions can convey meaning or reinforce themes; for instance, exploration systems that reveal story clues non-linearly.
- Art, encompassing all visual elements (characters, environments, lighting, effects). Beyond aesthetics, it communicates emotions and context; environmental details can tell parts of the story without words.
- Music and sound design fulfill emotional and functional roles. Music guides tension and attention, effects contextualize actions, voices convey story, and spatial audio enhances immersion, creating a unified sensory narrative.

CONCLUDING REMARKS

In conclusion, VR narrative blends traditional storytelling with the medium's interactive and spatial affordances. Classic elements such as characters, time, space, and plot are reconfigured so the user becomes an active participant with a subjective role in the story. The integration of focalization, ocularization, auricularization, and narrative tools (dialogue, decisions, mechanics, art, and sound) supports immersive and coherent experiences in which authors and users construct the narrative together (Figure 9.1).

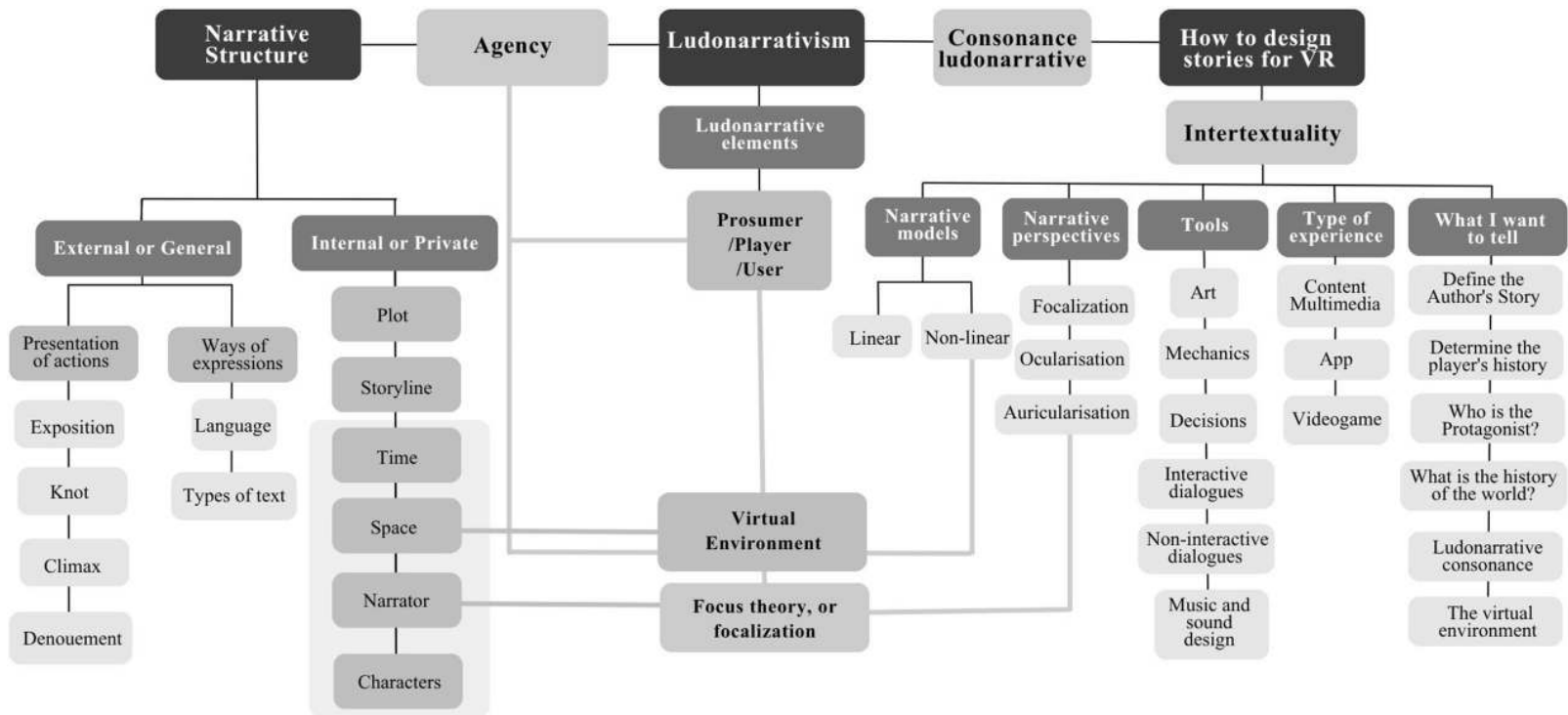


FIGURE 9.1 Conceptual map of narrative and storytelling in VR.

REFERENCES

- Araujo, D., & Lascurain, I. (2022). Robótica social, la ciencia detrás de la interacción humana. *+Ciencia* (28), 19–21.
- Bala, P., Dionisio, M., Nisi, V., & Nunes, N. (2016). IVRUX: A tool for analysing immersive narratives in virtual reality. In F. Nack & A. S. Gordon (Eds.), *Interactive storytelling* (pp. 3–11). Springer. https://doi.org/10.1007/978-3-319-48279-8_1.
- Brarda, M. (2016). *Motion graphics design, La dirección creativa en branding de TV*. Editorial Gustavo Gili.
- Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, 18(1), 1–21. <https://doi.org/10.1086/448619>.
- Bucher, J. (2017). *Storytelling for virtual reality: Methods and principles for crafting immersive narratives*. Taylor & Francis.
- Carpio Padilla, P. I. (2024). Análisis del discurso narrativo y estético de una pieza de motion graphics desde sus componentes. *Ciencia Latina Revista Científica Multidisciplinar*, 8(1), 10038–10054. https://doi.org/10.37811/cl_rcm.v8i2.10321.
- Černý, M., Plch, T., Marko, M., Gemrot, J., Ondráček, P., & Brom, C. (2017). Using behavior objects to manage complexity in virtual worlds. *IEEE Transactions on Computational Intelligence and AI in Games*, 9(2), 166–180. <https://doi.org/10.1109/TCIAIG.2016.2528499>.
- Connel, J. (2017). Ghost of Tsushima interview: Details on Sucker Punch's next open world adventure [Video]. *YouTube*. <https://www.youtube.com/watch?v=HftJIAidMo>.
- Cuadrado, A. (2013). Acciones y emoción: Un estudio de la jugabilidad en heavy rain. In J. Sánchez-Navarro, A. Planells de la Maza, & D. Aranda Juárez (Eds.), *Homo videoludens 2.0. De pacman a la gamification* (pp. 151–176). Laboratori de Mitjans Interactius, Universitat de Barcelona.
- Fantini, B., & Fantini, A. (1997). Artifacts, sociofacts, and mentifacts: A sociocultural framework. In A. Fantini (Ed.), *New ways in teaching culture*. TESOL Publications.
- Force Field & Vertigo Games. (2018). *Anne Frank House VR* [Aplicación de realidad virtual]. Force Field.
- Fournier Marcos, C. (2002). *Análisis literario*. Cengage Learning Latin America.
- Game Freak. (2022). *Pokémon scarlet and violet* [Nintendo Switch]. Nintendo.
- García Moreno, R. (2021). Videojuegos de mundo abierto: Propuesta de un modelo de análisis ludonarrativo. *Index. comunicación*, 11(2), 109–132. <https://doi.org/10.33732/ixc/11/02>.
- Genette, G. (1989). *Figuras III*. Lumen.
- Gómez-García, S., Paz-Rebollo, M. A., & Cabeza-San-Deogracias, J. (2021). Newsgames against hate speech in the refugee crisis [Newsgames frente a los discursos del odio en la crisis de los refugiados]. *Comunicar*, 67, 123–133. <https://doi.org/10.3916/C67-2021-10>.
- Grabarczyk, P., & Kampmann, B. (2022). A game of twisted shouting: Ludo-narrative dissonance revisited. *Eludamos: Journal for Computer Game Culture*, 13(1), 7–27. <https://doi.org/10.7557/23.6506>.
- Hocking, C. (2007, October 7). Ludonarrative dissonance in BioShock: The problem of what the game is about. *Click Nothing*. https://clicknothing.typepad.com/click_nothing/2007/10/ludonarrative-d.html.
- Hurtado-Mazeyra, A., Núñez-Pacheco, R., & Alejandro-Oviedo, O. (2023). El digital storytelling como herramienta y estrategia educativa en versión 2D y 3D para el desarrollo de la competencia narrativa en la educación infantil. *Pixel-Bit. Revista de Medios y Educación*, 67, 186–216. <https://doi.org/10.12795/pixelbit.97769>.
- Kershner, C. J. (2018). The lives of others: How NPCs can increase player empathy. In [Conference Presentation]. *Game Developers Conference*, San Francisco, California, United States. <https://tinyurl.com/t2dyrtud>.
- Lambert, Y. (2023). Las narrativas audiovisuales en el diseño de videojuegos: Aproximaciones teóricas, un acercamiento a sus particularidades. *Nawi: Arte Diseño Comunicación*, 7(1), 231–258.
- Landow, G. P. (1992). *Hypertext 2.0: The convergence of contemporary critical theory and technology*. Johns Hopkins University Press.
- Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2). <https://doi.org/10.1111/j.1083-6101.1997.tb00072.x>.
- Maldonado Nolivos, V. D. (2022). El valle inquietante, su ejecución y cómo evitarlo en animaciones 3D. *Revista Odigos*, 3(3), 9–26. <https://doi.org/10.35290/ro.v3n3.2022.675>.
- McKee, R. (2002). *El guion: Sustancia, estructura, estilo y principios de la escritura de guiones*. Alba.
- Montoya, M. I. V. (2018). Narrativas inmersivas para comunicadores. Realidad virtual, aumentada y mixta en propuestas audiovisuales de ficción y no ficción. *Comunicación*, 39, 7–12.
- Nintendo EPD. (2020). *animal crossing: New horizons* [Video game]. Nintendo.
- Nitsche, M. (2008). *Video game spaces: Image, play, and structure in 3D game worlds*. MIT Press.

- Orellana García, P., Valenzuela Hasenohr, M. F., Pezoa Sandoval, J., & Villalón Bravo, M. (2022). Competencia narrativa: Evidencias de validez para la estandarización de su evaluación en niños chilenos. *Psykhé*, 31(2). <https://doi.org/10.7764/psykhe.2020.22067>.
- Pérez, Ó. (2010). *Análisis de la significación del videojuego: Fundamentos teóricos del juego, el mundo narrativo y la enunciación interactiva como perspectivas de estudio del discurso* [Doctoral dissertation, Universitat Pompeu Fabra].
- Planells, A. (2013). *Los Videojuegos Como Mundos Ludoficciones: Una Aproximación Semántico-Pragmática a su Estructura y Significación*. Laboratori de Mitjans Interactius, Universitat de Barcelona.
- Polyarc. (2018). *Moss* [Video game]. PlayStation VR. Sony Interactive Entertainment.
- Rafols, R., & Colomer, A. (2010). *Diseño Audiovisual*. GG Diseño.
- Shaw, A. (2019). *Design for motion: Fundamentals and techniques of motion design*. Focal Press.
- Smeda, N., Dakich, E., & Sharda, N. (2014). The effectiveness of digital storytelling in the classroom: A comprehensive study. *Smart Learning Environments*, 1, 6. <https://doi.org/10.1186/s40561-014-0006-3>.
- Suárez Mourinho, A. (2021). El usuario interactivo y su rol como atacante y narrador en el consumo de narrativas ludoficciones con elementos transmedia. *Virtualis*, 12(22), 1–17.
- Szymanczyk, O., Dickinson, P., & Duckett, T. (2011). From individual characters to large crowds: Augmenting the believability of open-world games through exploring social emotion in pedestrian groups. In *Proceedings of DiGRA 2011: Think Design Play*, Hilversum, The Netherlands. <https://tinyurl.com/2djsc2s2>.
- Valve. (2020). *Half-life: Alyx* [Video game]. PC VR.

Manuel Caeiro

ABSTRACT

This chapter examines Social Virtual Reality (social VR) as an emerging medium that integrates immersive virtual reality (VR) technologies with real-time social interaction. It introduces the core components of social VR, including avatars, presence, immersion, rich multimodal communication, and tracking technologies, that enable embodied and socially meaningful experiences. This chapter reviews how these elements foster a strong sense of co-presence and interaction that goes beyond traditional text and video-based communication systems. This chapter then explores key application domains of social VR, including virtual meetings and collaboration, education, healthcare, virtual events and performances, collaborative co-design, gaming and entertainment, and socialization and networking. Through representative platforms and illustrative examples, it highlights how social VR is used to support learning, professional collaboration, creative practices, and social experiences at scale. Finally, this chapter addresses the main challenges faced by the adoption of social VR, addressing technical limitations, interoperability issues, user security and moderation, data privacy and security, as well as ethical and psychological concerns related to prolonged immersion and identity dissociation in avatars. Taken together, these perspectives offer a balanced view of the opportunities and constraints shaping the future of social VR.

INTRODUCTION

Social Virtual Reality (social VR) represents a significant evolution in immersive technologies by combining the possibilities of VR with the social dynamics of networked digital environments. Unlike the VR applications described in previous chapters, which focus primarily on individual experiences, social VR emphasizes shared presence, real-time interaction, and collective participation within persistent or semi-persistent virtual spaces. Through the use of avatars, spatialized audio, embedded interaction, and shared environments, social VR blurs the boundaries between physical and virtual social spaces, enabling forms of interaction that more closely resemble face-to-face encounters than conventional text- or video-based communication systems.

Over the past decade, advances in head-mounted displays, motion tracking, network infrastructure, and real-time rendering have facilitated the rapid emergence of social VR platforms geared toward socialization, collaboration, education, healthcare, entertainment, and creative co-production. These platforms allow geographically dispersed users to meet, communicate, collaborate, and perform activities together in immersive, three-dimensional environments. As a result, social VR has gained increasing attention from researchers, designers, educators, healthcare professionals, and industry professionals as a means capable of sustaining enriching social experiences beyond the limitations of existing digital communication systems.

This chapter offers a comprehensive overview of social VR from both a conceptual and applied perspective. It begins by introducing the fundamental components that define social VR experiences, including avatars as embodied representations of users, different dimensions of presence (spatial, social, and self-presence), levels of immersion, and enriched multimodal communication. Special attention is paid to the role of nonverbal interaction, such as gestures, gaze, and body movement, as well as tracking and interaction technologies that enable real-time social interaction in virtual environments.

FOUNDATIONS OF SOCIAL VR

Social VR platforms allow users to represent themselves through avatars and interact with others in virtual environments, providing a sense of social presence and interaction that goes beyond traditional text or video communication. To understand social VR in more depth, it is important to disaggregate its core components, such as avatars, presence, immersion, enriched communication, and the different types of tracking and interaction that create an engaging social experience.

Avatars: Digital Representation of Users

Users in social VR environments are represented by avatars, which are customizable virtual characters. An avatar is a virtual representation of a user. This depiction goes beyond appearance, which can be very realistic or highly imaginative, with cartoon characters, animals, or objects. In addition, avatars can mimic real movements and expressions, increasing the sense of realism in interactions.

The use of avatars is essential in social VR, as users interact with each other through them. Avatars mirror users' real-time positions and reproduce their movements through gesture tracking. Some researchers suggest that, when interacting with others through virtual avatars, social interactions in VR are perceived as more natural and closer than interactions that occur in other digital systems.

The design and functionality of avatars play an important role in shaping the social dynamics of VR spaces, and this can be achieved in a number of ways through personalization, identity expression, and support for highly detailed and realistic avatars that mimic human proportions and appearances.

Presence: The Feeling of “Being There”

Presence is defined as the feeling of “being there”, physically located within the virtual environment (Slater & Sanchez-Vives, 2016). This concept implies a perceptual illusion of non-mediation that causes the user to perceive, and usually also react, to the VR environment as if it were real, even when they know it is not.

A distinction is usually made between several types of presence (Barreda-Ángeles et al., 2023):

- *Spatial presence* refers to the feeling of being physically in a virtual environment.
- *Social presence* refers to the perception of being in the same space as others and being connected to them.
- *Embodiment (or presence of the self)* is described as the feeling that the avatar's virtual body is his or her own body.

Another key aspect of social VR is the *shared virtual space*. This concept of being able to occupy a common environment with others, whether it is a virtual cafeteria, a concert venue, or an art gallery, reinforces the illusion of being in a real environment. The ability to interact with others within these shared environments helps maintain a sense of presence, even if users are geographically distant.

Immersion: The Depth of Engagement

Immersion is the level of sensory and emotional engagement experienced by a user within a virtual environment. Social VR places users in a fully immersive 3D environment where they have a high level of autonomy and can interact with other users and available objects. The more immersive a VR experience is, the more it can fully capture the user's attention and make them feel like they are "inside" the virtual world, rather than just observing it.

Social VR can involve different types of immersion:

- **Visual and Auditory Immersion:** Immersion begins with the sensory aspects of VR. High-definition graphics, 3D soundscapes, and spatial audio technologies contribute to a realistic virtual environment. For example, if two people are speaking in a VR space, the volume and direction of their voices will change depending on their location in the virtual world. This auditory cue helps to create the illusion of proximity, adding realism and immersion to the experience.
- **Tactile Immersion:** While visual and auditory immersion are essential, haptic feedback and motion tracking are playing an increasing role in VR immersion. Advanced controls and accessories such as gloves, vests, and full-body suits provide users with tactile feedback that mimics real-world sensations, such as the sensation of holding an object or the vibration of distant sound.
- **Environmental Immersion:** The design of the virtual environment itself also contributes to immersion. High-fidelity storytelling, whether replicating real-life situations or creating completely fictional landscapes, helps create the illusion that users are present in a space, allowing for deeper interaction.

Enriched Communication: Beyond Text and Voice

One of the main objectives of social VR is to improve communication by offering more expressive and interactive forms of communication than traditional media such as text, voice, or video calls. Social VR offers a variety of forms of communication, many of which are designed to make virtual interactions more nuanced and dynamic. The following communication modalities can be recognized with social VR environments:

- **Nonverbal Communication:** In the real world, people rely on body language, facial expressions, and gestures to communicate. Social VR attempts to replicate this in the digital world through avatar animation. Features such as eye tracking, lip sync (which matches the avatar's mouth movements to the user's speech), and hand gestures allow for more natural interactions. Some platforms support a wide range of expressive gestures such as greeting, pointing, or dancing, which can be performed in real time.
- **Voice Interaction:** Voice communication in social VR is spatialized, meaning that the volume of a person's voice depends on their proximity to the speaker. This creates a more immersive and realistic communication experience. For example, a user speaking in the virtual environment will sound louder when they are close to another avatar and softer when they are farther away, simulating how sound works in the real world.
- **Multimodal Interaction:** Some social VR platforms also allow users to communicate through a combination of voice, gestures, and object manipulation. This multimodal approach enriches conversations by adding more layers of interaction, allowing users to interact with objects in the environment (e.g., offering someone a virtual drink) while talking.

Tracking and Interaction: Enabling Real-Time Social Interaction

In social VR, interaction tends to be more holistic and involves body movements. The ability to track a user’s movements and actions in real time is essential for creating an interactive and responsive social experience in VR. This tracking technology allows the user’s physical actions to be reflected by their avatar in the virtual world, increasing the sense of presence and social interaction. Several different types of tracking can be recognized on social VR platforms: head tracking, hand and body tracking, gesture recognition, or interaction with objects.

USE CASES OF SOCIAL VR

Social VR is being applied to various fields in which social interaction plays a relevant role. Educational, medical, collaborative co-creation scenarios, or virtual events and conferences in which several people must interact together are the most common.

Table 10.1 provides an overview of the most representative social VR platforms and their distinctive features and dominant use cases, which are further described in the following subsections.

Virtual Meetings: Remote Communication and Collaboration

Social VR is increasingly being adopted as a platform for professional collaboration. Platforms like Spatial and vSpatial allow individuals and teams to hold meetings in immersive virtual offices, breaking down geographic barriers. The ability to share virtual tablets, 3D objects, and real-time visualizations makes social VR an attractive option for remote work, education, and brainstorming sessions.

The improvement of communication and remote collaboration is one of the functionalities that generates the most interest in terms of social VR (Li & Cesar, 2023). It is about allowing a more enriching interaction, with a clearer social presence and forms of nonverbal communication, which can bring the experience of users closer to what they have in face-to-face situations, thus eliminating the need for

TABLE 10.1 Representative social VR platforms and key use cases

PLATFORM	PRIMARY USE CASES	DISTINCTIVE FEATURES
VRChat	Socialization and networking; gaming and entertainment; virtual events	Highly customizable avatars; strong embodiment and social presence; large-scale social worlds; HMD and desktop access
Meta Horizon Worlds	Socialization and networking; virtual meetings; virtual events	Stylized avatars; real-time gesture and voice interaction; HMD-based immersive social spaces
Hubs	Education; virtual meetings; socialization	Web-based access; supports HMD and non-HMD users; small-to-medium group collaboration
ENGAGE VR	Education	Virtual classrooms; experiential learning; immersive simulations for higher education and training
Spatial	Virtual meetings; co-design	Immersive virtual offices; shared 3D objects and whiteboards; collaboration-focused
CakeVR	Co-design	Avatar-based collaborative design; real-time co-creation between customers and professionals
Rec Room	Gaming and entertainment; socialization	Multiplayer mini-games; cooperative and competitive play; strong social layer

physical travel and reducing travel time and environmental impact. However, the objective of social VR is not to replicate reality, but to facilitate and expand the available communication channels.

Currently, there are two main types of systems available for remote communication and collaboration: video conferencing platforms, such as Zoom or Teams, commonly used in desktop computers and mobile devices, and more advanced systems, such as HP Halo or Cisco telepresence, which allow you to connect physically separated rooms using large screens with high-quality audio and video. Despite the capabilities of these systems, the user experience is not completely satisfactory, as their possibilities for interaction and collaboration are much more limited than those that occur in face-to-face meetings (Brubaker et al., 2012). All of these video conferencing tools limit users in front of screens with “talking head experiences” and limited physical interaction capabilities, with limited opportunities for embodied interaction compared to face-to-face meetings.

Social VR can enhance the social interaction available in video conferencing environments in a number of ways. Specifically, it can provide a more immersive experience and allow users to interact with shared virtual objects (Pidel & Ackermann, 2020). Several of the commercial social VR platforms (VRChat, Meta Horizon Worlds) implement different types of mechanisms and tools to stimulate social activities, such as creating a group discussion environment, implementing tools that allow users to focus on social tasks, etc.

Education

Social VR is also being explored for educational purposes, creating immersive environments for learning. The engaging and immersive nature of VR makes it ideal for experiential learning, simulations, and interactive lessons on topics such as history, science, and engineering.

Educational activities carried out with VR are generally considered more effective than other non-immersive interventions. The review conducted by Jensen & Konradsen (2017) shows that the use of VR equipment can provide advantages in terms of spatial and virtual information acquisition, training in psychomotor and observational skills, and development of affective self-regulation skills.

In higher education contexts, researchers have mainly focused on the cutting-edge that VR offers in terms of academic achievement and increased engagement (Nesenbergs et al., 2020). According to the results, VR can offer realistic interactive simulations of environments and objects that are not physically available or that would be impossible to use in reality. In this way, students can carry out internships that would not otherwise be possible. In addition, more engaging content can also be generated that helps students better understand complex concepts. On the other hand, VR can also improve student engagement by using content that they find more fun, motivating, and that generates positive emotions.

In general, there are multiple opportunities that VR environments can offer in education, including those related to social interaction (Barreda-Ángeles et al., 2023). First of all, it is important to recognize that most educational activities take place in social settings. Students do not learn in isolation and independently in most cases, but do so in scenarios that they share with other peers and in which a teacher or instructor also usually participates. Social VR can introduce positive effects in terms of improving social presence, communication between students, and feelings of connection and community. In addition, social VR technology shows great potential in terms of enhancing the social aspects of distance education, allowing the feeling of social presence to be realized. Holt et al. (2020) describe a study in which ecology students used the social VR platform Mozilla Hub to present their research work to their peers in an interactive session, using avatars to interact with them through HMDs. The students noted that the shared (virtual) space facilitated interactions with their peers. Another similar work is that of Gomes de Siqueira et al. (2021), in which VR spaces from Mozilla Hubs were also used, although in this case they did not use HMDs. However, there are still few studies that investigate the benefits and problems that the use of fully immersive social environments in VR can involve (Barreda-Ángeles et al., 2023).

Healthcare

Social VR systems have also been proposed as a way to reduce patient visits to medical centers. Examples of situations that have been explored are: when a patient needs a consultation to communicate a problem, when treatment should be performed, or when a patient needs to be informed about the preparation of surgery. In all cases, patients and medical staff can interact through social VR environments and therefore reduce the time spent on physical commuting.

The use of VR in medicine has been around for many years. One of the pioneering applications was launched in the 1990s, the purpose of which was to facilitate the visualization of complex medical data in a way that would facilitate the preparation of surgeries for medical teams (Satava, 1994). Currently there are many examples of applications of VR in medicine for the development of medical training (Riva, 2014), psychological consultation (Riva et al., 2015), remote therapy, (psycho)therapy (Bee et al., 2008), etc.

A concrete example carried out at the Delft University of Technology, in particular, is a social VR clinic for patients with knee and hip arthritis. According to the authors of this study, this type of solution allows patients with limited physical mobility to reduce the number of hospital visits, but maintain good communication with medical staff. The social VR clinic simulates a real consultation room in which patients can interact with doctors and nurses by visualizing information, such as procedures for the preparation of surgeries, 3D anatomical models, or take a virtual guided tour of the clinical facilities.

Virtual Events and Performances

Increasingly, concerts, conferences, art exhibitions, and other live events are held in social VR spaces. VR platforms allow users to attend these events in a virtual setting, often with the ability to interact with other attendees and participate in activities in real time.

In Li and Cesar (2023) an experience with VR is presented that allows a virtual visit to a museum. Currently, museums are exploring new ways to engage for their visitors, allowing them to do so remotely. The MediaScape XR application aims to offer an alternative to passive visits to museums, turning them into immersive experiences in which visitors can act remotely with the help of photorealistic representations. This application integrates a 3D replica of an object from the Dutch Institute for Sound and Vision (NISV) into an immersive virtual museum. Visitors wear HMDs to enter the virtual space and interact with the virtual object, being able to put on a virtual disguise.

In addition to museums, social VR is also explored in other entertainment activities, such as watching a movie (Li & Cesar, 2023). The aim is to reproduce and, to some extent, improve the experience of watching a movie together. When people are apart, they can watch the same movie and communicate via text message or video call, and even share their screens or use video streaming sync apps (e.g., Telparty). This differs substantially from the experience of sharing the same space. In this scenario, social VR also aims to offer an experience closer to the real situation, allowing social interactions to be closer to face-to-face and enabling participants to interact with each other and with available objects. The virtual space can be a generated scene or a real 360° scenario captured with an omnidirectional camera. Each user can be represented by an avatar, or in recent solutions, as a virtual representation of a user based on camera capture.

Co-Design

In CakeVR, a 3D virtual scenario is presented in which customers and pastry chefs represented through avatars can collaborate to create a custom cake (Li & Cesar, 2023). This is a co-design activity in which several participants interact together on the same stage.

Preparing a custom cake is not an easy task. In real life, several meetings are usually required in which customers express their preferences and ideas to the pastry chefs through images and general descriptions, and the pastry chefs prepare designs that they display through graphics and templates to the customers.

Games and Entertainment

One of the most common uses of VR is gaming. VR-based games such as Beat Saber and Population: One combine immersive environments with multiplayer social interactions, creating a more engaging experience than traditional 2D gaming. Social VR platforms also offer a variety of mini-games, from sports to party games, where users can compete or cooperate in virtual spaces.

Socialization and Networking

Social VR provides users with a digital space to socialize with friends and meet new people from all over the world. Platforms such as VRChat and Rec Room allow users to attend virtual parties, chat in informal lounges, or participate in themed events. These platforms are similar to online social networks, but with the added dimension of immersion and presence.

CHALLENGES OF SOCIAL VR

While social VR offers enormous potential for communication, entertainment, and collaboration, there are several major challenges to its widespread adoption and long-term viability. These challenges have technical, social, psychological, and ethical dimensions.

Technical Barriers

Despite its rapid growth, social VR remains limited by the availability of high-quality hardware. VR headsets, while becoming more affordable, still require a significant investment for an immersive experience. In addition, social VR platforms require powerful computing resources, reliable internet connections, and low latency to ensure smooth interactions.

The immersive experience that social VR promises requires sophisticated hardware, but current VR technologies are still evolving and have limitations that can affect both usability and accessibility:

- **Cost of VR Headsets:** High-quality VR headsets, which offer the best experience, can be expensive. While standalone headsets (such as the Meta Quest 3) have lower costs, higher-end systems (e.g., Valve Index or HTC Vive Pro) continue to have high prices. These costs may be prohibitive for some users and limit widespread adoption of social VR, especially in developing regions or for casual users.
- **Computational Power:** Many VR experiences require powerful computers or specialized consoles to run smoothly. To achieve a truly immersive experience with high graphics fidelity, VR headsets must be paired with high-end GPUs (Graphics Processing Units) and CPUs (Central Processing Units). This increases the cost of the technology and can create barriers for users who do not have access to sufficient computing resources.

- **Battery Life and Convenience:** Battery life remains a major limitation for standalone VR devices, as many headsets require recharging after a few hours of use. Comfort is also a challenge. VR headsets can be heavy and cause discomfort after prolonged use. This can make longer social interactions in VR (such as virtual meetings or events) physically exhausting, discouraging users from staying in virtual environments for long periods.
- **Latency and Bandwidth:** Social VR experiences rely heavily on low-latency interactions and high-bandwidth internet connections to ensure smooth, real-time communication between users. Latency can create a fragmented experience, with users experiencing delays in their actions or communication. This is especially problematic in multiplayer games or in live events, where real-time interaction is crucial. Areas with less developed internet infrastructures may struggle with the demands of VR.

In addition to hardware limitations, another problem that limits the development of social VR is the lack of standardization. Two main problems are recognized:

- **Fragmented Platforms:** There is a lack of interoperability between different social VR platforms. Users on one platform (e.g., VRChat) cannot easily interact with others on another (e.g., Meta Horizon Worlds), making it difficult to create a unified virtual social experience. If users want to access different experiences, they often need separate accounts, avatars, and virtual environments, leading to a fragmented social experience.
- **Cross-Platform Integration:** For social VR to be truly universal, you need to transport avatars, resources, and social connections across different platforms. This kind of integration between hardware and software ecosystems remains a challenge, as companies may not want to create open systems that allow for fluid interaction between competing platforms.

User Safety and Moderation

With the increasing presence of diverse groups of users in virtual environments, problems of harassment and inappropriate content are frequent. One of the biggest challenges in social VR is dealing with bullying, verbal abuse, and toxic behavior. In virtual spaces, users can feel a sense of anonymity and detachment, which leads some people to adopt inappropriate behaviors that they might not show in the real world. In addition, in some social VR platforms, issues related to unwanted sexual advances or inappropriate behavior are a growing concern. Since VR allows users to interact spatially and maintain proximity, there are unique challenges around physical boundaries and the consent of digital touch or proximity in virtual space.

Effective moderation systems and safety measures for users are essential to ensure that they can interact in a safe and respectful environment. Many platforms are implementing features such as avatar blocking, voice chat filters, and automated content moderation to address these challenges. In any case, social VR involves more complex forms of interaction, such as body language, voice communication, and spatial proximity. Detecting inappropriate behavior in these contexts is not as simple as monitoring inappropriate text-based content.

Privacy and Data Security

As with other online platforms, social VR raises concerns about user privacy and data security. VR systems and platforms collect a wide range of data, including location, behavior, and potentially biometric data (such as facial expressions and movement).

The following types of data include key personal information that must be properly managed:

- **User Behavior Tracking:** Social VR platforms typically track users' movements, voice data, and interactions within the virtual environment. This includes data such as gaze tracking, body movements, and even possibly facial expressions or biometric information. While this data is used to improve the immersive experience (e.g., avatar movement matching), it also poses a risk if mishandled or exploited.
- **Location and Biometric Data:** Some social VR platforms collect sensitive biometric data, such as facial expressions, heart rate, or even emotional responses, to enhance the realism of the interaction.

Ethical and Psychological Impacts

While social VR offers new ways to socialize and collaborate, it also presents potential risks to users' mental and emotional well-being:

- **Escapism:** One of the most significant psychological risks associated with social VR is the possibility of users disconnecting from reality. For some people, the immersive nature of VR can become a form of escapism, prioritizing virtual interactions over real-world relationships. This could exacerbate feelings of loneliness and isolation rather than alleviate them.
- **Addiction:** Excessive use of social VR could lead to addiction-like behaviors, in which users spend more time in virtual environments than interacting with people or activities in the physical world. This could lead to neglect of real responsibilities, such as work, school, or personal health, resulting in a deterioration in the balance between online and face-to-face life.
- **Social Anxiety and Stress:** For some users, interacting with others in a virtual world can increase the feeling of social anxiety or performance pressure, especially in crowded environments. The act of "acting" through an avatar can also exacerbate mental health disorders related to self-esteem or body image, as users may feel pressured to present idealized versions of themselves.
- **Depersonalization:** The use of avatars, while offering creative expression, can also cause a psychological disconnection from real identity. If a user immerses themselves excessively in their avatar, it can cause a dissonance between the digital and physical selves, contributing to feelings of depersonalization or identity confusion.
- **Identity and Self-Perception:** As VR becomes an integral part of daily life, the way people perceive themselves and others will continue to evolve. The ability to create and modify one's avatar, along with the immersive nature of VR interactions, raises questions about self-identity, body image, and virtual personalities.

CONCLUSION

Social VR represents a significant shift in the evolution of digitally mediated interaction by combining immersive technologies with embodied, real-time social presence. Unlike traditional communication systems, social VR enables users to share spatial environments, express identity through avatars, and engage in multimodal interaction that closely approximates face-to-face encounters while also extending beyond physical constraints. Nowadays, these applications span professional collaboration, education, healthcare, creative co-design, entertainment, and large-scale virtual events, illustrating its growing relevance across social and institutional contexts.

At the same time, the expansion of social VR raises complex technical, ethical, and psychological challenges. Hardware limitations, interoperability barriers, user safety concerns, data privacy risks, and the potential impact of prolonged immersion on identity and well-being are main concerns.

Looking ahead, the future of social VR will depend not only on technological advancement but also on thoughtful integration into everyday practices and institutional frameworks.

REFERENCES

- Barreda-Ángeles, M., Horneber, S., & Hartmann, T. (2023). Easily applicable social virtual reality and social presence in online higher education during the COVID-19 pandemic: A qualitative study. *Computers & Education: X Reality*, 2, 100024. <https://doi.org/10.1016/j.cexr.2023.100024>.
- Bee, P. E., Bower, P., Lovell, K., Gilbody, S., Richards, D., Gask, L., & Roach, P. (2008). Psychotherapy mediated by remote communication technologies: A meta-analytic review. *BMC Psychiatry*, 8, 1. <https://doi.org/10.1186/1471-244x-8-60>.
- Brubaker, J. R., Venolia, G., & Tang, J. C. (2012). Focusing on shared experiences: Moving beyond the camera in video communication. In *Proceedings of the Designing Interactive Systems Conference*, Newcastle Upon Tyne, United Kingdom (pp. 96–105). ACM. <https://doi.org/10.1145/2317956.2317973>.
- Gomes de Siqueira, A., Feijóo-García, P. G., Stuart, J., & Lok, B. (2021). Toward facilitating team formation and communication through avatar-based interaction in desktop-based immersive virtual environments. *Frontiers in Virtual Reality*, 2, 647801. <https://doi.org/10.3389/frvir.2021.647801>.
- Holt, E. A., Heim, A. B., Tessens, E., & Walker, R. (2020). Thanks for inviting me to the party: Virtual poster sessions as a way to connect in a time of disconnection. *Ecology and Evolution*, 10(22), 12423–12430. <https://doi.org/10.1002/ece3.6756>.
- Jensen, L., & Konradsen, F. (2017). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4), 1515–1529. <https://doi.org/10.1007/s10639-017-9676-0>.
- Li, J., & Cesar, P. (2023). Social virtual reality (VR) applications and user experiences. In A. Smolic & P. Kauff (Eds.), *Immersive video technologies* (pp. 609–648). Academic Press. <https://doi.org/10.1016/B978-0-32-391755-1.00028-6>.
- Nesenbergs, K., Abolins, V., Ormanis, J., & Mednis, A. (2020). Use of augmented and virtual reality in remote higher education: A systematic umbrella review. *Education Sciences*, 11(1), 8. <https://doi.org/10.3390/educsci11010008>.
- Pidel, C., & Ackermann, P. (2020). Collaboration in virtual and augmented reality: A systematic overview. In L. T. De Paolis & P. Bourdot (Eds.), *Augmented reality, virtual reality, and computer graphics* (pp. 141–156). Springer. https://doi.org/10.1007/978-3-030-58465-8_10.
- Riva, G. (2014). Medical clinical uses of virtual worlds. In M. Grimshaw (Ed.), *The Oxford handbook of virtuality* (pp. 649–665). Oxford University Press.
- Riva, G., Dakanalis, A., & Mantovani, F. (2015). Leveraging psychology of virtual body for health and wellness. In S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 528–547). Wiley-Blackwell. <https://doi.org/10.1002/9781118426456.ch24>.
- Satava, R. M. (1994). Emerging medical applications of virtual reality: A surgeon's perspective. *Artificial Intelligence in Medicine*, 6(4), 281–288. [https://doi.org/10.1016/0933-3657\(94\)90033-7](https://doi.org/10.1016/0933-3657(94)90033-7).
- Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3, 74. <https://doi.org/10.3389/frobt.2016.00074>.

Ethics and Sustainability

11

Itziar Goicoechea

ABSTRACT

This chapter explores the role of ethics and sustainability in virtual reality (VR) design, emphasizing designers' responsibilities towards social, environmental, and cultural challenges. It discusses ethical principles, including privacy, data transparency, psychological safety, and inclusiveness, and demonstrates VR's potential to transform human behaviour through case studies. Sustainability is addressed from environmental, social, and economic perspectives, highlighting circular economy principles, energy efficiency, and equitable access, aligned with the United Nations Sustainable Development Goals (SDGs). This chapter advocates integrating ethics and sustainability from the earliest stages of design, using practical frameworks, indicators, and evaluations to guide responsible innovation. By providing tools and case studies, it prepares designers to create VR experiences that are not only immersive and engaging but also socially responsible, environmentally conscious, and economically sustainable.

INTRODUCTION

Virtual reality (VR) has emerged as a powerful technology capable of transforming how we learn, communicate, work, and interact with the world. By creating immersive and interactive environments, VR offers opportunities for innovation across many sectors, including education, healthcare, entertainment, architecture, and social activism. However, as the technology advances, it also raises important ethical and sustainability challenges that must be addressed from the earliest stages of design.

The immersive nature of VR can strongly influence users' emotions, behaviours, and perceptions, which places significant responsibility on designers and developers to anticipate potential social, psychological, and cultural impacts. In addition, the reliance on hardware, intensive computational resources, and data-driven systems raises concerns about environmental impact and equitable access to technology.

- This chapter argues that ethics and sustainability should be considered fundamental principles in the design of VR experiences. Through a critical and reflective approach, designers can promote more inclusive and responsible uses of immersive technologies. In particular, this chapter highlights the importance of: protecting users' rights, privacy, and well-being.
- Promoting inclusion and cultural sensitivity.

- Minimizing environmental impacts through sustainable design decisions.
- Aligning projects with broader social goals, such as the United Nations Sustainable Development Goals (SDGs).

Practical tools, frameworks, and examples are also presented to support designers in integrating these principles into their creative processes.

ETHICAL CONSIDERATIONS IN VR DESIGN

This section examines the application of ethical principles in the design of VR experiences, addressing issues such as privacy protection, transparent management of personal data, and the prevention of psychological harm. Through case studies, it explores how immersive environments can alter users' perceptions, influence behaviour, and generate social effects.

First, it is necessary to define two fundamental concepts: ethics and morality, and their relationship in the design of VR experiences. Ethics refers to the systematic reflection on what is right or wrong and guides decision-making in professional practice. Morality, in contrast, refers to the values, norms, and beliefs shared by a community at a given time. Within this framework, technological ethics, as a branch of applied ethics, addresses the responsibilities arising from the design, development, implementation, and use of technologies.

Designing VR experiences is not a neutral task. It involves decisions that influence how people perceive their environment, interact with others, and represent themselves. What is shown, how it is represented, and who is included or excluded reflect ethical criteria, whether explicit or implicit. Therefore, ethical principles serve as a compass guiding responsible, inclusive, and user-respecting practices.

Principles such as respect for autonomy, non-maleficence, justice, and beneficence, derived from bioethics and professional ethics, must be adapted to the specific context of VR, where immersion and the capacity to alter perception create new ethical challenges. Consequently, ethics should not be considered an element added at the end of the process, but a transversal dimension guiding the design and evaluation of immersive experiences.

Among the key aspects to consider are as follows.

User Autonomy

Autonomy refers to individuals' capacity to make informed decisions about their own experiences without being manipulated or unduly influenced. In the context of VR, this involves allowing users to maintain control over their experiences, especially in environments that may produce intense emotional effects. Deep immersion can subtly and imperceptibly influence emotions and decisions, raising ethical dilemmas regarding psychological persuasion, whether for commercial, political, or educational purposes (Marloth et al., 2020).

Case Study 11.1

In VR advertising, such as immersive games or marketing experiences, influencing users' decisions can become problematic if explicit consent is not ensured or if emotions are manipulated to prompt actions such as purchasing a product.

Privacy and Data Protection

VR requires the collection of personal and biometric data, such as eye movements, physical location, emotional responses, and behavioural patterns. This raises ethical dilemmas regarding who controls this information, how it is stored, and how privacy policies are communicated (Kumarapeli et al., 2024). While technology offers significant advantages, it also carries risks related to the misuse of sensitive data, which necessitates transparency, informed consent, and robust security measures.

Case Study 11.2

VR platforms may collect emotional data through biometric sensors. If this information is used without informed consent or shared with third parties, user privacy is violated and trust in the technology is undermined.

Inclusivity and Social Justice

Inclusivity is essential, as VR can reflect and amplify existing social biases. Designers must consider potential racial, gender, cultural, or disability-related prejudices that may affect avatars, narratives, or accessibility (Zhang et al., 2025). Inclusive design not only broadens access but also promotes fair representation and a sense of belonging.

Case Study 11.3

In some VR environments, avatars do not adequately represent the diversity of human bodies and characteristics, which may exclude people with disabilities or ethnic minorities. Platforms such as Rec Room (Rec Room Inc., 2016) have introduced more inclusive avatar options, allowing users to customize their appearance to better reflect their identity.

Psychological and Emotional Impacts

VR has significant therapeutic and educational potential, such as treating phobias or reducing anxiety, but it can also cause adverse effects, including cybersickness, eye strain, emotional agitation, or excessive rumination. To ensure safety, designers should implement measures such as pre-use psychological assessments, giving users control options, monitoring reactions, adjusting technical parameters, limiting duration and intensity, and providing clear exit points or breaks.

Case Study 11.4

VR is used to treat disorders such as post-traumatic stress disorder (PTSD). While exposure therapy can be effective, it may also cause emotional distress if not carefully designed or supervised. Research shows that although these interventions can aid trauma habituation in controlled settings, their application requires rigorous protocols and trained personnel to avoid adverse effects (Maples-Keller et al., 2017).

Integrating Ethics into the Design Process

Ethical principles must be integrated from the earliest stages of design, not added as an afterthought. This requires adopting a critical stance that questions not only what technology allows us to do, but what ought to be done (Madary, M., & Metzinger, T. K., 2016). Ethical responsibility is amplified in immersive environments: what is designed is not merely observed, but experienced. This raises key questions:

- Is user autonomy respected in environments that may manipulate emotions or induce behaviours?
- What boundaries should be established to protect participants' psychological integrity?
- What are the consequences of excluding certain voices or cultures from the virtual universe?

To ensure ethical considerations are applied at every stage, VR design should:

- Be transparent about the collection and use of data.
- Avoid stereotypes and represent identities in a diverse and fair manner.
- Establish narrative and visual limits to prevent adverse psychological effects, including traumatization.
- Include the participation of communities affected by the themes or contexts represented.

Furthermore, ethics requires considering the real consequences of design: monitoring the impact after release, being willing to modify or remove harmful content, and fostering a culture of responsibility within multidisciplinary teams.

IEEE Ethical Principles for Autonomous and Intelligent Systems

The Institute of Electrical and Electronics Engineers (IEEE) has proposed a framework called *Ethically Aligned Design* (IEEE Standards Association, 2024), which highlights principles applicable to the development of immersive technologies such as VR:

- **Transparency:** Users should be able to understand how a system works, including what data it collects and how decisions are made.
- **Accountability:** Designers must be accountable for the impacts of their creations, including unforeseen consequences.
- **Privacy:** Respect for personal and biometric data is crucial, especially in systems that record movements, reactions, and behavioural patterns.
- **Inclusivity:** Technologies must ensure they do not reproduce or reinforce social or cultural inequalities.

The Design Justice Network Framework

The Design Justice movement (Design Justice Network, 2024) offers a critical perspective on design from social and political viewpoints. Its approach is particularly relevant in VR, where narratives, environments, and virtual bodies are actively designed. Key principles include:

- Centring the voices of communities most affected by design systems.
- Redistributing power within design processes.
- Designing for collective liberation and justice, not just efficiency or entertainment.

UNESCO Principles on the Ethics of Artificial Intelligence

Although focused on artificial intelligence (AI), UNESCO's principles provide useful guidance for any emerging technology, including VR. Among these are:

- **Sustainability:** Design should consider environmental impact and promote sustainable development.
- **Human Rights:** Technology must respect and promote fundamental rights.
- **Diversity and Non-Discrimination:** Avoid biases and ensure fair representation of different cultures, genders, ages, and abilities.

EXAMPLES OF ETHICAL APPLICATION IN VR PROJECTS

- **Clouds over Sidra (Arora & Milk, 2015):** A UN VR experience depicting a Syrian refugee girl's life. The team collaborated with the community to ensure respectful representation, avoiding sensationalism.
- **VR for Medical Training (Sanderson et al., 2023; Aizenberg & Van den Hoven, 2020):** Protocols limit overly realistic simulations that could cause anxiety or trauma among trainees.
- **Rec Room and Social VR Environments:** Platforms implement moderation and “safe zones” to protect users from harassment, showing how design decisions shape ethical experiences.

SUSTAINABLE DEVELOPMENT AND VR

VR is a transformative technology, but it also has significant environmental, social, and economic impacts. As VR grows across industries, it is essential to address sustainability by reducing environmental footprints, promoting equitable access, and fostering social well-being. This section examines VR's environmental and economic effects and emphasizes the creation of inclusive and accessible content.

Key areas include raising awareness of environmental impacts, promoting sustainability education, and aligning with the United Nations SDGs. Figure 11.1 illustrates the three sustainability dimensions: environmental, social, and economic, relevant to VR environments.



FIGURE 11.1 Sustainability dimensions in virtual reality contexts.

Environmental Sustainability in VR

The environmental impact of VR is increasingly urgent. Key factors include hardware production and disposal, energy consumption, and potential positive contributions in reducing other environmental impacts.

Hardware Production and Disposal

VR hardware, such as headsets, controllers, and computers, requires significant raw materials, energy, and labour. Demand for more powerful devices increases carbon footprints, and device obsolescence contributes to e-waste (Baldé et al., 2017).

Reducing environmental impact involves designing durable, repairable products and adopting a circular economy approach that facilitates disassembly and component reuse. Using sustainable materials and implementing recycling programmes can further mitigate adverse effects.

Energy Consumption of VR Platforms

VR applications, especially those with high-quality graphics or real-time simulations, can be energy intensive, particularly when powered by non-renewable sources (Stallkamp, 2020).

Energy use can be reduced by optimizing software to require less processing power without compromising user experience and by integrating VR systems with green energy solutions such as solar or wind power.

Reducing Environmental Impacts with VR

Although VR has environmental costs, it can reduce impacts elsewhere. For example, immersive VR allows meetings, conferences, and educational activities to occur remotely, lowering travel-related carbon emissions (Freeman et al., 2017). Similarly, virtual education reduces the need for physical attendance, further decreasing emissions.

Case Study 11.5

Virtual tourism allows users to explore iconic landmarks or fragile ecosystems without travelling, reducing carbon emissions and helping preserve vulnerable destinations.

Sustainability in VR Design

Technological sustainability is a challenge in a world striving to reduce its environmental impact. The energy consumption of VR platforms, hardware waste, and device production all carry significant environmental costs. Designers must consider these factors when creating VR experiences, seeking ways to minimize the technology's environmental impact (Ahern & Cairns, 2022).

Case Study 11.6

VR companies, such as Oculus and HTC, are beginning to consider sustainability in device design, but the product lifecycle remains a challenge. Recycling, component reuse, and lower-impact designs will be essential to reduce VR's long-term environmental footprint.

Social Sustainability in VR

In addition to environmental considerations, social sustainability is crucial to ensure that VR technology serves the broader goal of promoting equitable access to education, social experiences, and cultural exchange. VR has the potential to make significant contributions to social sustainability by promoting equity in access to technology, educating users about important social issues, and designing experiences that foster empathy and global awareness.

Promoting Equity in Access to Technology

One of the main goals of social sustainability in the context of VR is to ensure that all individuals, regardless of their background, income, or physical ability, can access and benefit from VR technology. For example, VR has the potential to reduce educational gaps by providing access to immersive learning experiences that would otherwise be out of reach for people in underserved or remote areas (Borghesi & Vercelli, 2020). To promote equity in access to VR, it is essential to address the digital divide. VR development should consider affordability, usability, and accessibility. Offering low-cost VR devices and free or subsidized educational content could help make the technology more inclusive, particularly for disadvantaged communities.

Case Study 11.7

In regions with limited access to traditional education, VR programmes provide students with immersive learning experiences. For example, ClassVR (Magalhães et al., 2025) integrates VR into school curricula, allowing students from diverse backgrounds to explore science, history, and geography in 3D environments.

Using VR for Sustainability Education

VR also offers a unique opportunity to educate people about sustainability in an engaging and memorable way. Through immersive experiences, users can explore vulnerable ecosystems, witness the consequences of climate change, and observe the effects of human activity on the planet. These experiences not only raise awareness but also inspire action towards sustainability (Ahn & Cairns, 2014).

Case Study 11.8

The VR climate change experience allows users to witness the impact of climate change on various ecosystems, such as coral reefs or polar ice caps. By immersing users in these environments, the experience creates a powerful emotional connection to the issue, motivating people to take action to reduce their carbon footprint and advocate for environmental policies.

Designing Experiences That Promote Empathy and Global Awareness

Social sustainability in VR includes designing experiences that foster empathy and global awareness. VR lets users step into others' perspectives and experience different cultures in ways traditional media cannot (Sutcliffe et al., 2019), helping people connect across geographical, cultural, and social boundaries.

For example, VR can simulate living with a disability, enabling users to understand associated challenges, reduce stereotypes, and promote inclusivity.

Case Study 11.9

The Empathy VR Project (VR Society, 2022) immerses users in scenarios such as homelessness or life as a refugee. Experiencing these perspectives helps users empathize and advocate for those in need, fostering a more inclusive society.

Economic Sustainability in VR

Sustainability in VR requires balancing environmental, social, and economic aspects. Economically sustainable projects ensure financial viability, enabling continuity and scalability over time.

Economic sustainability is not just about short-term profit but about managing costs and investments responsibly. It also involves considering environmental and social returns alongside financial indicators, promoting projects that combine profitability with social, cultural, and ecological value.

Designing economically sustainable VR projects requires strategies focused on resource optimization, financial transparency, and long-term solutions that benefit communities and the environment.

Case Study 11.10

VirtuFarm is a VR platform that trains small-scale farmers in sustainable agricultural techniques. Rather than relying on high one-time sales or intrusive advertising, it uses low-cost licences, affordable annual subscriptions, and partnerships with governmental organizations and nongovernmental organizations (NGOs) to ensure financial viability. Modular content allows scalable updates, optimizing investment and avoiding costly redesigns.

The project's Social Return on Investment (SROI) was 3.5, meaning every euro invested generated 3.5 euros of social and economic value through improved agricultural practices, productivity, and reduced environmental impact. This model keeps the platform accessible while ensuring operational sustainability, combining profitability with social and environmental responsibility.

Evaluation and Measurement of Ethical and Sustainable Impact for VR Designers

For VR designers, evaluating and measuring ethical and sustainable impact is essential. These practices enhance the quality of experience, promote responsible use of technology, and help guide projects towards positive outcomes for users as well as the social and environmental context.

Impact is understood as the influence that VR experiences have on users, including changes in awareness, understanding, attitudes, and behaviours related to ethical, social, and environmental aspects. Assessing this impact enables designers to identify areas for improvement, optimize outcomes, and demonstrate the effectiveness of the project to sponsors, communities, and other stakeholders.

Designers can combine quantitative and qualitative methods. Quantitative methods include surveys and usage metrics, such as session duration or interaction rates, which provide numerical data on the experience. Qualitative methods consist of interviews or focus groups conducted after the experience, offering insights into user perceptions, emotions, and behavioural impact. Combining both approaches provides a comprehensive and accurate view, ensuring that VR experiences are innovative, responsible, and sustainable.

To guide the evaluation, designers can use indicators that measure both ethical impact and sustainability:

- **Ethics:** Consider diversity, inclusion, and social justice; assess how representations influence the perception of different communities and cultures.
- **Sustainability:** Measure the environmental impact throughout the product’s lifecycle and promote efficiency in resource use, encouraging recycling and reuse in hardware and content development.

Table 11.1 presents relevant case studies and the lessons learned, which can serve as a guide to improving design practices.

TABLE 11.1 Case studies: VR projects with effective evaluations

PROJECT	DESCRIPTION	EVALUATION	KEY LESSON LEARNED
Earth: the power of the planet VR	This project uses virtual reality to demonstrate the effects of climate change and environmental degradation across various ecosystems worldwide. Users witness habitat destruction and gain a clearer understanding of the scale of the environmental crisis.	Pre- and post-experience surveys were conducted to measure changes in users’ environmental awareness. Results showed a significant increase in interest in participating in conservation activities.	Measuring changes in attitudes and behaviours after the experience is crucial, enabling adjustments to content and narrative to maximize educational impact.
VR for Good: project X	This initiative employs virtual reality to address social issues such as poverty and education. Through immersive storytelling, users experience the daily challenges faced by individuals in disadvantaged situations.	User engagement and participation were monitored, alongside interviews assessing the depth of empathy generated by the experience.	Personal and emotionally resonant narratives are highly effective in fostering empathy and understanding, highlighting the importance of incorporating real stories into the design process.
Ocean cleanup VR	This project raises awareness about marine pollution and the work of The Ocean Cleanup in clearing the oceans. Users can virtually participate in collecting plastic waste from the sea.	Interaction metrics and surveys were used to evaluate how the experience influenced users’ perceptions of marine pollution and their willingness to act.	Providing users with a direct, action-oriented experience can motivate behavioural change, such as supporting cleanup initiatives and adopting more sustainable habits.
Virtual reality climate change education (VRCC)	This educational programme uses virtual reality to teach students about climate change and its impacts. Students can explore various climate scenarios and potential solutions.	Learning assessments were implemented to measure information retention and students’ ability to apply knowledge in real-world contexts.	Including formative and summative learning assessments during and after the experience strengthens educational effectiveness and identifies areas requiring further instructional support.

LEGAL CONSIDERATIONS IN THE CREATION OF VR CONTENT

The development of immersive experiences must comply with an ethical and regulatory framework that ensures social responsibility, sustainability, and user protection. The main legal challenges include intellectual property, data protection, content regulation, educational and labour legislation, and health and safety.

Intellectual Property

In VR environments, multiple elements are protected by copyright, including virtual worlds, avatars, music, code, graphics, and animations. In interactive spaces, several creators may coexist (developers, platforms, and users), so ownership depends on:

- The platform's terms of service.
- The degree of creative authorship by the user.
- Specific contracts or agreements (for example, in educational or workplace contexts).

In general, moral rights usually belong to the human author, while exploitation rights may be subject to contracts or licences.

For example, if a student creates an interactive simulation within an educational VR platform, they may retain moral rights as the author, whereas the university or the software provider could hold the exploitation rights of the content, depending on the terms of use or established agreements.

Therefore, who is the legal author of creations generated by AI or users?

This issue is especially relevant in VR experiences that incorporate AI-generated content, such as design assistants, procedural modelling, or creative tools powered by AI. If a user creates an environment or avatar with the assistance of AI, do the rights belong to the user, the AI, or the platform?

Currently, only natural people can be considered legal authors. Works generated exclusively by AI, without significant human intervention, are not protected by copyright in most jurisdictions. If the user actively participates through detailed instructions, selection, editing, or creative decisions, they can be considered the author. Platforms offering AI tools generally allow the use of the generated content, although they may reserve certain licences or rights according to their terms of use. Table 11.2 presents a summary of the different scenarios addressed.

TABLE 11.2 Scenarios and copyright ownership

SCENARIO	WHO HOLDS THE RIGHTS?
Content generated solely by AI, without human intervention	No legal protection (no human author)
User contributes ideas, edits, or combines creatively	The user may be considered the legal author
Platform provides the AI and sets conditions	The user generally holds rights, but subject to limits set by the terms of use

RECOMMENDATION: ETHICS AND LEGAL CONSIDERATIONS

- Clearly define intellectual property rights in the Terms of Use.
- Offer co-ownership models or shared licences to promote equity and collaborative innovation.

Use of Third-Party Material (Licences and Copyright)

The use of music, 3D models, videos, graphics, or trademarks without authorization may constitute an infringement. It is common for resources to be used improperly, such as downloading content without checking licences or including real-world brands in virtual environments without permission.

Recommendations:

- Use open licences (Creative Commons, MIT, GNU).
- Always review the terms of use.
- Avoid decorative use of registered trademarks without authorization.

RECOMMENDATION

Legal and cultural sustainability promotes a culture of original creation, respect for authorship, and responsible reuse of digital resources.

Data Protection and Privacy in Immersive Experiences

VR can collect sensitive data, including biometrics, voice, eye-tracking, and body movements. This requires compliance with regulations such as:

- General Data Protection Regulation (GDPR) – European Union
- California Consumer Privacy Act (CCPA) – United States
- XRSI Privacy Framework – voluntary standards

Organizations must ensure transparency, obtain informed consent, minimize data collection, and guarantee the security of its storage and processing.

RECOMMENDATION

- **Ethical Approach:** Respect user autonomy, obtain informed consent, and minimize data collection. Users should be clearly informed and provide explicit consent for the collection of their data. This is especially important in VR environments, where immersion can make users less aware of how their data is being used.
- **Sustainable Approach:** Avoid data overexploitation and promote transparency. Organizations should be open about what data is being collected and how it is used, including informing users about any tracking or analysis performed during the VR experience.

Legal Responsibility and Content Regulation

Immersive environments can amplify risks such as harassment, discrimination, behavioural manipulation, or exposure to inappropriate content. Creators and platforms may be held responsible for:

- Copyright infringements
- Illegal content (hate speech, violence, child pornography, etc.)
- False or misleading information that causes harm

Platforms must moderate content, establish clear terms of use, and provide reporting mechanisms. Key challenges include determining responsibility among users, the global scope of VR, the lack of specific legislation, and regulatory gaps concerning minors exposed to extreme or persuasive content.

RECOMMENDATION: SOCIAL SUSTAINABILITY

- Design VR environments that are safe and inclusive, with particular attention to vulnerable groups.
- Establish clear behavioural standards, such as a virtual code of conduct, along with effective moderation and control mechanisms.
- Carefully assess the psychological impact of content, for example, in hyper-realistic violent games or simulated trauma experiences.

Educational, Labour, and Accessibility Legislation

When VR is used for educational or professional purposes, additional regulations apply:

- **Child Protection:** Strict compliance with privacy laws and assessment of the psychological impact of content
- **Digital Labour Rights:** Respect for privacy, protection of biometric data, and limitation of excessive monitoring
- **Accessibility:** Adaptation of environments for people with disabilities, ensuring inclusive interfaces and compatibility with assistive technologies, in accordance with current regulations

RECOMMENDATION: ETHICAL AND SUSTAINABLE APPROACH

- Avoid practices that overload or manipulate users through hyperstimulation or excessive tracking of their data and behaviours.
- Adopt universal design principles to ensure accessibility and usability for people with diverse abilities.
- Limit immersive exposure time to prevent cognitive fatigue and minimize potential adverse effects on users' mental and physical health.

Health, Safety, and Well-Being Legislation

Immersive experiences can pose physical and psychological risks:

- Visual and cognitive fatigue
- Motion sickness (kinetosis)

- Emotional impact from sensitive content
- Physical hazards in the real environment
- Vulnerability of minors or individuals with neurological conditions

Although there is no global legislation specific to VR, general product safety standards and user-centred design principles should be applied, including ergonomic standards and technical requirements. In the European Union, devices must comply with regulations on electronic safety, CE marking, electromagnetic compatibility, hazardous substances, and cybersecurity, which are particularly relevant in educational and workplace contexts.

CONCLUSIONS: VR AS A CATALYST FOR ETHICS AND SUSTAINABILITY

VR has substantial potential to promote ethics and sustainability. By reducing travel and education impacts, fostering social equity, empathy, and global awareness, and prioritizing energy efficiency, circular economy practices, and inclusive content, VR can help build a more sustainable and equitable future.

VR designers must recognize their responsibility to create experiences that are innovative, ethical, inclusive, and sustainable. Implementing robust ethical principles and continuously assessing social impacts ensures VR benefits society responsibly.

To support this, a practical checklist guides designers in integrating responsibility, inclusion, environmental awareness, and user well-being throughout the creative and technical process, aligning projects with global standards and values (Figure 11.2).



FIGURE 11.2 Checklist of ethical and sustainable best practice for VR designers.

REFERENCES

- Ahern, S., & Cairns, P. (2022). Sustainable interaction design: Considering the environmental impact of immersive technologies. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1–22. <https://doi.org/10.1145/3512963>.
- Ahn, S. J., Bailenson, J., & Park, D. (2014). Short- and long-term effects of embodied experiences in immersive virtual environments on environmental locus of control and behavior. *Computers in Human Behavior*, 39, 235–245. <https://doi.org/10.1016/j.chb.2014.07.035>.
- Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, 7(2), 1–14. <https://doi.org/10.1177/2053951720949566>.
- Arora, G., & Milk, C. (Directors). (2015). *Clouds over Sidra* [Virtual reality film]. United Nations & Samsung. <https://unvr.sdgactioncampaign.org/clouds-over-sidra/>.
- Baldé, C. P., Forti, V., Gray, V., Kuehr, R., & Stegmann, P. (2017). *The global e-waste monitor 2017: Quantities, flows and resources*. United Nations University, International Telecommunication Union, and International Solid Waste Association. <https://doi.org/10.1007/s40831-018-0201-5>.
- Borghesi, S., & Vercelli, A. (2020). Digital inclusion and equity in emerging technologies: the case of virtual reality in education. *Journal of Technology and Society*, 12(3), 45–58.
- Design Justice Network. (2024). *Principles of design justice*. <https://designjustice.org/read-the-principles/>.
- Freeman et al. (2017). *Virtual reality in the assessment, understanding, and treatment of mental health disorders*. *Psychological Medicine*, 47(14), 2393–2400. <https://doi.org/10.1017/S003329171700040X>.
- IEEE Standards Association. (2024). *Best practice guide to developing ethically aligned standards*. IEEE. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf.
- Kumarapeli, D., Jung, S., & Lindeman, R. W. (2024). Privacy threats of behaviour identity detection in VR. *Frontiers in Virtual Reality*, 5, 1197547. <https://doi.org/10.3389/frvir.2024.1197547>.
- Madary, M., & Metzinger, T. K. (2016). Real virtuality: a code of ethical conduct recommendations for good scientific practice and the consumers of VR-technology. *Frontiers Robotics and AI*, 3, 3. <https://doi.org/10.3389/frobt.2016.00003>.
- Magalhães, C., Bento, M., & Lencastre, J. (2025). Immersive virtual learning spaces for emotional engagement in education with the classroom-ready virtual reality device CLASSVR. In Paper presented at the *11th International Conference of the Immersive Learning Research Network (iLRN 2025)*, Chicago, Illinois, USA, June 15–19.
- Maples-Keller, J. L., Bunnell, B. E., Kim, S. J., & Rothbaum, B. O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review of Psychiatry*, 25(3), 103–113. <https://doi.org/10.1097/HRP.0000000000000138>.
- Marloth, M., Chandler, J., & Vogeley, K. (2020). Psychiatric interventions in virtual reality: Why we need an ethical framework. *Cambridge Quarterly of Healthcare Ethics*, 29(4), 663–675. <https://doi.org/10.1017/S0963180120000328>.
- Rec Room Inc. (2016). *Rec Room* [Virtual reality social platform]. <https://recroom.com/>.
- Sanderson, et al. (2023). *AI Ethics Principles in Practice: Perspectives of Designers and Developers*. *IEEE Transactions on Technology and Society*, 4(2), 171–187. <https://doi.org/10.1109/TTS.2023.3257303>.
- Stallkamp, J. (2020). Energy consumption of virtual reality systems: A review. *Journal of Sustainable Computing*, 25, 100377. <https://doi.org/10.1016/j.suscom.2020.100377>.
- Sutcliffe, et al. (2019). Reflecting on the Design Process for Virtual Reality Applications. *International Journal of Human-Computer Interaction*, 35(2), 168–179. <https://doi.org/10.1080/10447318.2018.1443898>.
- VR Society. (2022). *Empathy VR project: Immersive experiences for social change*. Retrieved from <https://www.vrsociety.org/empathy-vr-project>.
- Zhang, K., Scott Spencer, E. G., Manikandan, A., Li, A., Yao, Y., & Zhao, Y. (2025). Inclusive avatar guidelines for people with disabilities: Supporting disability representation in social virtual reality. *arXiv*. <https://doi.org/10.48550/arXiv.2502.09811>.

Accessibility and Inclusive Design

12

Tatiana Santos Gonçalves

ABSTRACT

Accessibility has been an important topic thoroughly investigated in the last decade, especially in the context of the transformation from print to Internet content. However, immersive technologies like virtual reality (VR) are fundamentally very different from 2D interfaces and therefore pose new accessibility challenges that have not yet been fully understood or investigated. In VR, users move through three-dimensional spaces and manipulate objects, which creates both obvious and very subtle barriers that were not present before and hence need to be investigated. In this study, we aim to examine accessibility issues in the context of immersive environments through the lens of the ability spectrum, which recognizes the full range of human capabilities rather than focusing only on a particular type of disabled people. We discuss practical strategies for inclusive design, including ability-informed personas, adaptable interaction methods, spatial adjustments for comfort and reach, and the use of captions and audio description. We conclude that all these aspects should be taken into account to derive a useful guide for designers to identify accessibility issues that would otherwise remain undetected until later stages of deployment.

INTRODUCTION

Accessibility has become a central concern in the design of contemporary digital systems and as such it has been thoroughly investigated in the last decades (Santos Gonçalves, 2020, 2021). Accessibility has been defined in various ways, but here we will regard it as the “usability of a product, service, environment or facility by people with the widest range of capabilities”. This means that all users can perceive, understand, navigate, and interact with it (W3C, 2018). Ensuring equitable access to the emerging technologies further reinforces the need to consider the diverse needs of users, maximizing the accessibility, inclusion, and enjoyment capacities of users.

In a United Nations report published in 2018, the United Nations Committee on the Rights of Persons with Disabilities emphasized the importance of making greater efforts to ensure that everybody with and without disabilities enjoys equal rights, particularly regarding social integration and access to products and services (WHO, 2018). This effort toward equality and democratic participation is even more critical in the context of technology as people with physical disabilities or sociodemographic limitations often encounter significant barriers when trying to access the Internet. Hence, ensuring the accessibility of technological content is essential so that all users can fully benefit from the opportunities provided by novel

ways of communication and interaction. However, having technological infrastructure alone is not sufficient. Resources must also be implemented effectively in order to improve the real and practical accessibility of web services for these users. In this sense, it is paradoxical that the change from print to Internet content has generated barriers in terms of access to local information for certain groups of users, such as the elderly and disabled people. For these reasons, accessibility is a major concern in today's society as elderly and disabled people are at risk of being excluded from the use of technologies.

The large body of research has led to a broad consensus that digital technologies should be usable by the widest possible range of people, regardless of physical, sensory, or cognitive differences. For instance, within web environments, this principle has been assessed through internationally recognized standards such as the Web Content Accessibility Guidelines developed by the World Wide Web Consortium (W3C, 2018). In fact, most documents that establish rules and standards for web accessibility are based on the guidelines developed by the World Wide Web Consortium (W3C). The W3C is the leading international organization responsible for standardizing requirements for the development and interpretation of web content. One of its most important initiatives is the Web Accessibility Initiative (WAI), whose objective is to promote web accessibility for people with disabilities. To define accessibility standards for websites, the WAI provides guidelines that include rules and recommendations for the development of platforms, browsers, and web content. These requirements include, among others, offering equivalent alternatives for alternative text, keyboard navigation, captioning, and semantic structuring of content. As a result, the W3C guidelines (2018) have served as the foundation for the development of various recommendations and regulatory obligations adopted by member states of the European Union. In summary, these guidelines have been very useful, providing important guidance to improve users' experiences of 2D interfaces by promoting practices such as alternative visual content, providing contextual information and guidance to assist users with special needs in understanding web pages, and ensuring clear and consistent navigation mechanisms.

All this underlines the fact that accessibility issues derived from the transformation from print to Internet content have been thoroughly investigated. However, virtual reality (VR) is fundamentally different from 2D user interfaces/experiences. The emergence of new immersive technologies raises novel questions and challenges regarding accessibility that should be investigated. In this new context, interaction occurs within a three-dimensional environment where perception, movement, and spatial orientation are integral components of the experience (Dudley et al., 2023). Users do not simply read or listen to information; they move within a simulated space, manipulate virtual objects, and navigate environments through gestures, controllers, or other input modalities (Martínez-Cano, 2024; Wobbrock, 2011). As a consequence, novel barriers appear that were not present before. These can be generated from missing information or even from the physical and perceptual assumptions embedded in the design of interaction techniques. Hence, there is a need for a more inclusive design with a focus on the possible accessibility issues that these very new and changing technologies can generate.

The possible accessibility barriers that appear in these immersive environments are much more nuanced and involved than the ones we have in, for instance, computers, tablets, or mobile devices. For this reason, they are frequently not detected by the designers. For example, many VR applications implicitly assume a particular user profile: a person standing upright, with full upper-body mobility, comfortable reach, and tolerance for rapid visual motion. These assumptions frequently remain invisible during development because they reflect normative expectations about the body. This is particularly important in, for instance, immersive systems deployed for educational or scientific purposes. Such general assumptions may restrict participation for many users whose physical or sensory conditions differ from those of the statistical model. Accessibility in immersive environments requires further investigation as it is not solved by simply adapting existing web-based accessibility guidelines. It demands a design approach that acknowledges the diversity of human abilities and the variability of interaction contexts (La Viola et al., 2017; Mott et al., 2019).

Our research aims to shed light on these important issues in order to provide designers with practical information to improve the democratization of access to immersive technologies for different people. To achieve this goal, it is necessary to design applications considering accessibility issues, so as to reduce potential barriers. We believe that, for a better and more democratic use of immersive technologies, it is important that these can be reached by the largest possible number of users. In this regard, we start from the premise that all these new emerging technologies are essential sources of information or entertainment and believe that an improvement in accessibility would stimulate their use in a more balanced and

democratic way among all users. This effort toward equality should be even greater when it comes to the use of the technology by the most disadvantaged people, as the elderly or people affected by a disability. These people face numerous accessibility challenges that must undoubtedly be addressed.

However, this study focuses on accessibility and inclusive design in VR from the perspective of the *ability spectrum*. Rather than treating accessibility as a set of specialized adaptations for a minority of users, the ability spectrum considers all types of people, recognizing that the different conditions of users might vary along multiple dimensions, including reach, visual acuity, hearing, motor precision, and tolerance to motion or sensory stimulation (Creed et al., 2024; Mott, 2019). Hence, it is important that when designing these systems, this variability of conditions is taken into account, allowing their use by the widest possible range of people, and this work aims to provide guidance in this regard.

To this goal, the discussion that follows focuses on several practical dimensions of inclusive immersive design. First, we analyze how ability spectra can inform the construction of personas and scenarios used during early design stages. Second, we consider the role of alternative inputs, interaction remapping, and tolerance windows in expanding the range of possible interactions. We pay particular attention to the adaptation regarding the spatial interaction to individual differences in height, reach, and comfort which is addressed in the “Calibration Flows (Height, Reach, Comfort)” section. The “Visual/Audio Accessibility and Captions/Audio Description” section is devoted to the study of visual and auditory accessibility, including the integration of captions and audio description in spatial environments. Finally, we discuss methodological approaches for testing immersive systems with diverse users and documenting accessibility barriers in a systematic manner. We conclude that situating accessibility within the broader design process, rather than addressing the problems once it has been developed, seems to help immersive technologies to support more inclusive participation across a wide spectrum of users.

ABILITY SPECTRUM, PERSONAS, AND SCENARIOS

Accessibility is a major concern in today’s society as elderly and disabled people are at risk of being excluded from the use of modern technologies. Hence, it is important to make an effort toward equality and democracy, and this effort should be even greater when it comes to the use of technology, as people affected by a physical disability or sociodemographic limitation face numerous challenges. This is a crucial topic that deserves to be thoroughly investigated. Here, however, we take a different approach and consider a wider view on the accessibility barriers that appear in the new context of immersive technologies. Our analysis focuses on the notion of ability spectrum which is concerned with accessibility issues of all users rather than focusing on a particular type.

The concept of ability spectrum acknowledges that distinct categories, physical, sensory, and cognitive capacities, exist among all users, and these should be taken into account when aiming at an inclusive design (Szentirmai, 2024). Because interaction in immersive systems relies on coordinated movement, spatial orientation, and perceptual processing (Martínez-Cano, 2024; Dudley et.al., 2023), the effects of this variability are amplified as opposed to the 2D technology where these effects are much less relevant. As a result, the level of immersive engagement may be significantly impacted by variations in, for instance, shoulder mobility, visual acuity, auditory discrimination, or motion sensitivity.

For all these reasons, inclusive persona development needs to be further developed and investigated. For that, we need to first identify measurable characteristics, for instance, we can mention that functional reach envelope, maximum comfortable rotation range, arm elevation endurance thresholds, tolerance to abrupt changes in field of view, and dependence on assistive input technologies should all be included in personas for immersive design (Creed et al., 2024; Mott, 2019). It is crucial to develop a scenario for testing the actual embodied diversity rather than theoretical statistical averages and for this we need to consider such variables.

Research has been carried out in this direction. For instance, empirical research by Hartfill et al. (2025) has shown how adaptive mapping techniques that maintain performance while taking into account unilateral motor profiles can be used to redesign interaction schemes based on symmetrical bimanual

manipulation. Their results highlight an important conclusion: rather than being a result of intrinsic technological limitations, immersive exclusion usually arises from constrained interaction logic.

Hence, the environmental context should be considered when modeling scenarios. For example, someone interacting in a small space at home is not the same as using VR while seated at work. Therefore, designing considering the environmental variability should improve the experience of inclusivity in real-world scenarios and decrease deployment fragility. Moreover, as mentioned above, the ability spectrum perspective is a wider concept than the one typically considered in the literature, where the focus is on particular types of users (elderly or disabled, for instance). It highlights that inclusive design in immersive environments must address not only permanent disabilities but also situational and temporary variations in human capability in more general terms. The abilities of any user fluctuate depending on fatigue, context of use, stress, and environmental conditions. For example, prolonged arm elevation during mid-air interaction may lead to muscular fatigue even in users with no motor impairments, a phenomenon often referred to as the “gorilla arm effect” in mid-air interaction research (La Viola et al., 2017). In the same way, contexts with rapidly changing images may present challenges for some users. These are accessibility issues that can affect a wide diversity of users and not only people affected by a physical disability or a sociodemographic limitation. The latter issue deserves further investigation and lies beyond the scope of this work.

Within immersive systems, users must continuously coordinate body posture, hand movement, gaze direction, and spatial awareness to interact with virtual objects and navigate environments. Hence, these considerations are of particular relevance in this context, and research has shown that physical constraints significantly influence how users manipulate objects and explore virtual spaces (Bowman et al., 2004; Dudley et al., 2023). These constraints can be, for example, limited reach range, reduced joint mobility, or restricted locomotion. So, when a designer assumes unrestricted movement, he/she implicitly encodes an “average” user model that does not reflect the variability present across real populations. This is an example of something that must be avoided.

For all these reasons an inclusive design must take into account the ability spectrum that requires the integration of functional capability parameters into design artifacts. Instead of representing users primarily through demographic characteristics, personas can incorporate measurable dimensions such as reach radius, comfortable interaction height, grip stability, visual contrast sensitivity, or motion tolerance thresholds. So, it is important that the designer incorporates all these features to evaluate interaction techniques against a wider range of embodied conditions. As argued in accessibility research, designing for variability in physical and perceptual abilities often leads to solutions that benefit broader user populations, including individuals experiencing temporary impairments or constrained contexts of use (Mott et al., 2019; Creed et al., 2024).

Scenarios developed from such general users can also facilitate systematic testing of interaction robustness. Hence, instead of developing prototypes exclusively for users who fit normative interaction assumptions, scenario-based design can simulate diverse capability configurations. For example, interaction tasks may be evaluated under conditions of restricted reach, reduced field of view, or limited input modalities. We will consider this approach in this work as it enables us to guide designers to identify interaction breakdowns that would otherwise remain undetected until later stages of deployment. In immersive systems, where interaction techniques often combine multiple input channels—gesture, gaze, locomotion, and voice—such scenario-based evaluation is particularly valuable.

We shall consider further another important dimension that concerns sensory tolerance and motion perception. This is motivated by the fact that immersive technologies can induce discomfort in users. For instance, there are a large number of users that are sensitive to motion-visual discrepancies or rapid changes in viewpoint. Some of the scarce research devoted to cybersickness seems to point out that individual tolerance to motion varies considerably and is influenced by vestibular sensitivity, visual processing, and prior exposure to immersive systems (La Viola, 2000; Keshavarz et al., 2014). Personas that encode motion sensitivity thresholds or preferred locomotion modalities can therefore guide the selection of interaction techniques, such as teleportation, incremental turning, or gaze-directed navigation, that minimize discomfort for a wider range of users.

Finally, we shall consider environmental context into scenario modeling as it acknowledges that immersive interaction does not occur in a vacuum. In fact, the physical setting in which VR systems are used—home environments, classrooms, laboratories, or public spaces—imposes different spatial constraints, lighting conditions, and levels of physical mobility. Already in Bowman & McMahan (2007) it was shown that interaction techniques designed for large open spaces may not translate effectively to domestic environments where users operate within limited spatial boundaries. This is an important observation that underlines the fact that considering environmental variability during persona and scenario development should help to ensure that immersive systems remain usable across a wide range of real-world contexts.

In conclusion, integrating the ability spectrum into persona construction and scenario modeling can enormously help designers to develop systems beyond the assumption of a single normative user profile or focusing on a particular group of users. Instead, immersive systems can be evaluated and developed taking into account a wide range of human variability, hence supporting interaction designs that remain functional, comfortable, and meaningful across a broad range of user experiences.

ALTERNATIVE INPUTS, REMAPPING, AND TOLERANCE WINDOWS

We shall consider a wide range of input options in immersive systems, namely, modalities, such as hand-held controllers, optical hand tracking, voice commands, gaze-based selection, adaptive switches, and hybrid configurations. The relevance of these may vary significantly among the systems and situations. In other words, dependency on a single physical capacity should be reduced by designing these modalities as interchangeable rather than hierarchical. This provides a lot of flexibility in the design. Further, every interactive action, such as selection, scaling, rotation, and locomotion, has to be related to a command that can be initiated via various options. This is in contrast with being connected to a fixed gesture and again provides more flexibility to the designer. Accommodation is made easier by this modular architecture without compromising coherence. In fact, Hartfill et al. (2025) empirically showed that such a way of interaction can improve the completion time and precision of the tasks.

We shall consider yet another important design factor: the tolerance windows. Within these contexts, timing constraints, alignment precision requirements, and gesture recognition management are frequently determined by technical parameters rather than human variability. Hence, this is an area that can be further improved by extending these windows, for example, by using spatial buffer zones, confirmation prompts, or adjustable sensitivity, we aim at decreasing the exclusion caused by weariness, tremor, or irregular tracking. Not surprisingly, Pérez et al. (2023) showed that adaptive interaction parameters in immersive environments improve both general usability and accessibility outcomes. In this way we seek that accessibility is kept traceable instead of being incidental, and we achieve this by recording remapping structures and tolerance modifications during development.

In order to clarify the relationship between ability variability and interaction design we make use of a structured mapping of potential barriers and corresponding design strategies to mitigate them. Indeed, within immersive environments, accessibility often emerges from the alignment between user capabilities, interaction techniques, and configurable system parameters.

CALIBRATION FLOWS (HEIGHT, REACH, COMFORT)

It is important to point out that in the contexts of VR, calibration processes are often restricted to boundary definition and device alignment. However, again, we believe that calibration must function as a negotiated configuration process between the user and the system from the standpoint of inclusive design. For example, height calibration should allow for the explicit declaration of seated, standing, or mixed posture modes, along with a corresponding recalibration of interaction zones and object placement. As a consequence, users who interact while seated are implicitly marginalized by systems that make standing assumptions by default. This must be considered in our analysis of calibration flows in order to improve the experience of all users.

Furthermore, it is readily apparent that real-time measurement will help us to take care of the reach calibration. In fact, this is already possible as systems can evaluate real reach envelopes and modify spatial distribution instead of assuming uniform arm length and shoulder mobility. For these reasons, this adaptation will help to avoid disadvantage for users with limited mobility in, for instance, educational simulations that require manipulation of laboratory staff.

As for the measurement of the comfort calibration, we first point out that it encompasses both sensory intensity and movement mechanics. While teleportation may interfere with spatial continuity for some users, continuous forward motion may cause cybersickness in those with increased vestibular sensitivity. On the other hand, we note that different comfort thresholds are supported by the field-of-view modulation, turn increments, and movable parameters. For instance, in Szentirmai (2024), these modifications ought to be viewed as accessibility-critical components of immersive educational infrastructures rather than as user preferences. Hence, we believe that calibration should stop being a technical routine and instead become the center of inclusive negotiation with the users.

VISUAL/AUDIO ACCESSIBILITY AND CAPTIONS/AUDIO DESCRIPTION

Before addressing the texting, it is important to discuss about the spatial challenges related to visual and auditory accessibility that may occur in the immersive systems. As opposed to the so far classical two-dimensional interfaces, VR distributes information within a three-dimensional field and so it is subject to constant head movement. Again, we face a novel situation that did not appear in the previous contexts and so it must be taken into consideration. For instance, captioning strategies need to consider orientation variability as when the user shifts focus, static subtitles that are anchored to a fixed spatial coordinate might vanish from view. This problem was addressed in Brescia-Zapata et al. (2025), and the authors showed that dynamic caption anchoring greatly enhances comprehension and lessens cognitive load when it is either affixed to the speaking avatar or moved within the user's field of view.

Moreover, different visual profiles can be considered to provide the user control over caption size, contrast, persistence duration, and placement. In a similar way, we can incorporate spatial references to audio descriptions. Again, critical environmental information is usually not properly communicated through descriptive narration and, as a consequence, rules out the possibility of directional context. For this reason, Creed et al. (2024) stressed in their work the necessity of iteratively improving spatial audio strategies. On the other hand, we need to avoid users from being overloaded with too many simultaneous information. Hence, it is important to ensure that sensory diversity does not result in a disadvantage. This can be achieved, for instance, by incorporating multimodal accessibility layers as modifiable elements of the immersive interface rather than as fixed components of the environment.

TESTING WITH DIVERSE USERS AND REPORTING BARRIERS

Although methodologically challenging, testing immersive systems with a variety of users is still essential, see for instance Martínez-Cano (2024). This is crucial in order to understand and address any of the issues mentioned in this work. To achieve this goal, we propose the use of structured evaluation protocols as they can provide useful information even in situations where large-scale trials are not feasible. Within these protocols, qualitative barrier analysis, experiential measurement, and functional validation should all be incorporated in order to obtain a proper testing framework.

It is important that functional validation evaluates whether calibration processes successfully modify ambient layout and whether all essential system interactions can be carried out across available input modalities. Moreover, structured task analysis techniques can help to keep track of completion rates, task duration, error frequency, and assistance required across user profiles.

Even more complicated are the subjective aspects like presence, mental strain, and physical comfort captured by experiential measurement. For this reason, we need to establish standardized metrics that are frequently used and tested in VR research. These are provided by tools like the Simulator Sickness Questionnaire (SSQ), the Slater–Usoh–Steed (SUS) presence scale, and the Igroup Presence Questionnaire (IPQ). Even when functional outcomes seem similar, differences may be found when comparing these measures among diverse participants (Gerling et al., 2025). For example, elevated SSQ scores within particular user groups might be a sign of poorly calibrated locomotion parameters.

Another interesting tool that we mention here is the NASA Task Load Index (NASA-TLX), which gauges perceived mental, physical, and temporal demand. This can be used to assess cognitive workload. Differences between user groups' NASA-TLX scores could be an indication of unseen interaction friction. Measurable strain and articulated difficulty can be triangulated by combining quantitative tools with think-aloud protocols.

We remark that documentation of qualitative barriers is equally important. Structured interviews conducted following immersive sessions might uncover subtle exclusion mechanisms, such as discomfort with gesture precision requirements or trouble locating spatial audio cues. Barrier classification frameworks can be used to categorize problems into motor, visual, auditory, cognitive, or environmental domains. As per Pérez et al. (2023), it is recommended to document not just challenges that have been surmounted but also persistent constraints and the rationale for design compromises.

Finally, we note that variability in mobility, sensory processing, and interaction experience should all also be considered in our testing and valuation framework. Recurring design assumptions that limit inclusion can be found in even small, purposively sampled groups. Further, we underline that enhancing inclusivity are participatory design approaches, where users provide feedback during iterative prototyping instead of post hoc evaluation.

Obviously, we know that accountability is clearly improved when testing procedures are reported in a transparent way. Hence, we propose that reports include details such as participant characteristics, the tools that were used, the calibration parameters that were tested, any challenges that arose, the adjustments that were made, and it is also important to point out any possible limitations that still remain. We believe that when carrying out research in the area of immersive accessibility, adopting consistent reporting practices like these can help refine methodologies progressively over time.

CONCLUSION

In this chapter we have reviewed the issue of accessibility in the context of the new immersive environments. Virtual and augmented reality are emerging as major technologies that will soon become mainstream. For this reason, it is very important to address the accessibility issues that arise within these new

environments. Accessibility considerations often develop more slowly than the technology itself, but it is now time for researchers to focus on these issues to ensure these technologies are inclusive as adoption grows.

In this chapter we have reviewed and analyzed accessibility in immersive interfaces from a general point of view. We argued that calibrating established digital accessibility paradigms is necessary for VR accessibility and inclusive design. Pure representational equivalency is insufficient in immersive systems because they reorganize interaction around embodied spatial participation. We urge for a more mature accessibility architecture for immersive technologies that should be defined by designing for an ability spectrum, incorporating modular remapping strategies, extending tolerance windows, putting adaptive calibration flows into place, incorporating spatially aware sensory supports, and institutionalizing structured testing procedures.

Finally, we conclude that accessibility and technological legitimacy are closely related as VR increasingly influences scientific, industrial, and educational engagement, as already pointed out by numerous researchers. We believe that systems that do not take embodied variability into account will fail to develop inclusive designs and therefore promote exclusion. In contrast, inclusive immersive design enhances emerging technological infrastructures' resilience and flexibility, and emphasizes that different people and groups should have equal opportunities.

REFERENCES

- Bowman, D., & McMahan, R. (2007). Virtual reality: How much immersion is enough? *Computer*, 40(7), 36–43. <https://doi.org/10.1109/MC.2007.257>.
- Bowman, D., Kruijff, E., La Viola, J., & Poupyrev, I. (2004). *3D user interfaces: Theory and practice*. Addison-Wesley.
- Brescia-Zapata, M., Krejtz, K., Duchowski, A. T., Hughes, C. J., & Orero, P. (2025). Subtitles in VR 360° video: Results from an eye-tracking experiment. *Perspectives*, 33(2), 357–379. <https://doi.org/10.1080/0907676X.2023.2268122>.
- Chen, J., Galindo Esparza, R. P., Garaj, V., Kristensson, P. O., & Dudley, J. (2026). EnVisionVR: A scene interpretation tool for visual accessibility in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 32(2), 2007–2019. <https://doi.org/10.1109/tvcg.2025.3617147>.
- Creed, C., Al-Kalbani, M., Theil, A., Sarcar, S., & Williams, I. (2024). Inclusive AR/VR: Accessibility barriers for immersive technologies. *Universal Access in the Information Society*, 23, 59–73. <https://doi.org/10.1007/s10209-023-00969-0>.
- Dudley, J., Yin, L., Garaj, V., & Kristensson, P. O. (2023). Inclusive Immersion: A review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse. *Virtual Reality*, 27, 2989–3020. <https://doi.org/10.1007/s10055-023-00850-8>.
- Gerling, K., Meiners, A.-L., Schumm, L., Rixen, J., Wolf, M., & Yildiz, Z. (2025). An equitable experience? How HCI research conceptualizes accessibility of virtual reality in the context of disability. *ACM Transactions on Accessible Computing*, 18(4), 1–35. <https://doi.org/10.1145/3770755>.
- Hartfill, J., Hajahmadi, S., Schmidt, S., Marfia, G., & Steinicke, F. (2025). Embracing differences in virtual reality: Inclusive user-centered design of bimanual interaction techniques. *Frontiers in Virtual Reality*, 6. <https://doi.org/10.3389/frvir.2025.1586875>.
- Keshavarz, B., Hecht, H., & Lawson, B. D. (2014). Visually induced motion sickness: Causes, characteristics, and countermeasures. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environment: Design, implementation, and applications* (2nd ed., pp. 648–681). CRC Press.
- La Viola, J. (2000). A discussion of cybersickness in virtual environments. *SIGCHI Bulletin*, 32(1), 47–56. <https://doi.org/10.1145/333329.333344>.
- La Viola, J., Kruijff, E., McMahan, R., Bowman, D., & Poupyrev, I. (2017). *3D user interfaces: Theory and practice* (2nd ed.). Addison-Wesley Professional.
- Martínez-Cano, F. (2024). The stigma machine: A study of the prosocial impact of immersive VR narratives on youth in Spain and Canada. *Media and Communication*, 12, 1–19. <https://doi.org/10.17645/mac.8548>.

- Mott, M. E., Cutrell, E., Gonzalez-Franco, M., Holz, C., Ofek, E., Stoakley, R., & Morris, M. R. (2019). Accessible by design: An opportunity for virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Beijing, China (pp. 451–454). doi:10.1109/ISMAR-Adjunct.2019.00122.
- Pérez, D., Ronchi, R., Giroux, A., Brenet, F., Serino, A., Tadi, T., & Blanke, O. (2023) An immersive virtual reality system for ecological assessment of peripersonal and extrapersonal unilateral spatial neglect. *Journal of Neuroengineering and Rehabilitation*, 20(1), 33. <https://doi.org/10.1186/s12984-023-01156-1>.
- Santos Gonçalves, T., Ivars-Nicolás, B., & Martínez-Cano, F. J. (2021). Mobile applications accessibility: An evaluation of the local Portuguese press. *Informatics*, 8(3), 52. <https://doi.org/10.3390/informatics8030052>.
- Santos Gonçalves, T., Simelio Solà, N., & Moreno Sardá, A. (2020). Web accessibility for citizens with reduced capacities on science portals: Evaluation of MCTIC and Ciência Viva. *The Journal of International Communication*, 27(1), 106–125. <https://doi.org/10.1080/13216597.2021.1885463>.
- Szentirmai, A. B. (2024). Universally designed virtual reality: Creating inclusive and immersive learning experiences with ‘VRinDanger’. *Studies in Health Technology and Informatics*, 18, 320, 461–468. <https://doi.org/10.3233/SHTI241042>.
- WHO. 2018. *Disability and health*. <https://www.who.int/en/news-room/fact-sheets/detail/disability-and-health>.
- Wobbrock, J. O., Kane, S. K., Gajos, K., Harada, S., & Froehlich, J. (2011). Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing*, 1–27, <https://doi.org/10.1145/1952383.1952384>
- World Wide Web Consortium. (2018). *Web content accessibility guidelines (WCAG) 2.1*. <https://www.w3.org/TR/WCAG21/>.

PART IV

Heuristics, Performance, and Validation



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Grzegorz Zwoliński

ABSTRACT

The rapid development of immersive technologies challenges traditional usability evaluation methods rooted in screen-based human–computer interaction. While classical heuristic frameworks, such as Nielsen’s principles, have proven effective for graphical user interfaces, they do not sufficiently address the embodied, spatial, and multisensory nature of virtual reality (VR). This chapter introduces *VR Heuristics*, a structured expert-oriented framework designed for the systematic evaluation of immersive experiences. The proposed model extends conventional usability principles by incorporating key dimensions specific to VR, including perceptual consistency, spatial orientation, interaction naturalness, embodiment and agency, comfort and safety, cognitive load, system transparency, and learning support.

To operationalize the framework, we propose a comprehensive evaluation procedure integrating structured checklists, severity rating scales, and a weighted scoring model inspired by NASA-task load index (TLX). This leads to the formulation of the *VR-Heuristic Weighted Index (VR-HWI)*, a standardized 0–100 metric that enables context-sensitive weighting of evaluation criteria while maintaining cross-application comparability. The model supports both expert-based assessment and user-informed validation, bridging qualitative diagnosis with quantitative reporting.

The framework is positioned not as a rigid standard but as an adaptable methodological tool supporting iterative design, interdisciplinary collaboration, and evidence-based improvement of VR applications. By structuring the evaluation of perception, embodiment, and immersive interaction, VR Heuristics contributes to advancing methodological rigor in extended reality research and practice.

INTRODUCTION

Usability heuristics have long played a fundamental role in the design and evaluation of interactive systems by enabling the early identification of problems affecting user experience (UX) quality. Within Human–Computer Interaction (HCI), heuristic evaluation has become a widely adopted practice due to its efficiency, relatively low implementation cost, and suitability for iterative development processes. Classical heuristic frameworks, most notably those proposed by Nielsen (1994), have significantly shaped usability design philosophies across graphical user interfaces (GUIs), web applications, and mobile systems by promoting predictability, consistency, and user control.

However, the emergence of immersive technologies—particularly virtual reality (VR)—challenges several assumptions underlying traditional usability evaluation. Interaction in VR is no longer mediated primarily through indirect input devices such as a mouse or touchscreen. Instead, it unfolds through spatial movement and bodily engagement. In immersive environments, the interface is not merely an

informational layer but becomes the environment in which experience itself takes place. Consequently, design flaws may lead not only to operational inefficiencies but also to spatial disorientation, cognitive overload, or even physiological discomfort.

A defining characteristic of VR is embodied interaction, in which the user's body becomes an integral component of the interaction system. Gestures, posture, gaze direction, and locomotion directly shape the unfolding experience. Designing immersive interfaces therefore requires consideration not only of cognitive processes but also of motor and proprioceptive modalities. Unlike traditional screen-based systems, VR is experienced less as a tool and more as a place for action and participation (Slater, 2009).

This shift is closely related to spatial cognition. Users construct mental representations of virtual environments in ways analogous to physical spaces, relying on innate mechanisms of orientation, spatial memory, and depth perception. Poor environmental layout, insufficient navigational cues, or inconsistent object scaling may disrupt orientation and reduce users' sense of control over their experience. Research in immersive environments has demonstrated that spatial coherence strongly influences presence and task performance (La Viola et al., 2017).

The multisensory nature of VR further complicates the design process. Contemporary immersive systems integrate visual, auditory, and increasingly haptic and kinesthetic feedback channels. Maintaining coherence across sensory modalities is essential for user comfort, whereas inconsistencies may generate perceptual conflicts associated with cybersickness symptoms (Stanney et al., 2021). At the same time, presence and immersion remain central constructs defining the level of psychophysiological engagement experienced within virtual environments (Slater, 2009; Jerald, 2015).

Despite the growing number of Extended Reality (XR) design guidelines and best-practice recommendations, a coherent and systematic heuristic framework for expert evaluation of immersive experiences comparable to established HCI usability methods remains lacking. Existing approaches frequently address isolated aspects of VR interaction, such as locomotion comfort or manual interaction techniques, while overlooking the holistic nature of immersive experience encompassing perception, embodiment, emotion, and cognition.

This chapter addresses this gap through the introduction of VR Heuristics, a structured framework for expert evaluation of immersive applications. The proposed approach extends classical usability thinking by integrating spatial perception, embodied interaction, multisensory consistency, and user comfort into a unified evaluation perspective.

Specifically, this work contributes to:

1. A multidimensional heuristic framework grounded in embodied interaction and spatial cognition theories.
2. A systematic evaluation procedure combining expert analysis with user-informed observation.
3. A weighted quantitative assessment model supporting iterative comparison of VR applications.

The framework aims to support both researchers and practitioners working with XR technologies in systematically diagnosing design issues, improving experiential quality, and developing safer, more intuitive, and engaging immersive environments.

BACKGROUND AND THEORETICAL FOUNDATIONS

The development of heuristic frameworks for VR environments requires grounding both in the traditions of HCI research and in emerging theories of embodied interaction and spatial cognition. VR is not merely an alternative channel for presenting information but represents a fundamentally different mode of engaging with digital systems. Consequently, the design and evaluation of immersive experiences must integrate classical usability models with theories of perception, action, and presence within digital environments.

HCI and Usability Heuristics

Usability heuristics constitute one of the most widely recognized tools developed within HCI research. Their emergence addressed the need for systematic evaluation of interface quality without requiring extensive user testing at every stage of development. Classical usability heuristics—most notably those proposed by Nielsen (1994)—emphasize predictable system behavior, interface consistency, error prevention, and clear communication between users and systems.

These principles proved highly effective in screen-based environments because they aligned with dominant interaction paradigms grounded in symbolic manipulation. Users interpreted graphical representations, made cognitive decisions, and executed actions through indirect input devices such as keyboards and pointing devices. However, as interactive systems evolved toward increasingly complex and spatially situated environments, limitations of heuristics rooted in two-dimensional interface metaphors became increasingly apparent.

Heuristic evaluation functions as an expert-based diagnostic method in which specialists analyze systems using structured principles to identify potential usability problems, ambiguities, or risks of user error. One of its primary advantages lies in applicability during early development stages, including prototyping and initial design iterations. This makes heuristic evaluation particularly valuable in research and development contexts where rapid iteration is essential.

It is important to distinguish usability from UX, although the two concepts remain closely related. Usability primarily concerns efficiency, effectiveness, and task performance, whereas UX encompasses broader experiential dimensions including emotional engagement, satisfaction, and perceived meaning of interaction (Hassenzahl, 2010). This distinction becomes especially relevant in immersive environments. In VR, successful experiences do not necessarily minimize cognitive effort; instead, they often deliberately modulate effort to sustain engagement, exploration, or presence.

XR and Embodied Interaction

XR technologies—including VR, Augmented Reality, and Mixed Reality—shift interaction from symbolic interfaces toward experiential spaces, fundamentally restructuring the relationship between users and digital systems. In VR, users do not observe an interface externally but act within an environment that responds dynamically to their movements and decisions.

VR therefore functions not only as a medium for information delivery but as a medium of experience in which perception, action, and interpretation form an integrated whole (Slater, 2009).

Central to this transformation is embodied interaction. Interaction no longer consists primarily of selecting options or manipulating icons; instead, users act through bodily movement, gaze direction, posture, and spatial positioning relative to objects and events. Designers consequently move beyond arranging graphical layouts toward constructing environments for action.

This transformation is closely connected to spatial cognition—the cognitive processes involved in orientation, navigation, and mental mapping of environments. Users in VR rely on strategies learned in physical space, remembering locations, interpreting distances and directions, and responding to environmental cues as navigational signals. Spatial inconsistencies, insufficient landmarks, or ambiguous environmental structure may lead to confusion or increased cognitive load even when systems remain technically functional (La Viola et al., 2017).

The concept of affordances also plays a critical role. Originally introduced by Gibson and later adapted to design theory by Norman (2013), affordances describe action possibilities suggested by the properties of objects and environments. In three-dimensional immersive environments, affordances extend beyond visual representations such as buttons or icons. Object shape, scale, placement, animation, and physical behavior communicate whether objects can be grasped, moved, or manipulated. Properly designed affordances support exploration-driven learning, reducing reliance on explicit instruction.

Challenges Unique to VR Systems

Despite their significant potential, immersive systems introduce challenges rarely encountered in traditional interfaces. One of the most extensively studied issues is cybersickness—a collection of adverse symptoms including nausea, dizziness, and disorientation resulting from conflicts between visual motion cues and vestibular system signals (Stanney et al., 1997). Even minor latency or inconsistencies between expected and perceived motion may significantly reduce user comfort and shorten session duration.

Closely related is perception mismatch, occurring when user expectations based on physical-world experience conflict with virtual environment behavior. Incorrect object scaling, unnatural acceleration, or insufficient sensory feedback may reduce plausibility and weaken the sense of presence (Slater, 2009).

Locomotion represents another major design challenge. Physical space limitations frequently require teleportation techniques, artificial movement, or perspective manipulation. While effective in extending navigable environments, such approaches may disrupt spatial orientation or induce physiological discomfort if not carefully designed (La Viola et al., 2017).

VR design must also account for proprioception—the body’s ability to perceive its own position and movement. Discrepancies between physical motion and avatar representation may reduce perceived control and realism. Closely related concepts include agency and body ownership, referring respectively to the sense of influencing events and identifying with a virtual body representation. Accurate motion mapping and coherent multisensory feedback strengthen engagement and embodiment, whereas inconsistencies may create cognitive distance between the user and the experience (Kilteni et al., 2012).

METHODOLOGICAL APPROACH TO DEVELOPING VR HEURISTICS

Research Design and Conceptual Foundations

The development of VR Heuristics followed a research-driven design methodology integrating theoretical grounding, empirical observation, and iterative design validation. Rather than deriving heuristics solely from existing usability frameworks, the process aimed to capture the experiential specificity of immersive environments, where perception, embodiment, and spatial interaction fundamentally reshape HCI.

Classical heuristic evaluation methods emerged within screen-based interaction paradigms characterized by indirect manipulation and symbolic interfaces. However, immersive environments introduce embodied interaction, spatial cognition, and multisensory integration as primary interaction mechanisms. Consequently, the methodological objective was not to extend existing usability heuristics incrementally but to redefine evaluation criteria around the experiential structure of VR.

The framework development was guided by three complementary perspectives:

1. Human-centered interaction theory, emphasizing cognition, perception, and usability evaluation traditions within HCI.
2. Embodied cognition and spatial interaction research, addressing the role of bodily movement and environmental perception in immersive systems.
3. Practice-based XR design knowledge, derived from iterative development of real-world immersive applications.

This triangulation ensured that the resulting heuristics reflect both theoretical rigor and practical applicability.

Evidence Sources and Knowledge Integration

The framework emerged through the synthesis of multiple evidence sources rather than a single empirical dataset, reflecting the interdisciplinary nature of immersive technology research, where design knowledge often develops through iterative practice. A structured review of XR and HCI literature informed the identification of recurring design challenges related to presence, locomotion, sensory consistency, and cognitive load, with particular attention to cybersickness, spatial orientation, and embodied interaction as recurring determinants of immersive experience quality. Additional insights were derived from the analysis of immersive applications developed in educational, training, and experimental contexts, including expert walkthroughs, pilot user sessions, and iterative design reviews conducted during development cycles. Rather than treating observations as isolated findings, recurring user behaviors such as hesitation during interaction, repeated attempts to manipulate objects, spatial disorientation, and reliance on environmental cues were analyzed as patterns informing heuristic formulation. The process also incorporated expert reflection by XR designers and researchers involved in immersive system development, which supported identification of micro-friction moments—subtle disruptions in interaction flow that do not necessarily cause task failure but significantly influence experiential quality.

Iterative Heuristic Construction Process

The formulation of VR Heuristics followed an iterative four-stage process. First, recurring design challenges observed across immersive projects were identified and categorized, with emphasis placed not only on task completion errors but also on experiential disruptions affecting presence, comfort, and agency, such as inconsistent object scaling, ambiguous affordances, and loss of spatial reference during locomotion. Next, these problems were grouped into conceptual domains representing dimensions of immersive experience rather than interface components, resulting in categories related to perception, orientation, interaction, embodiment, comfort, cognition, system transparency, and learning support. This clustering reduced the risk of overfitting heuristics to specific interaction techniques or hardware platforms. In the third stage, each domain was translated into evaluation questions designed for observable assessment during expert walkthroughs, following principles of observability, contextual adaptability, and independence from specific hardware implementations. Finally, the heuristics were repeatedly applied during internal evaluations of immersive prototypes, and feedback from these applications informed refinement of wording, scope, and category overlap, with particular attention to maintaining low cognitive overhead for evaluators.

Structure of the Heuristic Framework

The final framework organizes evaluation across eight experiential dimensions: perceptual consistency, spatial orientation and navigation, interaction naturalness, embodiment and agency, comfort and safety, cognitive load and information design, feedback and system transparency, and learning and adaptation support. Together, these dimensions describe immersive experience as a relationship between the user's body, perception, and environment rather than as a sequence of interface operations. The framework does not assume equal relevance of all domains across applications; instead, it allows contextual weighting depending on application goals, user populations, and specific risk profiles.

Reproducibility and Adaptability

A central methodological objective was reproducibility. VR Heuristics therefore defines not only heuristic categories but also evaluation procedures, checklist construction principles, and weighting mechanisms.

The framework is intentionally adaptable. Evaluators may select subsets of heuristics or adjust weighting according to context, provided that the rationale for modification is documented.

This flexibility acknowledges the diversity of immersive systems while maintaining methodological transparency.

Methodological Contribution

VR Heuristics proposes a shift from interface-centered evaluation toward experience-centered assessment in immersive environments. By integrating theoretical foundations, reflective practice, and operational evaluation procedures, the framework provides a structured yet adaptable methodology for diagnosing experiential quality in VR applications.

Rather than replacing empirical user studies, the framework supports earlier identification of design risks and complements user-centered evaluation throughout iterative development cycles.

VR HEURISTICS

As an expert evaluation tool, the VR Heuristics framework was developed to address the specific characteristics of immersive environments, including spatial perception, embodied interaction, and multisensory experience. While traditional usability heuristics primarily emphasize task efficiency and interface clarity, the proposed framework conceptualizes VR experience as a constellation of relationships between the user's body, the digital environment, and cognitive as well as emotional processes.

The heuristics are organized into eight domains representing key dimensions of immersive experience: perceptual consistency, spatial orientation and navigation, interaction naturalness, embodiment and agency, comfort and safety, cognitive load and information design, system transparency and feedback, and learning and adaptation support. These domains reflect recurring design challenges observed across VR applications regardless of context, including entertainment, education, therapy, rehabilitation, and professional training.

Perceptual Consistency

Perceptual consistency constitutes a fundamental requirement for maintaining the credibility and plausibility of immersive environments. The virtual world should exhibit coherent physical logic, perceptually appropriate scale, and predictable system responses. For example, environmental objects should behave according to expected spatial constraints (see Figure 13.1). Doors or solid elements should not intersect the user's body without justification.

Even minor visual or sensory inconsistencies may reduce presence and increase cognitive effort. Visual consistency includes image stability, coherent lighting behavior, and logical object motion relative to environmental conditions and user movement.

Spatial audio also contributes significantly to perceptual coherence. Sound should reinforce the perceived location of events rather than function as an isolated communication channel disconnected from spatial context.

Correct object scale and proportion are equally critical. Inconsistent scaling may disrupt depth perception and undermine spatial credibility. Because users experience VR through embodied perception, even subtle deviations from natural scale may negatively affect realism.

Latency—defined as the delay between user movement and system response—represents another key factor. Increased latency directly affects physiological comfort and may contribute to sensory conflict. Therefore, VR design should prioritize minimizing latency and maintaining rendering stability, particularly during dynamic interaction.



FIGURE 13.1 Example of perceptual inconsistency introducing cognitive disruption: conflicting object behavior, unstable spatial cues, or delayed system response may weaken presence and increase interpretative effort in immersive environments.



FIGURE 13.2 Example of spatially coherent environmental design supporting orientation: recognizable landmarks, directional structure, and stable spatial cues help users construct mental maps and maintain navigational continuity.

Spatial Orientation and Navigation

Spatial orientation is one of the core components of immersive experience because users interpret virtual environments using mechanisms similar to those employed in physical reality (see Figure 13.2). Design should support the construction of mental maps through recognizable landmarks and coherent environmental organization.

Environmental cues such as lighting gradients, contrast, distinctive objects, or architectural directionality help users understand their location and movement direction. The absence of such cues may lead to confusion, particularly in exploratory scenarios.

Users should maintain orientation throughout interaction. Sudden perspective shifts, teleportation without contextual feedback, or ambiguous scene transitions may disrupt experiential continuity. This heuristic emphasizes perceptual fluency supported by stable transitions and logically consistent spatial relationships.

Interaction Naturalness

Interaction naturalness refers to the degree to which user behavior aligns with intuitive expectations derived from physical-world experience. VR interaction should minimize reliance on abstract control schemes requiring explicit learning.

Gestures should correspond to task context and human biomechanics. Complex or unintuitive motion sequences increase both physical fatigue and cognitive burden.

Object manipulation should reflect familiar physical properties such as resistance, weight implication, or constrained movement. Equally important is clear system communication regarding interaction outcomes—whether an action succeeded, failed, or is unavailable.

Effective feedback may combine controller vibration, visual highlighting, animation, or spatial sound. Multisensory confirmation operating across visual, auditory, and tactile channels improves comprehension and accessibility.

Even simplified physical simulations enhance perceived control over the environment. Affordances—properties suggesting possible actions—play a crucial role. In three-dimensional environments, affordances emerge through object shape, texture, depth, animation, and placement. Well-designed affordances reduce the need for explicit instruction and encourage exploratory learning (see Figure 13.3).

Embodiment and Agency

Embodiment represents a central component of immersion in VR. Regardless of visual realism, body representation should remain consistent with user movement and proprioceptive expectations (see Figure 13.4).

Motion feedback—including gesture synchronization and environmental responsiveness—strengthens the sense of agency. Even minor delays or missing reactions may erode confidence in the user's ability to influence events.



FIGURE 13.3 Real-world interaction metaphors in virtual reality.



FIGURE 13.4 Illustration of disrupted embodiment caused by inconsistencies between user movement and virtual body representation.

Body ownership, understood as identification with a virtual body representation, emerges from sensory coherence and predictable system behavior. Accurate mapping of hand movement, gaze direction, and body position increases engagement and emotional credibility of the experience.

Comfort and Safety

User comfort in VR extends beyond traditional interface ergonomics and encompasses physiological as well as psychological factors. A primary design objective is minimizing cybersickness risk through reduction of sensory conflict and stabilization of visual motion.

Movement speed should align with natural perceptual expectations. Excessive acceleration, abrupt directional changes, unstable horizons, or lack of locomotion control may produce discomfort.

Ergonomic considerations also include session duration, required motion range, and accessibility for users with diverse physical abilities. Designers must consider muscle fatigue, real-world spatial constraints, and physical safety during interaction (see Figure 13.5).

Cognitive Load and Information Design

Immersive environments inherently present high stimulus density, increasing the risk of cognitive overload (see Figure 13.6). Effective design requires selective information presentation and reduction of competing attention demands.

VR rapidly engages vision, balance, and attentional resources. Consequently, interfaces should favor diegetic solutions and spatial context over text-heavy overlays whenever possible.

Attention may be guided using lighting, sound cues, motion, or narrative signals directing perception toward key events. Unlike traditional two-dimensional interfaces, designers cannot fully control the user's field of view. Attention must therefore be guided subtly rather than enforced.

Narrative structure further supports temporal and spatial coherence. Storytelling helps users interpret actions and construct meaning without extensive instruction.

Feedback and System Transparency

System transparency refers to the predictability of system behavior and clear communication of interaction consequences. Users should understand why the environment responds in a particular way.



FIGURE 13.5 Visual illustration of cybersickness risk during unstable virtual movement.



FIGURE 13.6 Cognitive overload caused by excessive simultaneous stimuli in a virtual environment.

Multisensory feedback—visual, auditory, and haptic—strengthens causal relationships between action and outcome. Without adequate feedback loops, uncertainty and loss of control may occur.

Predictability does not imply determinism. Systems may intentionally surprise users, particularly in educational or narrative scenarios. However, unexpected events should remain consistent with the internal logic of the virtual world (see Figure 13.7).

Learning and Adaptation Support

The final heuristic domain concerns facilitating user learning and adaptation within immersive environments. This aspect is particularly important in educational, therapeutic, and rehabilitation contexts where VR serves as a medium for skill acquisition or behavioral change.

Design should introduce interaction mechanics progressively, enabling users to acquire competence through action rather than instruction alone. Adaptive support may include dynamic adjustment of difficulty, pacing, or stimulus intensity.

Personalization and progress tracking become especially critical in rehabilitation scenarios. Integration of behavioral or biometric data additionally enables responsive environments capable of adapting to the user's real-time condition.



FIGURE 13.7 User disorientation caused by insufficient system transparency and unclear feedback.

EVALUATION PROCEDURE USING VR HEURISTICS

The VR Heuristics framework was designed to support systematic evaluation of immersive experience quality across different stages of development, including concept design, prototyping, and validation of final VR applications. The procedure assumes that expert judgment and UX are complementary: experts can diagnose design risks rooted in system structure and interaction mechanics, while users reveal practical breakdowns emerging during real use.

Unlike traditional evaluations focused primarily on task efficiency and error rates, VR Heuristics explicitly account for the multidimensional nature of immersive experience, including sensory perception, spatial orientation, physiological comfort, and cognitive load. The evaluation process is therefore treated as an iterative component of design, enabling continuous refinement of VR environments.

Application of Heuristics: Experts and Users

The primary mode of applying VR Heuristics is expert heuristic evaluation conducted by interaction designers, UX researchers, or XR specialists with knowledge of perception and immersive interaction. Experts analyze a VR experience across heuristic domains to identify potential sources of perceptual mismatch, spatial disorientation, interaction ambiguity, or excessive cognitive load.

A key advantage of expert-based evaluation is that it can be performed before broad user deployment, reducing the likelihood of issues that may compromise user comfort or safety. This is particularly relevant for experiences involving dynamic locomotion or high-intensity sensory stimulation, where early detection of risk factors (e.g., cybersickness triggers) is ethically and practically important.

At the same time, the framework is intended to be compatible with user-centered studies. In this mode, heuristics do not need to be presented explicitly to non-expert participants. Instead, they function as an analytic structure for research observation and post-session interviews, enabling researchers to interpret user feedback and behavior through the framework's categories. Combining expert and user perspectives supports the identification of issues that may be underestimated by experts, such as idiosyncratic spatial strategies, differences in sensory tolerance, or early "micro-friction" moments during interaction.

This approach follows the broader tradition of heuristic evaluation as a rapid diagnostic method that complements (rather than replaces) user studies and iterative design cycles (e.g., Nielsen & Molich, 1990).

Checklists and Evaluation Instruments

In practice, VR Heuristics can be operationalized using structured evaluation checklists mapped to the framework's domains. Each heuristic area is represented through a set of evaluation questions that guide the evaluator through the immersive experience in a systematic and repeatable way.

Checklists serve two key functions:

1. Reducing subjectivity by encouraging consistent attention to the same experiential dimensions across evaluators.
2. Supporting reproducibility by providing a shared structure that enables comparisons across iterations, evaluators, or alternative design solutions.

Example questions may address perceived consistency of object scale, clarity of navigation cues, predictability of system response, adequacy of multisensory feedback, or the user's ability to infer available actions from object affordances.

Checklists are particularly useful in multidisciplinary teams where designers, developers, and researchers conduct parallel evaluations. A shared evaluation structure improves communication of issues and facilitates prioritization of design changes.

Rating Scales and Severity Assessment

To enable comparison across iterative design cycles, identified issues are rated using a severity assessment scale grounded in established heuristic evaluation practices, such as Nielsen's severity model (Nielsen, 1994), while extending interpretation criteria to VR-specific concerns. Severity assessment may consider frequency of occurrence, impact on user comfort and continuity of experience, potential physiological risk such as cybersickness, and influence on achieving the intended experience goal, including learning outcomes, training objectives, or therapeutic engagement. Problems associated with cybersickness risk or loss of spatial orientation should be treated as high priority because they may require immediate design intervention. Severity scales also support partial quantification of findings, for example, when reporting progress across development milestones or comparing alternative interaction techniques.

Example Workflow of VR Experience Evaluation

A typical VR Heuristics evaluation workflow consists of several iterative stages.

- **Step 1:** Define the evaluation scenario. The team specifies the goal of the experience, key user interactions, and expected user paths. Potential risk areas are identified, such as dynamic locomotion, time-critical tasks, or intensive object manipulation.

- **Step 2:** Conduct expert walkthroughs. Experts perform individual exploration sessions using the heuristic checklist. Multiple walkthroughs are recommended to capture different exploration strategies and to account for the fact that users frequently deviate from anticipated paths.
- **Step 3:** Aggregate and synthesize issues. Identified problems are grouped by heuristic domain and rated by severity. A structured team discussion helps reduce individual evaluator bias and increases interpretive consistency.
- **Step 4:** Optional user-informed validation. The process may be extended with user studies involving behavioral observation, interviews, and, where applicable, telemetry or interaction logs. User findings are compared with expert assessments to confirm, refine, or reinterpret identified issues.
- **Step 5:** Generate recommendations and iterate. Design recommendations are produced, implemented, and re-evaluated. This positions VR Heuristics as a tool for continuous improvement rather than a one-time quality audit.

Example Evaluation Checklist and Severity Scale

To improve repeatability and reduce subjective interpretation, VR Heuristics can be applied through structured checklists and severity rating scales. This approach supports both qualitative diagnosis of immersive experience issues and partial operationalization of results, enabling comparisons across design iterations or alternative interaction solutions.

Evaluation questions are intentionally not purely closed-form. Their goal is to focus expert observation during exploration and to structure subsequent synthesis. When user studies are included, the same checklist can support observational coding and interpretation of qualitative data.

It is recommended that multiple evaluators conduct assessments independently, followed by structured aggregation of findings. This practice strengthens methodological reliability and makes design decisions easier to justify in research reporting (Table 13.1).

The checklist may be extended or adapted depending on the application context, such as educational, therapeutic, or training-oriented scenarios.

VR Heuristics Severity Rating Scale

Identified issues are assessed using a five-level severity rating scale that extends classical heuristic evaluation models by incorporating factors related to physiological comfort and immersive experience quality (Table 13.2).

Example Evaluation Workflow with Severity Aggregation

In practice, evaluation using VR Heuristics may proceed according to the following steps:

1. Individual exploration of the VR experience by three to five experts.
2. Completion of the evaluation checklist during or immediately after the VR session.
3. Assignment of a severity rating to each identified issue.
4. Aggregation of results by calculating the mean or median severity scores.
5. Prioritization of design changes based on their impact on comfort, safety, and immersion.

TABLE 13.1 Example evaluation questions supporting VR Heuristics expert assessment

HEURISTIC AREA	EVALUATION QUESTION	OBSERVATION FOCUS	EXAMPLE INDICATORS
Perceptual consistency	Does head movement remain fully synchronized with the visual response?	Image stability	Latency, jitter
Perceptual consistency	Does object scale correspond to natural proportions?	Depth perception	Objects appearing excessively large or small
Spatial orientation	Can the user easily recognize direction and target location?	Environmental cues	Wandering behavior, aimless turning
Spatial orientation	Do scene transitions preserve spatial continuity?	Disorientation	Loss of contextual awareness
Interaction naturalness	Are gestures biomechanically intuitive?	Motion fatigue	Unnatural hand movements
Interaction naturalness	Do objects clearly suggest how they should be interacted with?	Affordances	User attempts incorrect actions
Embodiment and agency	Do avatar movements accurately match user movements?	Motion synchronization	Delayed or drifting hands
Embodiment and agency	Does the user feel in control of events?	Sense of agency	Repeated attempts to trigger actions
Comfort and safety	Does locomotion cause discomfort?	Cybersickness risk	Session interruption or avoidance behavior
Cognitive load	Does the scene contain excessive stimuli competing for attention?	Cognitive overload	Ignoring key elements
Feedback and transparency	Do user actions produce an immediate system response?	Feedback clarity	Lack of response or delayed reaction
Learning support	Can users understand new mechanics without textual instructions?	Adaptation and onboarding	Reliance on researcher assistance

TABLE 13.2 VR Heuristics severity rating scale

SEVERITY LEVEL	DESCRIPTION	IMPACT ON EXPERIENCE	RECOMMENDED ACTION
0 No issue	No observable problem	No impact	No action required
1 Cosmetic	Minor perceptual or visual inconsistency	Minimal impact	Optional refinement
2 Minor	Noticeable obstacle affecting interaction or orientation	Occasional frustration	Improvement recommended
3 Major	Significant disruption of task completion or immersion	Disorientation or reduced sense of agency	High-priority correction
4 Critical	Risk of cybersickness or user safety compromise	Experience interruption	Immediate corrective action required

This approach enables not only identification of qualitative design issues but also systematic monitoring of project progress across iterative development cycles.

VR-Heuristic Weighted Index: Concise Formalization

To operationalize context-sensitive expert evaluation, we define the VR-Heuristic Weighted Index (VR-HWI) as a weighted aggregation of normalized quality scores derived from heuristic assessment.

Let

$$Q = \{1, \dots, n\}$$

be the set of selected evaluation questions.

For each question i :

r_i —severity rating (e.g., 0–4 scale),

w_i —normalized weight, such that $\sum_{i=1}^n w_i = 1$.

Quality normalization

Severity ratings are rescaled to a quality score in the range [0,100]:

$$s_i = 100 \left(1 - \frac{r_i}{R} \right)$$

where R is the maximum severity level (e.g., $R=4$).

Thus:

$$r_i = 0 \Rightarrow s_i = 100 \text{ (no issue)}$$

$$r_i = R \Rightarrow s_i = 0 \text{ (critical issue)}$$

Weight derivation (pairwise comparison)

Weights are derived using a NASA-task load index (TLX)-inspired pairwise comparison procedure (Hart & Staveland, 1988).

For each pair of questions (i, j) , evaluators assign a relative importance score $a_{ij} \in [0,10]$, with the complementarity constraint:

$$a_{ij} + a_{ji} = 10$$

Raw weights are computed as

$$t_i = \sum_{j \neq i} a_{ij}$$

and normalized:

$$w_i = \frac{t_i}{\sum_{k=1}^n t_k}$$

This guarantees:

$$\sum_{i=1}^n w_i = 1$$

VR-HWI definition

The final immersion quality index is defined as

$$\text{VR-HWI} = \sum_{i=1}^n w_i s_i$$

Because $s_i \in [0, 100]$ and $\sum w_i = 1$, the resulting index satisfies:

$$\text{VR-HWI} \in [0, 100]$$

Higher values indicate greater overall experiential quality.

Interpretation

VR-HWI integrates expert qualitative diagnosis expressed through heuristic severity, context-sensitive weighting, and standardized quantitative output. The model extends classical heuristic evaluation (Nielsen, 1994) by incorporating weighted aggregation inspired by NASA-TLX workload assessment (Hart & Staveland, 1988), adapted to immersive experience evaluation.

Reliability Considerations

Because the VR-HWI relies on expert judgments in both severity ratings and pairwise comparison scores, assessment of inter-rater reliability is essential. Agreement in quality ratings may be quantified using Cohen's kappa, Fleiss' kappa, or the Intraclass Correlation Coefficient (ICC), depending on the number of evaluators and rating scale assumptions. For weights derived from pairwise comparisons, reliability may be examined through Spearman's rank correlation between individual weight vectors, analysis of variance relative to aggregated mean weights, or optionally ICC calculated on raw weight scores. High agreement strengthens the credibility of VR-HWI as a synthetic evaluation index, whereas substantial disagreement indicates the need for calibration sessions in which evaluators jointly review interpretation criteria for selected heuristics. VR-HWI does not eliminate the inherent subjectivity of expert assessment, but structures and operationalizes it, while reliability analysis helps determine whether results reflect stable system evaluation rather than individual evaluator preferences. Internal consistency may additionally be assessed using Cronbach's alpha (Cronbach, 1951), particularly when multiple evaluation questions are grouped within a single heuristic domain. Values of $\alpha \geq 0.70$ are generally acceptable in exploratory research, whereas $\alpha \geq 0.80$ indicates stronger consistency (Nunnally & Bernstein, 1994). Low alpha values may suggest the need to revise item wording, reduce redundancy, or reconsider domain structure, thereby strengthening the methodological robustness of VR-HWI.

CONCLUSION

The rapid evolution of immersive technologies requires evaluation methods capable of addressing experiences that extend beyond traditional interface interaction. This chapter introduced VR Heuristics as a structured framework for expert assessment of immersive environments, integrating spatial perception, embodied interaction, multisensory feedback, and user comfort within a unified evaluation perspective. The framework is intended not as a rigid normative standard, but as an adaptable checklist supporting expert reasoning across applications that differ in purpose, user population, and interaction logic.

A key strength of the approach lies in its iterative use within design cycles, enabling early identification of experiential risks and progressive refinement of immersive systems. Evaluation should remain context-sensitive and focus on observable indicators such as hesitation, repeated interaction attempts, or loss of spatial orientation rather than general impressions alone. Combining expert assessment with user-informed observation further strengthens the identification of subtle interaction problems.

Several limitations remain important. Expert evaluation retains a degree of subjectivity, and outcomes may vary depending on evaluator background and hardware characteristics such as tracking precision, field of view, or latency. For this reason, transparent reporting and reliability analysis remain essential.

Compared with classical usability methods, VR Heuristics extend evaluation toward experiential dimensions specific to immersive systems, including comfort, agency, body ownership, and multisensory coherence. Rather than replacing usability testing, the framework complements it by supporting early-stage diagnosis of immersive design risks.

Future development may extend the framework toward mixed reality, adaptive systems, and artificial intelligence (AI)-supported evaluation. In summary, VR Heuristics and VR-HWI provide an adaptable methodological basis for more transparent, iterative, and human-centered design of virtual environments.

REFERENCES

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in psychology* (Vol. 52, pp. 139–183). North-holland.
- Hassenzahl, M. (2010). *Experience design: Technology for all the right reasons* (Vol. 8). Morgan & Claypool Publishers.
- Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.
- Kiltner, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4), 373–387.
- La Viola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: Theory and practice*. Addison-Wesley Professional.
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Nielsen, J., & Molich, R. (1990, March). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seattle, WA, United States (pp. 249–256).
- Norman, D. (2013). *The design of everyday things*. MIT Press.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). MacGraw-Hill.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549–3557.
- Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997, October). Cybersickness is not simulator sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41(2), 1138–1142.
- Stanney, K. M., Lawson, B. D., & Oman, C. M. (Eds.). (2021). *Cybersickness in virtual reality versus augmented reality*. Frontiers Media SA.

Performance and Optimization

14

Grzegorz Zwoliński

ABSTRACT

Performance in virtual reality is not solely a technical parameter but a fundamental determinant of experiential quality. Unlike screen-based systems, immersive environments directly engage the user's perception, bodily movement, and spatial cognition, making frame stability, latency, and tracking accuracy critical conditions for presence and comfort. This chapter examines performance and optimization from a design-oriented perspective, positioning real-time constraints as an integral component of immersive experience rather than a post-production concern.

This chapter introduces frame timing and computational budgets as key conceptual tools for understanding real-time systems, explaining how CPU–GPU balance, rendering strategies, and asset management influence perceptual continuity. Rendering optimization is discussed through the lens of perceptual efficiency, including lighting design, levels of detail, and precomputation as methods for achieving visual credibility within limited computational resources. Special attention is given to memory management and streaming workflows, emphasizing how spatial design and narrative pacing can support efficient data loading without disrupting immersion.

Latency and tracking stability are analyzed in relation to embodiment and agency, demonstrating how temporal inconsistencies can undermine the user's sense of control and physical integration within virtual space. This chapter further presents profiling workflows and quality assurance gates as iterative practices that align technical evaluation with experiential outcomes.

Finally, optimization is framed as a matter of responsible design. Stable performance contributes not only to comfort and safety but also to accessibility and sustainable technological development. By integrating perceptual psychology, system constraints, and ethical considerations, this chapter argues that optimization should be understood as a core design competency essential for creating reliable, inclusive, and meaningful immersive environments.

INTRODUCTION

In virtual reality (VR), performance cannot be reduced to frame rate or rendering efficiency alone. Unlike screen-based systems, immersive environments depend on continuous synchronization between visual output, bodily movement, and spatial perception, making system stability a prerequisite for perceptual coherence.

Because users experience VR from within the environment rather than observing it externally, delays or irregular frame timing directly affect presence, comfort, and control. Drops in frame rate or increased latency may introduce sensory conflict, weaken embodiment, and in some cases induce cybersickness (Slater, 2009; LaViola et al., 2017). Performance therefore affects not only visual quality but also physiological safety.

In immersive systems, every design decision generates computational cost. Spatial complexity, lighting, interaction logic, and motion all influence whether the system can maintain real-time stability. Performance should therefore be understood as an integral part of experience design rather than a final technical adjustment.

This chapter examines performance as a relationship between perceptual requirements and computational constraints. It discusses frame budgets, CPU–GPU balance, rendering efficiency, memory management, latency, and profiling as elements of responsible immersive design.

Performance as Experience Quality

In immersive environments, system performance is not secondary to experience design—it is its precondition. Unlike traditional screen-based interfaces, where performance issues may be perceived as aesthetic inconveniences, in VR any temporal instability directly affects perception, physiology, and the user’s relationship with virtual space.

To understand the significance of performance in VR design, its impact must be examined across three key experiential dimensions: presence, comfort, and agency. These constructs have been discussed in previous chapters; however, in the context of performance they acquire a systemic character—they depend on the stability of real-time system operation.

Presence—perceptual continuity as the foundation of immersion

Presence in VR is grounded in the illusion of being located within an alternative space that responds according to the user’s perceptual expectations. As Slater (2009) argues, both place illusion and plausibility illusion depend on the coherence of environmental responses to user actions. This means that the system must deliver stable, predictable, and immediate responses to head and body movements.

Performance influences presence primarily through temporal stability. Irregular frame intervals, drops in frame rate, or delays in image updates introduce micro-disruptions in perceptual continuity. Although users may not consciously identify these as technical issues, the brain detects inconsistencies between expected and received visual signals. As a result, the environment loses credibility.

Neurocognitive research suggests that the sense of “being in” an environment emerges from multi-sensory integration and predictive mechanisms of the brain (Sanchez-Vives & Slater, 2005). A real-time system must therefore provide stimuli consistent with these predictions. If the world reacts with delay or irregularity, predictive processing is disrupted and presence weakens.

For this reason, stability is more important than momentary peaks of visual quality. A consistent frame rhythm and predictable responsiveness create stronger immersion than a visually detailed environment operating with temporal instability.

Comfort—sensory congruence and physiological safety

User comfort in VR is directly linked to system performance. One of the most extensively studied phenomena in this domain is cybersickness—a set of symptoms resulting from sensory conflict between visual input and vestibular signals (LaViola et al., 2017; Stanney et al., 1997).

Latency between head movement and image update (motion-to-photon latency) may produce subtle discrepancies between what the user sees and what their body feels. Even small temporal offsets increase the cognitive load of the nervous system, which attempts to reconcile inconsistent signals. This may lead to disorientation, cognitive strain, or physiological discomfort.

Irregular rendering is equally problematic. Even if the system maintains a high average frame rate, occasional frame drops can cause micro-disruptions in motion continuity. Research on visually induced motion indicates that stability of trajectory and rhythmic consistency is critical for perceptual comfort (Keshavarz et al., 2011). Designers should therefore aim not only for high performance but, more importantly, for performance stability.

Performance thus becomes a matter of experiential safety. A temporally unstable system not only diminishes aesthetic quality but may also shorten session duration and limit accessibility for more sensitive users.

Agency—synchronization as the basis of control

In immersive environments, the user’s body becomes an integral part of the interface. Head movement, hand gestures, and gaze are directly mapped onto actions within the virtual world. The sense of agency depends on the belief that no disruption occurs between intention and outcome.

If system responses are delayed, synchronization between physical motion and digital representation is compromised. Research on embodiment in VR demonstrates that temporal and spatial congruence between the body and its virtual representation is crucial for identification and perceived control (Kilteni et al., 2012). Even minor inconsistencies may undermine trust in the system.

Importantly, agency is sensitive not only to average latency but also to variability. Irregular response timing increases cognitive load, as users begin monitoring system stability rather than focusing on the task. The experience shifts from intuitive to effortful.

In educational, training, or rehabilitation applications, response stability carries additional significance—it directly influences learning and skill acquisition. Here, performance conditions include not only comfort but also effectiveness.

Performance as a dimension of experience quality

Presence, comfort, and agency are interrelated. Instability in one dimension affects the others, producing a cascading effect. A drop in frame rate may weaken presence, increase discomfort, and undermine the sense of control (see Figure 14.1). From a design perspective, this means performance cannot be treated as a separate technical concern.

In VR, performance is an indicator of experiential quality—not because it measures computational power, but because it defines the perceptual conditions under which the user operates. Designers of immersive environments should therefore understand performance as a relational construct—emerging from the interaction between human perception and real-time system behavior.

Real-Time Constraints and Frame Budgets

VR is a real-time medium in the strictest sense of the term. While short delays may be tolerated or masked in many digital systems, in VR every frame must be generated within a precisely defined time interval. This constraint does not arise solely from display technology parameters—it is a consequence of perceptual biology and the necessity of synchronizing movement with visual feedback.

In immersive environments, images are generated in response to the user’s current head position. This means the system cannot prepare a complete animation sequence in advance. Each frame must be computed dynamically based on up-to-date tracking data. If this process takes too long, image updates fall out of synchronization with bodily movement, leading to a loss of perceptual stability.



FIGURE 14.1 Example distribution of the frame time budget within an 11 ms rendering cycle.

Refresh rate and frame time

Display refresh rate determines how frequently the system must deliver a new image frame. In practice, this means that:

- At 90 Hz, a new frame must be generated approximately every 11 ms.
- At 72 Hz, the available time increases to about 13.9 ms.
- At 120 Hz, it decreases to roughly 8.3 ms.

This time interval is known as the frame time budget. It represents the maximum amount of time in which all operations related to updating the world state and generating the image must be completed.

It is essential to understand that this budget is a hard limit rather than an average value. If a single frame is generated too late, the system cannot compensate by producing the next frame faster. In VR, the critical factor is not average frames per second but consistency of frame delivery. Irregular timing disrupts perceptual continuity.

What fits within the frame budget

The frame time budget includes all processes required to generate the image. Although implementation details differ between systems, the general structure is similar and typically includes:

1. Reading and predicting user position (tracking)
2. Updating application logic and physical simulation
3. Processing interaction and animation systems
4. Preparing rendering data
5. Generating stereoscopic images (one for each eye)
6. Synchronizing and presenting the image on the display

Each of these stages consumes part of the available time. If their combined duration exceeds the budget, the system cannot maintain the required refresh rate.

It is important to emphasize that VR requires stereoscopic rendering. This involves generating images from two slightly different viewpoints. Even when certain computations can be shared, the visualization cost remains higher than in single-screen systems.

An additional burden arises from the continuously changing viewpoint. Unlike traditional applications, where the camera may remain relatively stable, even micro-movements of the user's head in VR require recalculating the scene.

Exceeding the budget and its consequences

When the system exceeds the frame time budget, the consequences are immediate: frame delivery becomes unstable, frame intervals become irregular, and latency between movement and visual response increases. From the user's perspective, these are not merely visual imperfections, because temporal irregularity disrupts synchronization between bodily motion and visual updates, directly affecting presence and comfort. Even short-lived budget overruns may weaken experiential stability. Although some systems employ compensatory mechanisms such as motion prediction or frame reconstruction to reduce the visibility of performance issues, these techniques do not eliminate their underlying causes. From a design perspective, maintaining stable performance remains more important than relying on corrective measures after the fact.

Frame budget as a design framework

From a designer's perspective, the frame time budget should not be understood solely as a technical limitation. It constitutes the structural framework within which immersive experience emerges. Every

decision that increases environmental complexity—additional light sources, a greater number of interactive objects, or more elaborate simulations—consumes part of the available time.

VR design therefore requires thinking in terms of temporal resource allocation. This involves consciously balancing visual detail, interaction dynamics, and system stability. Time becomes one of the primary materials of design—equally important as space, light, or sound.

Computational Budgets: CPU–GPU Thinking in VR Design

The frame time budget defines how much time the system has to generate a single update of the virtual world, but time alone does not explain why some environments remain stable while others with similar visual complexity experience performance drops. The key lies in how computational work is distributed between the two primary processing units: the CPU and the GPU. Although architectural details differ across platforms, the general principle remains the same—the CPU manages the logic and organization of the world, while the GPU is responsible for visual representation. For VR designers, the essential point is not technical mastery of these components but understanding that both operate within the same limited frame budget.

The CPU can be understood as the part of the system responsible for world simulation and decision-making. It updates application logic, processes user interactions, runs physics simulations, manages procedural animation, controls artificial intelligence, and handles spatial audio. Every interactive object introduces computational cost: a touch-sensitive button, a physics-enabled object, or a reactive character all require continuous state calculation. This means that interaction complexity may burden the system as heavily as visual detail, and even visually simple scenes may exceed CPU capacity when many dynamic relationships are active. The CPU also prepares data for the GPU, so delays at this stage may leave graphical resources underused while overall performance still declines.

The GPU is responsible for generating the image presented to the user. It processes geometry, calculates lighting and shadows, manages materials and textures, generates visual effects, and prepares the final image for display. In VR, this workload is especially demanding because the scene must be rendered stereoscopically, from two viewpoints. Combined with a wide field of view and close observation distance, this increases the visibility of visual artifacts and often encourages excessive model detail or texture resolution. Yet greater visual complexity does not necessarily improve the experience. Users are generally more sensitive to scene stability and perceptual clarity than to maximum graphical detail, meaning that excessive rendering cost may reduce comfort without delivering clear experiential benefit.

Performance therefore depends on systemic balance rather than on a single component. If the CPU spends too much time simulating interactions, the GPU receives data too late; if the GPU is overloaded by geometry or lighting, the CPU finishes early and waits. In both cases, the frame budget is exceeded. For designers, this means thinking not only about how much detail to include, but also about which part of the system bears the cost of each decision. Additional light sources, autonomous characters, or interaction layers may all compete for the same temporal resources.

Understanding CPU–GPU balance is therefore a design competency rather than a programming skill. Stable immersive environments emerge when visual complexity, simulation, and interaction remain in equilibrium. Optimization then becomes a process of deliberate prioritization—deciding which elements most meaningfully contribute to the experience and where perceptual strategies such as lighting, composition, or attention guidance can replace computational expense.

Rendering and Perceptual Efficiency

Rendering performance in VR is not achieved by maximizing image detail, but by balancing computational cost with perceptual quality. Users do not analyze immersive scenes pixel by pixel; instead, they

interpret them as coherent spaces in which they are situated. For this reason, designers should aim for perceptual efficiency—using available resources so that the environment remains stable, legible, and credible even when technical representation is simplified. Three strategies are particularly important in this context: lighting design, levels of detail (LOD), and precomputation.

Lighting in VR serves both perceptual and computational functions. It shapes atmosphere and directs attention, but it is also one of the most demanding elements of rendering. Dynamic light sources, realistic shadows, and complex reflections can significantly increase system load, although full physical accuracy is rarely necessary to maintain credibility. From the user's perspective, clarity of form, contrast between spatial planes, consistency of light direction, and shadow stability are often more important than precise physical simulation. Carefully designed static or semi-static lighting can therefore produce convincing realism at lower computational cost while also improving perceptual comfort by reducing flicker and rendering instability.

LOD provide another essential strategy for managing scene complexity. Objects farther from the user do not require the same geometric precision as nearby elements, especially in VR where attention remains concentrated in the central visual field and close spatial range. Peripheral and distant objects function mainly as contextual references, so simplified models often remain perceptually credible as long as their silhouettes and proportions stay legible. LOD should therefore be understood not as quality reduction but as adaptation to spatial perception, provided that transitions remain smooth and imperceptible.

A further optimization method is precomputation, which shifts selected calculations outside real-time execution. Elements such as global illumination, shadows of static objects, or reflections on fixed surfaces can be prepared in advance, reducing runtime load and increasing predictability. Dynamic computation should remain reserved for elements that genuinely require change or interaction, while static architecture and scenographic components can rely on precomputed data.

Together, these strategies show that rendering in VR is not a pursuit of maximum detail but a process of managing perception. Lighting guides attention, LOD preserve spatial clarity, and precomputation stabilizes runtime behavior. Perceptual efficiency therefore becomes a design strategy in which computational constraints are treated not as obstacles, but as tools for sustaining presence and comfort.

Assets and Memory Management

The performance of a VR environment depends not only on rendering speed but also on how digital assets are managed. Geometric models, textures, spatial audio, and animations occupy memory and must be available exactly when needed during runtime. If data cannot be delivered fast enough, even an otherwise optimized rendering pipeline may become unstable. Because VR users can freely look in any direction and move dynamically through space, the system must be prepared to reveal different parts of the environment at any moment. Missing or delayed assets may therefore appear as empty objects, temporary texture degradation, or short application stalls, all of which disrupt perceptual continuity and weaken presence.

Memory functions as an invisible design constraint because every scene element carries resource cost, including geometry, textures, lighting maps, animation data, audio resources, and particle systems. Textures are particularly demanding in VR, as close viewing distances often encourage high material resolution. However, multiple large textures quickly consume available memory and may force the system to replace or unload data dynamically during longer sessions, causing sudden visual changes. Asset design should therefore consider not only appearance but also how resources behave over time in memory.

A key strategy for managing this complexity is streaming, which dynamically loads required assets while removing those no longer needed. Instead of keeping the entire environment in memory, the system maintains only the portion relevant to the user's current location or context. This is especially important in large exploratory environments, training applications, and narrative experiences involving movement between areas. Effective streaming remains invisible: transitions feel seamless, and new elements appear naturally as the user moves through space.

Because streaming depends on timing, it also requires intentional spatial design. Corridors, turns, controlled visibility, lighting changes, or narrative pauses allow the system to prepare upcoming content before it becomes visible. In this sense, scenography and narrative pacing directly support optimization.

Excessive memory load often appears not as dramatic failure but as subtle instability, such as delayed textures, temporary blurring, or sudden material changes. Although visually minor, these effects make users notice the system instead of the environment itself. Memory management therefore becomes part of experience design, where careful asset selection, appropriate resolution planning, and effective streaming help maintain stability without sacrificing visual richness.

Latency and Tracking Stability

In immersive environments, latency and tracking stability are among the most critical determinants of experience quality because they directly affect the relationship between bodily movement and spatial perception. Latency is the delay between a user's physical movement and its visual update in the virtual world, including motion capture, system processing, rendering, and display. Even small delays may create discrepancies between visual and proprioceptive signals: the body moves immediately, while the environment responds slightly later. This loss of synchrony is a major source of discomfort and can weaken experiential credibility.

Latency also directly affects embodiment—the sensation that the virtual body or perspective functions as an extension of the user's own body. Embodiment depends on immediate and predictable sensory correspondence. When head movement is followed by delayed environmental response, or when virtual hands fail to match controller motion, users become aware of technological mediation rather than experiencing the space directly. In this sense, the system must behave as if it were a continuation of the body; delay makes that illusion fragile.

Tracking stability is equally important. Even with low latency, unstable positional data may produce micro-vibrations, controller jumps, or temporary tracking loss. Such irregularities disrupt motion continuity, reduce interaction precision, and weaken the credibility of physical relationships within the environment. From a perceptual perspective, smooth predictability is often more tolerable than perfect speed: users adapt more easily to slight delay than to sudden, irregular positional errors.

Although designers do not control hardware parameters directly, they can reduce perceptual consequences through design choices. Useful strategies include avoiding interactions that demand extreme precision, enlarging activation zones, providing clear visual or haptic feedback, and limiting sudden camera or object movements. Objects can also be designed with tolerant interaction logic, such as subtle snapping or larger detection areas, so that minor tracking inaccuracies do not interrupt the experience.

Latency and tracking stability are therefore not merely technical parameters but conditions of presence, agency, and comfort. When synchrony between body and environment is preserved, users experience the world as responsive and credible; when it fails, even milliseconds become perceptually significant because they define the boundary between embodiment and technological awareness.

Profiling Workflow and Quality Assurance

Performance optimization in VR should not be treated as a final development stage or a reaction to technical problems, because performance is itself a component of experiential quality. For this reason, profiling must accompany design throughout the entire development process and function as a design instrument rather than merely a technical procedure. Profiling means analyzing how the system uses available resources during runtime in order to understand which design decisions affect experiential stability, not only which technical processes are slowest.

An effective profiling workflow is iterative and based on realistic test scenarios. Evaluation should reflect actual use conditions rather than simplified scenes, including moments of peak load such as locomotion, multi-object interaction, or transitions between environments. Key indicators include frame time, frame rate stability, CPU and GPU utilization, memory consumption, and the frequency of data-loading operations. More important than isolated values is their consistency over time, since short performance spikes are often more disruptive than constant, predictable load.

When instability appears, the next step is identifying its cause. Performance drops may result from visual complexity, excessive interaction logic, or inefficient asset handling, and without contextual analysis it is easy to misidentify the source. Any adjustment should therefore be followed by repeated measurement, because profiling is inherently cyclical and aims at gradual convergence toward stable performance.

Within this process, quality assurance gates are especially valuable. At defined production stages, the application should meet clear performance criteria before further development continues. Such criteria may include stable frame timing across full scenarios, uninterrupted streaming, reliable interaction tracking, and the absence of perceptible visual irregularities affecting orientation. This prevents technical debt from accumulating as new features are introduced.

Ultimately, profiling is not only about eliminating errors but about building awareness of how design decisions influence presence, comfort, and embodiment. In VR, optimization does not limit creativity; it protects the conditions that make immersive experience possible.

Optimization as Responsible Design

Optimization in VR is not only a technical concern but also a matter of design responsibility, because every performance-related decision directly affects comfort, safety, and accessibility. In immersive environments the system is no longer separated from the user by a screen; it operates in direct relation to the body, perception, and sense of balance. As a result, image instability, latency, and sensory overload are not merely technical imperfections—they may produce discomfort, disorientation, and even premature termination of the session. From this perspective, optimization becomes part of responsible design practice.

Stable frame rates, reliable tracking, and predictable responsiveness should therefore be treated as baseline conditions of immersive quality rather than technological advantages. When performance is insufficient, the risk of cybersickness, loss of orientation, cognitive fatigue, and reduced control increases. In this sense, performance decisions influence not only system behavior but also the physiological and psychological state of the user.

Optimization is equally important for accessibility. VR applications are experienced by users with different perceptual tolerances, motor abilities, and hardware conditions. Unstable performance may be especially problematic for those sensitive to visual stimuli, affected by balance limitations, or using less powerful devices. For this reason, reducing excessive visual effects, stabilizing lighting, and limiting abrupt motion should be understood not as aesthetic compromise but as inclusive design.

A further dimension concerns sustainability. Resource-efficient applications consume less energy and may extend hardware lifespan, which becomes increasingly relevant as VR expands across education, health-care, training, and entertainment. Optimization therefore also supports more sustainable technological ecosystems.

Ultimately, performance in VR is not a final-stage technical adjustment but an integral part of experience design from the earliest decisions. Designers who understand the relationship between computational limits, perception, and embodiment create environments that are more stable, safer, and more accessible. In this sense, optimization does not restrict creativity—it enables responsible immersive design.

CONCLUSION

Designing for VR requires moving beyond the traditional understanding of performance as a purely technical concern. In immersive environments, performance becomes a fundamental condition of experience quality, as it directly influences spatial perception, physiological comfort, and the user's sense of agency.

Unlike screen-based systems, where temporary drops in smoothness may be perceived merely as inconvenience, temporal instability in VR disrupts the relationship between the body and the digital world. Frame rate, latency, and tracking stability determine the credibility of movement and the continuity of presence. Milliseconds therefore become a design material as significant as lighting, spatial composition, or narrative structure.

This chapter has demonstrated that real-time constraints should not be viewed as obstacles to creativity. CPU and GPU budgets, asset management strategies, and rendering techniques are not solely mechanisms for reducing computational cost. Rather, they are tools for consciously shaping perceptual experience. Carefully designed lighting, appropriate LOD, and selective precomputation allow designers to achieve a high degree of environmental plausibility without imposing unnecessary system load.

Systemic thinking thus becomes a key competence for VR designers. Decisions concerning scenography, interaction design, and narrative structure simultaneously influence both user perception and application stability. Optimization is not the removal of quality, but its selective amplification through attention guidance and the informed use of perceptual mechanisms.

Production workflow also plays a crucial role. Continuous profiling, testing within realistic usage scenarios, and the introduction of performance quality gates enable teams to identify risks before they become perceptible to users. In VR, stability is rarely the result of a single technical fix; rather, it emerges from an iterative design process.

Performance also carries ethical and social implications. Stable experiences reduce the risk of physiological discomfort, improve accessibility for users with diverse perceptual capabilities, and enable applications to function across a broader spectrum of hardware configurations. Optimization therefore becomes an element of responsible design practice—a commitment to user well-being as well as to the sustainable development of immersive technologies.

Ultimately, efficient design does not mean less ambitious design. It means conscious design. VR creators do not design merely images or interfaces, but dynamic systems engaged in continuous dialogue with the human body and perceptual processes. Understanding the constraints of real-time computation allows technology to become a transparent medium of experience—one that disappears from the user's awareness, leaving only the sensation of presence.

REFERENCES

- LaViola Jr J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: Theory and practice*. Addison-Wesley Professional.
- Keshavarz, B., Hecht, H., & Zschuschke, L. (2011). Intra-visual conflict in visually induced motion sickness. *Displays*, 32(4), 181–188.
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332–339.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549–3557.

Usability Testing and Iterative Development in VR—Methods for Conducting Usability Testing in VR

Anna Laska-Leśniewicz and Dorota Kamińska

ABSTRACT

Usability testing is a critical component in the development of virtual reality (VR) applications, where immersive interaction introduces new challenges related to spatial navigation, embodied interaction, and multisensory feedback. This chapter presents methods and metrics commonly used for conducting usability testing in VR environments. Both subjective and objective evaluation approaches are discussed. Subjective methods include standardized questionnaires that capture users' perceptions of usability, workload, user experience, and cybersickness. Objective metrics focus on measurable indicators of interaction, such as task performance, navigation behaviour, and system interaction logs collected during VR sessions. This chapter also highlights the growing role of behavioural and physiological measurements—such as eye-tracking, facial expression analysis, and wearable sensor data—in understanding users' cognitive and emotional responses during immersive interaction. In addition to methodological considerations, this chapter discusses the role of usability testing within iterative development processes. Practical examples from VR development projects illustrate how insights obtained from usability studies support the refinement of interaction design and contribute to the creation of more intuitive, efficient, and engaging immersive applications.

USABILITY TESTING

Usability testing is a fundamental method used to evaluate how effectively, efficiently, and satisfactorily users can interact with a system. Its primary goal is to identify usability issues, understand user behaviour, and provide evidence that supports improvements in interface design. By observing users performing

representative tasks, researchers and developers can detect problems that may not be apparent during the design process. Usability testing therefore plays a crucial role in ensuring that digital systems are intuitive, accessible, and aligned with users' expectations.

Traditional usability evaluation methods have been widely applied to desktop, web, and mobile applications. However, immersive technologies such as virtual reality (VR) introduce new interaction paradigms that differ significantly from conventional graphical user interfaces. In VR environments, users interact through body movements, spatial navigation, head orientation, and controller-based manipulation of virtual objects. These characteristics create unique usability challenges related to spatial orientation, interaction design, sensory feedback, and cognitive load. As a result, usability testing becomes particularly important in VR applications to ensure that users can understand and effectively navigate immersive environments.

In addition to identifying interface problems, usability testing in VR helps developers understand how users perceive environmental cues, guidance mechanisms, and interaction affordances within a three-dimensional space. Because VR systems involve complex combinations of visual, auditory, and sometimes haptic feedback, even small design decisions may significantly influence user behaviour and task performance. Systematic usability evaluation therefore provides valuable insights into how users explore virtual environments, how they respond to different interaction techniques, and which design elements support or hinder efficient task completion.

An important aspect of usability testing, particularly in the development of immersive systems, is its integration within an iterative design process. Instead of evaluating the system only after development is completed, usability testing is typically conducted at multiple stages of the design cycle. Feedback obtained from users during early prototypes allows developers to refine interaction mechanisms, improve guidance cues, and simplify navigation structures before the system reaches its final form. Iterative testing (see Figure 15.1) and refinement significantly reduce the risk of usability problems and contribute to the development of more intuitive and effective VR applications.

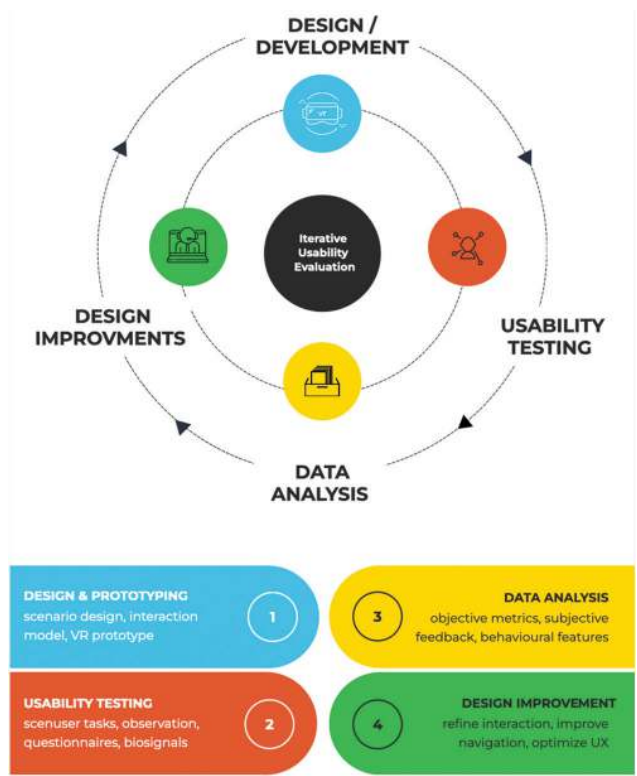


FIGURE 15.1 Iterative usability evaluation as a key element of the iterative design process.

For these reasons, usability testing is widely recognized as a critical component of VR development, supporting both the identification of design issues and the continuous improvement of immersive experiences. In this chapter, we present a range of usability evaluation methods applicable to VR environments, including subjective measures based on user feedback, objective performance and behavioural metrics derived from system data, as well as affective and physiological indicators that provide insights into users' emotional responses during interaction.

SUBJECTIVE METRICS IN VR USABILITY TESTING

The evaluation of VR usability often relies on subjective metrics, which capture users' perceptions and experiences during interaction with immersive environments. While objective metrics provide measurable information about user performance—such as task completion time or error rates—subjective measures reveal how users perceive usability, workload, comfort, and overall interaction quality. This perspective is particularly important in VR, where immersive interaction may evoke cognitive, emotional, and physiological responses that cannot be fully captured through behavioural observations alone.

Subjective evaluation typically relies on self-reported data collected through standardized questionnaires, in which participants rate their experience using structured scales, most commonly 5- or 7-point rating scales. These questionnaires are usually administered after completing tasks or the entire usability test and are most informative when interpreted alongside objective performance metrics.

In usability research, standardized questionnaires are preferred over ad hoc instruments because they have demonstrated validity, reliability, and sensitivity to meaningful differences between systems. By collecting structured feedback on users' perceptions, subjective metrics help identify usability issues that may not be visible in performance data alone.

The following subsections present selected subjective metrics commonly used in VR usability testing and user experience evaluation, illustrating how these instruments support the systematic assessment of immersive systems.

System Usability Scale

The **System Usability Scale (SUS)** is a widely used instrument for evaluating the perceived usability of interactive systems. Developed by Brooke (1996), the SUS is a standardized questionnaire consisting of 10 statements rated on a 5-point Likert scale, typically ranging from strongly disagree to strongly agree (Figure 15.2). The responses are combined to produce a single usability score ranging from 0 to 100, representing the user's overall perception of the system.

The questionnaire captures several usability aspects, including ease of use, system complexity, learnability, interaction consistency, and user confidence. Although SUS provides a single overall score, the items collectively reflect both usability and learnability characteristics of the interface.

The SUS score is calculated using a standardized procedure and serves as a relative usability measure, allowing comparison between different systems or design iterations. A score of approximately 68 is typically considered average usability, while higher values indicate better perceived usability.

Due to its simplicity, efficiency, and reliability, SUS is widely used in usability research. In VR usability studies, participants typically complete the questionnaire after performing predefined tasks in the virtual environment. When combined with objective performance metrics, SUS helps identify usability issues and evaluate improvements across iterative design cycles.

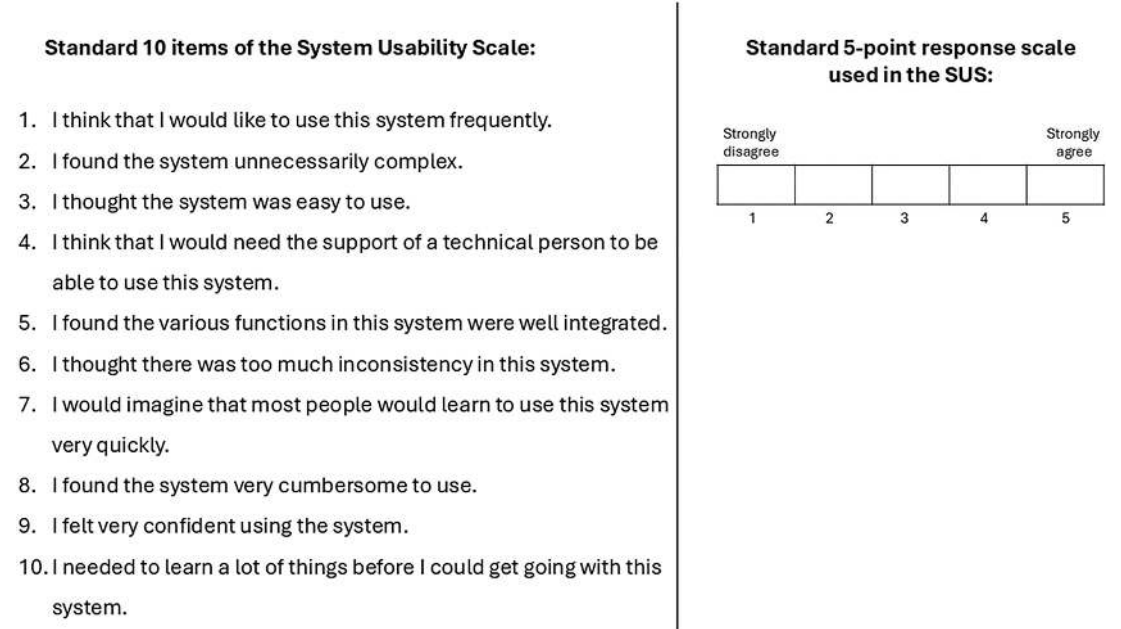


FIGURE 15.2 SUS questionnaire. (Adapted from Brooke, 1996.)

User Experience Questionnaire

The *User Experience Questionnaire (UEQ)* is a widely used instrument for assessing users’ subjective experiences during interaction with interactive systems. It evaluates both usability-related qualities and emotional aspects of interaction, providing a structured method for collecting users’ impressions after system use.

The original UEQ (see Figure 15.3) consists of 26 items presented as pairs of opposite adjectives (semantic differentials), such as understandable–not understandable or exciting–boring. Participants rate each item on a 7-point scale ranging from –3 to +3, where the endpoints represent negative and positive attributes. Completing the questionnaire typically takes 3–5 minutes, making it suitable for usability and user experience studies.

The UEQ measures user experience across six dimensions: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. The first four dimensions mainly reflect pragmatic qualities related to usability, while stimulation and novelty capture hedonic aspects of interaction, such as engagement and innovation (Schrepp et al., n.d.).

A shorter version of the questionnaire, UEQ-S, was later developed to support rapid evaluations. It contains eight items that capture the essential aspects of user experience while significantly reducing completion time, making it suitable for repeated measurements or studies involving multiple prototypes (Schrepp et al., 2017).

In the context of VR usability research, UEQ and UEQ-S are useful for evaluating how users perceive immersive systems beyond traditional usability metrics. VR applications aim not only to support efficient task performance but also to provide engaging and stimulating experiences. By measuring both pragmatic and hedonic qualities, the UEQ enables a comprehensive evaluation of interaction quality in virtual environments.

NASA Task Load Index

The *NASA Task Load Index (NASA-TLX)* is a widely used instrument for measuring subjective workload experienced by users while performing tasks in interactive systems. The method was developed by Hart and Staveland to assess human workload in complex human–machine interactions, originally in aviation

Assess by ticking one circle per line.

	1	2	3	4	5	6	7		
annoying	☐	☐	☐	☐	☐	☐	☐	enjoyable	1
not understandable	☐	☐	☐	☐	☐	☐	☐	understandable	2
creative	☐	☐	☐	☐	☐	☐	☐	dull	3
easy to learn	☐	☐	☐	☐	☐	☐	☐	difficult to learn	4
valuable	☐	☐	☐	☐	☐	☐	☐	inferior	5
boring	☐	☐	☐	☐	☐	☐	☐	exciting	6
not interesting	☐	☐	☐	☐	☐	☐	☐	interesting	7
unpredictable	☐	☐	☐	☐	☐	☐	☐	predictable	8
fast	☐	☐	☐	☐	☐	☐	☐	slow	9
inventive	☐	☐	☐	☐	☐	☐	☐	conventional	10
obstructive	☐	☐	☐	☐	☐	☐	☐	supportive	11
good	☐	☐	☐	☐	☐	☐	☐	bad	12
complicated	☐	☐	☐	☐	☐	☐	☐	easy	13
unlikable	☐	☐	☐	☐	☐	☐	☐	pleasing	14

FIGURE 15.3 Example items from the User Experience Questionnaire (UEQ). (Adapted from Schrepp et al., n.d.)

and aerospace contexts (Hart & Staveland, 1988). Today, NASA-TLX is commonly applied in human–computer interaction, usability studies, and VR research.

The questionnaire evaluates perceived workload across six dimensions:

- **Mental Demand:** cognitive effort required for the task (e.g., thinking or decision-making)
- **Physical Demand:** physical effort required during interaction
- **Temporal Demand:** time pressure experienced during task execution
- **Performance:** self-assessment of task success
- **Effort:** amount of work required to accomplish the task
- **Frustration Level:** stress or irritation experienced during interaction

Participants complete the NASA-TLX assessment (see Figure 15.4) after performing tasks in the evaluated system, rating each dimension on a numerical scale, typically from 0 to 100.

The original NASA-TLX includes a pairwise weighting procedure to determine the relative importance of each workload dimension. However, many usability studies apply a simplified version known as *Raw TLX (RTLX)*, where the overall workload score is calculated as the average of the six ratings without weighting (Hart, 2006).

In VR usability studies, NASA-TLX helps identify tasks or interaction mechanisms that impose excessive cognitive or physical effort. When used alongside usability and user experience measures, it provides valuable insight into the workload associated with interaction in immersive environments and supports the iterative improvement of VR applications.

Virtual Reality System Usability Questionnaire

The *Virtual Reality System Usability Questionnaire (VRSUQ)* is a usability evaluation instrument designed specifically for assessing the perceived usability of VR systems. While general tools such as the SUS are widely used, they were not developed with immersive environments in mind and may not fully

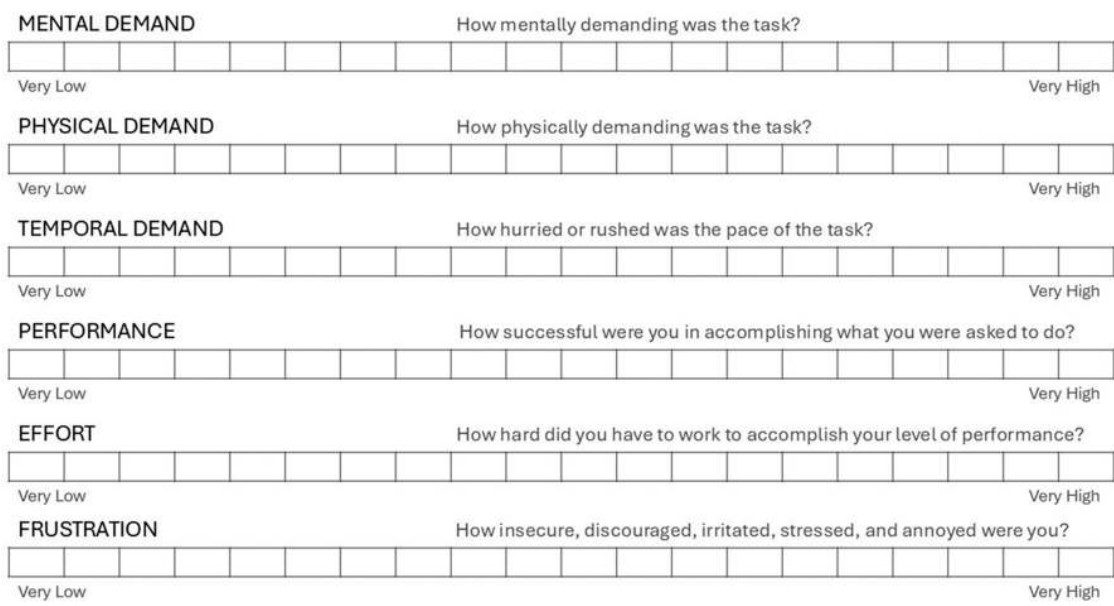


FIGURE 15.4 NASA Task Load Index (NASA-TLX) rating scale. (Adapted from National Aeronautics and Space Administration, n.d.)

capture aspects of VR interaction, such as spatial navigation, embodied movement, or sensory immersion. As a result, VR studies often combine multiple questionnaires to evaluate presence, workload, or motion sickness.

To address these limitations, Kim and Rhiu (2024) developed the VRSUQ as a concise instrument tailored to VR environments. The questionnaire consists of nine items rated on a 5-point Likert scale, allowing participants to evaluate their experience after interacting with a VR application (see Figure 15.5).

The questionnaire evaluates usability across three dimensions aligned with the ISO 9241-11 usability framework: effectiveness, efficiency, and satisfaction. Effectiveness refers to how well the system supports successful task completion and includes indicators such as compatibility between user actions and system responses, system feedback, and error prevention. Efficiency reflects how easily users can understand and learn to operate the system and how effectively they can recover from errors during interaction. The satisfaction dimension captures the user’s subjective experience of interacting with the VR application, including enjoyment of the experience as well as potential physical or emotional discomfort.

VRSUQ Items Across Usability Dimensions (Effectiveness, Efficiency, Satisfaction)

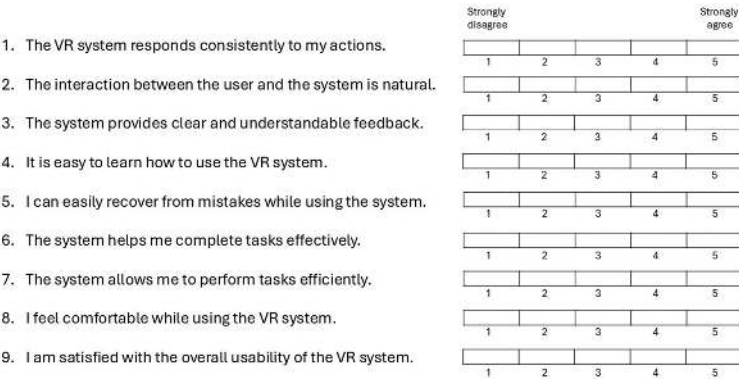


FIGURE 15.5 Virtual Reality System Usability Questionnaire (VRSUQ) items. (Adapted from Kim & Rhiu, 2024.)

Due to its VR-specific focus and concise structure, VRSUQ is well suited for usability studies of immersive systems. It enables researchers to identify usability issues related to interaction design and user comfort, supporting the iterative improvement of VR applications.

Motion Sickness Questionnaires

Exposure to immersive virtual environments may induce cybersickness or simulator sickness, manifested by symptoms such as nausea, dizziness, visual fatigue, or general discomfort. These effects are typically caused by sensory conflicts between visual motion cues in the virtual environment and signals from the vestibular system. Because such physiological responses can strongly influence users' perceived usability and overall experience, assessing motion sickness is an important component of VR usability testing.

Two commonly used instruments for evaluating these symptoms are the *Simulator Sickness Questionnaire (SSQ)* and the *Virtual Reality Sickness Questionnaire (VRSQ)*. The SSQ, developed by Kennedy et al. (1993), measures 16 symptoms grouped into three categories: nausea, oculomotor disturbances, and disorientation. Participants rate the severity of each symptom after exposure to the virtual environment, allowing researchers to calculate both subscale scores and an overall sickness score.

Because VR systems involve head-mounted displays and immersive interaction, instruments specifically adapted to VR have also been proposed. The VRSQ, developed by Ames et al. (2005), provides a more concise assessment focused on oculomotor disturbances and disorientation, capturing symptoms such as visual fatigue, blurred vision, and dizziness. Its shorter structure makes it particularly suitable for VR studies involving repeated measurements or multiple experimental conditions.

In VR usability evaluation, motion sickness questionnaires provide important insight into user comfort and physiological responses to immersive interaction. High cybersickness levels may indicate issues related to locomotion design, visual latency, or mismatches between visual and physical motion, and therefore complement traditional usability and user experience measures.

OBJECTIVE METRICS IN VR USABILITY TESTING

Objective metrics play a key role in usability evaluation because they provide measurable and reproducible indicators of user interaction with a system (Jerald, 2015). Unlike subjective measures based on self-reports, objective metrics rely on observable behaviour and system-generated data. In VR environments, such data can be automatically collected through sensors integrated in VR hardware and software.

In VR usability studies, objective metrics are commonly used to assess task performance, interaction behaviour, and user responses during the experience. Modern VR systems enable the recording of diverse data such as spatial movement, controller interactions, gaze direction, and other behavioural or physiological signals.

Selecting appropriate metrics depends on the goals of the study. Some indicators focus on task efficiency and success, while others reveal navigation patterns and interaction behaviour within the virtual environment. Together, these measures provide valuable evidence for evaluating usability and supporting the iterative improvement of VR applications.

Performance-Based Metrics

Performance-based metrics are among the most used indicators in usability testing. In VR environments they are typically derived from tasks that participants are asked to complete during the test scenario. When designing a usability study, researchers should first define a set of representative tasks that reflect the intended use of the application.

The simplest and most informative metric is task completion time (Rubin & Chisnell, 2008; La Viola et al., 2017). Measuring how long users need to complete a task helps identify interaction elements that slow users down or create confusion. For example, unusually long completion times may indicate unclear instructions, poorly visible interactive objects, or unintuitive navigation mechanisms.

Another key metric is *task success rate*, which indicates whether users can achieve the intended goal. A low success rate usually signals critical usability issues such as unclear interaction logic or poorly designed interface elements.

Researchers should also record *errors* and *repeated attempts*. These metrics reveal where users struggle during interaction. Frequent errors often indicate that the system's *affordances* are not sufficiently clear or that users misunderstand how to interact with virtual objects.

Finally, *navigation efficiency* can be analysed by comparing the path taken by the user with the optimal path defined by the system designers. Large deviations may suggest that spatial cursor environmental guidance mechanisms are insufficient. Figure 15.6 presents heat maps of user navigation paths together with the time spent in specific locations during the task. The two examples shown in the figure illustrate navigation behaviour of two participants performing the same task with the same LIGHT-type visual guidance indicating the direction of the target. In one case, the participant largely overlooked the visual cue and explored the environment extensively, resulting in a total search time of 466.52 s. In the other case, the participant noticed the guidance early and followed a near-optimal path, completing the task in 35.56 s. This comparison demonstrates how differences in perceiving the same environmental cue can significantly affect navigation efficiency and task completion time.

When analysed together, these performance indicators provide a clear picture of how efficiently users interact with the system and where iterative improvements should focus.

Behavioural and Interaction Metrics Captured by VR Systems

One of the major advantages of usability testing in VR is the ability to automatically collect detailed behavioural data from the headset and controllers. Modern VR systems continuously track the position and movement of the user, allowing researchers to reconstruct how participants explore and interact with the virtual environment.

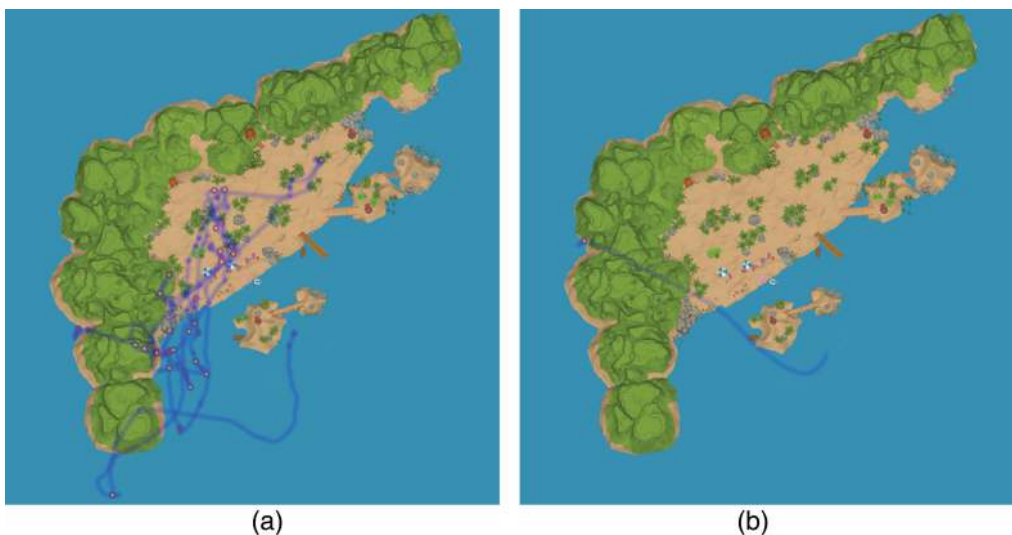


FIGURE 15.6 Heat maps of user navigation paths and time spent in specific locations during the task: (a) total search time 466.52 s and (b) 35.56 s.

These data can be used to analyse navigation behaviour and interaction patterns (Jerald, 2015). For example, tracking spatial trajectories can reveal whether users explore the environment efficiently or repeatedly return to the same locations due to uncertainty. Such patterns often indicate problems with spatial orientation or unclear task guidance.

Another useful indicator is head orientation and movement behaviour. By analysing where users direct their head movements, researchers can infer which elements attract attention and which objects may be overlooked. If critical elements are consistently ignored, this may indicate problems with visual salience or interface placement.

An additional and increasingly important source of behavioural data in VR usability studies is eye-tracking, which is now integrated into many modern head-mounted displays. Eye-tracking technology enables precise recording of users' gaze direction, fixation duration, and visual attention patterns during interaction with the virtual environment (Duchowski, 2017). While head orientation provides only an approximation of where users may be looking, eye-tracking allows researchers to determine exactly which elements of the scene attract visual attention.

Gaze data can be used to analyse how users visually explore the environment, how long they focus on specific objects, and whether important interface elements or guidance cues are noticed. If users repeatedly fail to fixate on critical objects, this may indicate insufficient visual salience or ineffective placement of interface elements. Eye-tracking data can also be visualized using gaze heat maps, which illustrate how visual attention is distributed across the scene.

An illustrative example of this type of analysis is presented in Figure 15.7. The figure shows gaze heat maps obtained during a VR task in which participants attempted to locate a mosquito in a dark virtual environment. The task was performed under two conditions: without auditory guidance and with spatial sound cues indicating the direction of the target. Although both conditions required the same visual search behaviour, the presence of spatial audio influenced the distribution of visual attention. When auditory cues were available, gaze patterns were more concentrated and directed toward the target area, indicating a more efficient search strategy. In contrast, without auditory guidance, the search behaviour appeared more dispersed, suggesting greater uncertainty during the task.

Interaction logs generated by the system can also provide information about object manipulation frequency, menu usage, or controller input patterns. Repeated or unnecessary interactions may signal that the interface requires refinement. These automatically recorded behavioural traces are particularly valuable because they capture user behaviour continuously and unobtrusively, providing insights that may not emerge from questionnaires or interviews.

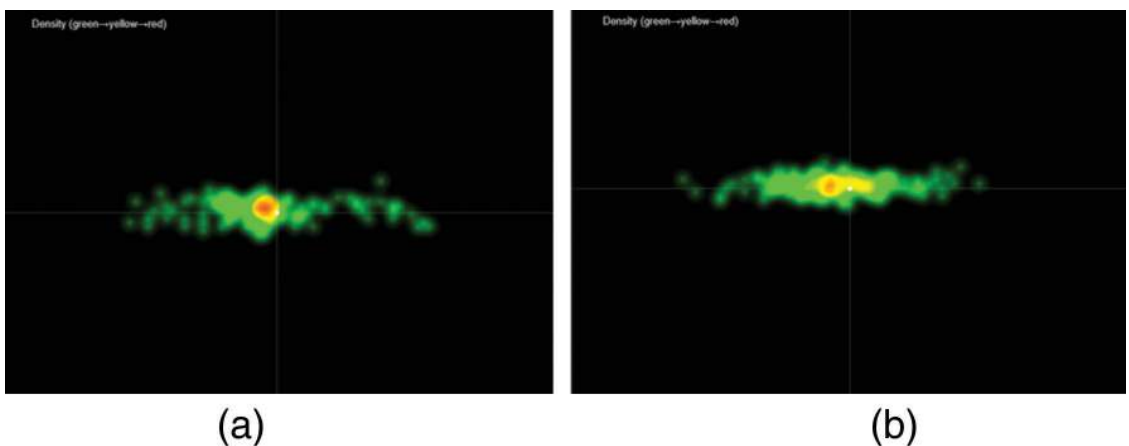


FIGURE 15.7 Eye-tracking heat maps showing gaze distribution during the mosquito-search task in VR: (a) muted condition (no spatial audio) and (b) spatial sound guidance.

Physiological and Affective Metrics

In more advanced usability studies, researchers may extend the analysis by incorporating physiological and affective measurements. These metrics provide additional insights into users' emotional states, stress levels, and cognitive load during interaction with VR systems (Cacioppo et al., 2007). Unlike behavioural metrics derived from movement or task performance, physiological signals allow researchers to capture internal responses that may not be directly observable during interaction.

One increasingly important source of physiological and cognitive information in VR studies is eye-tracking, which is now integrated into many modern head-mounted displays. Eye-tracking technology enables the recording of gaze direction, fixation duration, and visual attention patterns during interaction with the virtual environment. Beyond analysing attention distribution, eye-tracking can also provide indirect indicators of cognitive load, for example, through changes in fixation behaviour or pupil dilation. These measures help researchers understand how users process visual information in immersive environments and whether certain interface elements require excessive effort or attention.

Another source of affective data comes from facial tracking technologies, which allow the detection of subtle facial muscle movements and micro-expressions. In modern VR headsets such as Meta Quest Pro, facial expression tracking is enabled through inward-facing infrared cameras and sensors embedded inside the headset. These sensors capture movements of facial muscles around the eyes, eyebrows, cheeks, and mouth while the user interacts with the virtual environment.

The recorded signals are processed by the headset's tracking algorithms, which estimate facial expression parameters and translate them into a set of facial movement indicators. These parameters can then be analysed to identify emotional reactions such as frustration, confusion, surprise, or engagement occurring during specific stages of interaction. Because the sensors operate continuously during the VR session, facial tracking enables unobtrusive monitoring of affective responses without interrupting the immersive experience.

In addition to analysing emotional responses, facial tracking can also be used to animate realistic avatars that mirror the user's facial expressions in real time. An example of such facial expression tracking and avatar mapping in the Meta Quest Pro system is illustrated in Figure 15.8.



FIGURE 15.8 Example of facial expression tracking and real-time avatar animation in Meta Quest Pro.

Another increasingly adopted approach in VR usability research is the integration of *electroencephalography* (EEG) sensors. EEG devices allow researchers to record brain activity and analyse neural signals associated with attention, cognitive workload, emotional engagement, mental fatigue, and presence during interaction (Kober & Neuper, 2012). With the rapid development of lightweight wearable EEG systems, integrating brain–computer interface technologies into VR studies has become increasingly feasible.

Different EEG devices provide varying levels of measurement detail and user comfort. More advanced systems, such as multi-channel EEG headsets, can capture brain signals from multiple regions of the scalp, providing richer information about neural activity. At the same time, newer lightweight devices with a smaller number of frontal electrodes focus primarily on signals associated with attention and cognitive load. Although such devices collect less detailed neural data, they are often less intrusive and more comfortable for users, making them particularly suitable for usability studies conducted in immersive environments. Examples of these two approaches are illustrated in Figure 15.9, showing advanced multi-channel EEG systems and lightweight frontal EEG devices that can be used during VR experiments.

In addition to EEG, other physiological sensors may also be integrated into VR usability studies. Wearable devices such as wristband sensors are increasingly used to monitor physiological responses during immersive interactions. These devices can record signals such as heart rate, *heart rate variability* (HRV), and *electrodermal activity* (EDA), which provide valuable indicators of physiological arousal, stress, and emotional activation (Healey & Picard, 2005).

EDA reflects changes in skin conductance associated with sympathetic nervous system activity and is often used to measure stress levels or emotional engagement during tasks. Similarly, heart rate and HRV can provide insights into cognitive workload and emotional responses experienced by users during interaction with a virtual environment.

An example of such a wearable physiological monitoring device is shown in Figure 15.10. Devices such as the Empatica wristband allow researchers to collect physiological signals continuously while users perform tasks in VR. Because the sensor is worn on the wrist and does not interfere with head-mounted displays or controllers, it enables unobtrusive monitoring of physiological responses during immersive experiments.

These signals are particularly useful in applications related to training, healthcare, or emotionally engaging VR experiences, where understanding users' physiological responses can provide important insights into stress, engagement, and cognitive load.

Although the integration of physiological sensors requires additional equipment and more complex data analysis, such measurements can significantly enrich usability evaluation by revealing aspects of the user



FIGURE 15.9 Examples of EEG devices used in VR studies: (a) DSI-24 multi-channel EEG system and (b) lightweight frontal EEG headset.



FIGURE 15.10 Example of a wearable physiological sensor (Empatica wristband) used to monitor heart rate and electrodermal activity during VR experiments.

experience that are not directly observable through behavioural metrics alone. Combining physiological, behavioural, and performance data allows researchers to obtain a more comprehensive understanding of user interaction and experience in immersive environments.

ITERATIVE DEVELOPMENT OF THE VR APPLICATION IN PRACTICE

At the Voxel Research Lab, VR applications are developed using an iterative design methodology, in which prototypes are repeatedly tested with users and refined based on usability evaluation results. As illustrated in Figure 15.1, iterative design integrates development, usability testing, and analysis in successive cycles, allowing researchers to progressively improve interaction mechanisms and the overall user experience of immersive systems. This approach has been applied across several research initiatives conducted at the lab, including the *ATOMIC*, *MrUD*, and *VRxanny* projects (Voxel Research Lab, n.d.).

In the *ATOMIC* project, immersive training scenarios were developed to support management skills such as planning, problem-solving, teamwork, and communication in simulated business environments. The project implemented multiple extended-reality solutions—including VR scenarios—and evaluated them through usability testing with students acting as end users. One of the exercises focused on constructive criticism, where participants interacted with virtual characters in a managerial communication scenario. The evaluation results helped identify strengths and limitations of the interaction design and supported improvements in the training environment and interface mechanisms (Zwoliński et al., 2022).

In the *MrUD* project, immersive simulations were created to raise awareness of universal design and accessibility by allowing users to experience everyday situations from the perspective of individuals with different physical limitations. User testing with students (see Figure 15.11) provided feedback on both the educational impact and usability of the VR application (Raposo et al., 2023). The *VRxanny* project applies a similar iterative approach to develop immersive environments aimed at reducing anxiety and stress among university students, where user feedback supports improvements in environmental design and interaction simplicity (Kamińska et al., 2025).

In addition to traditional usability evaluation methods, the research team also explored data-driven and partially automated approaches to VR usability testing, combining behavioural signals collected from VR tracking systems with subjective feedback from participants. As shown in previous research (Kamińska et al., 2022), integrating multiple data sources—such as behavioural metrics, physiological signals, and user-reported feedback—can support a more comprehensive evaluation of interaction in virtual environments. This multimodal approach facilitates the identification of usability issues at early stages of VR application development and contributes to more effective iterative improvements of immersive systems.

SUMMARY OF THE APPLIED METRICS

Usability evaluation in VR benefits from combining subjective and objective metrics, which capture complementary aspects of user interaction. Subjective measures reflect users' perceptions and experiences, while objective metrics provide measurable indicators of performance and behaviour during interaction with the system.

Tables 15.1 and 15.2 summarize selected metrics commonly used in VR usability testing, indicating how they can be collected and what aspects of usability they help interpret.

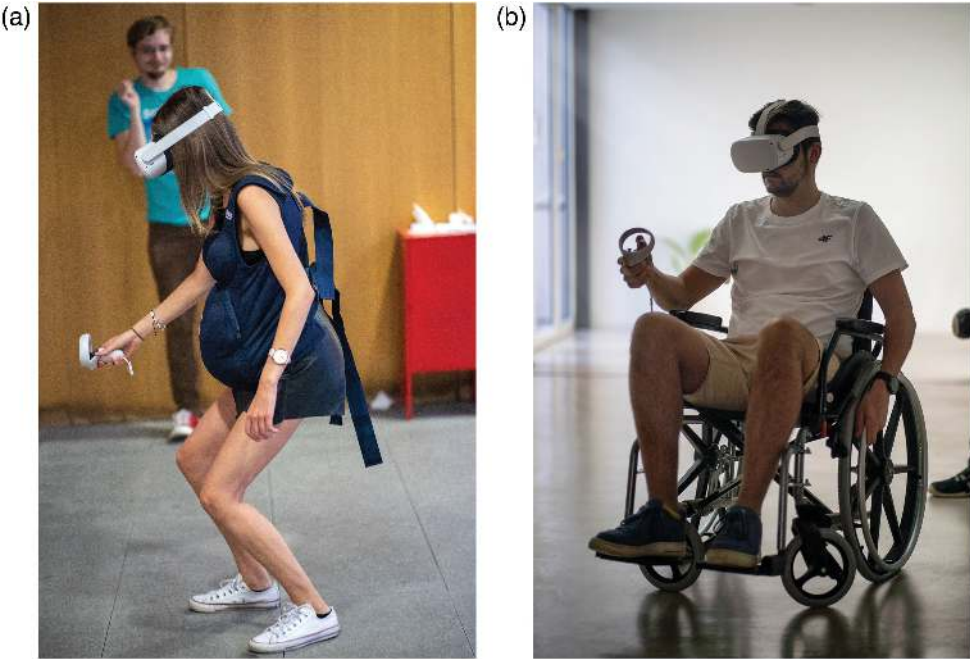


FIGURE 15.11 User testing with students as end users of the VR applications developed in the MrUD project. (a) Scenario using a pregnancy-belly simulator; (b) scenario using a wheelchair.

TABLE 15.1 Subjective metrics commonly used in VR usability testing

<i>METRIC</i>	<i>HOW TO COLLECT</i>	<i>WHAT IT INDICATES</i>
System usability scale (SUS)	Participants complete a ten-item questionnaire rated on a 5-point Likert scale after interacting with the VR application.	Overall perceived usability; measures ease of use, learnability, and user confidence when interacting with the system.
User experience questionnaire (UEQ)	Participants evaluate the system using bipolar adjective scales (e.g., efficient–inefficient, exciting–boring) after the VR session.	User experience aspects including attractiveness, efficiency, perspicuity, dependability, stimulation, and novelty.
NASA task load index (NASA-TLX)	Participants rate six workload dimensions (mental demand, physical demand, temporal demand, performance, effort, frustration) after completing tasks.	Perceived cognitive and physical workload during interaction; helps identify tasks or interface elements that require excessive effort.
Virtual reality system usability questionnaire (VRSUQ)	Participants complete a questionnaire specifically designed for VR systems after finishing interaction tasks.	VR-specific usability factors such as interaction naturalness, ease of control, comfort of the interface, and clarity of VR interaction mechanisms.
Simulator sickness questionnaire (SSQ)/virtual reality sickness questionnaire (VRSQ)	Participants report symptoms such as nausea, dizziness, and visual discomfort after VR exposure.	Level of cybersickness and physical discomfort experienced during the VR session.

TABLE 15.2 Objective metrics commonly used in VR usability testing

<i>METRIC</i>	<i>HOW TO COLLECT</i>	<i>WHAT IT INDICATES</i>
Task completion time	Measure time required for a participant to complete a predefined task; usually recorded automatically in application logs.	Interaction efficiency; long times may indicate unclear interfaces, navigation problems, or cognitive load.
Task success rate	Record whether the participant completed the assigned task successfully.	Interface effectiveness; low success rates suggest major usability issues.
Error rate	Log incorrect actions such as wrong selections or failed interactions.	Identifies confusing interface elements and problematic interaction design.
Number of attempts	Count how many times users repeat an action before succeeding.	Suggests unclear affordances or insufficient interaction feedback.
Navigation path efficiency	Compare the user's path with the optimal path defined in the scenario.	Reveals spatial orientation issues or poor environmental guidance.
Head orientation and movement	Use headset tracking data to record head rotation and direction.	Indicates attention direction and which elements attract or fail to attract user attention.
Spatial movement patterns	Analyse user trajectories in the virtual environment using positional tracking.	Shows exploration behaviour and possible disorientation in the space.
Interaction frequency	Log controller inputs, object manipulations, and menu interactions.	Reveals interaction complexity and possible usability bottlenecks.
Gaze behaviour (eye-tracking)	Use integrated or external eye-tracking modules to record gaze direction and fixation duration.	Shows visual attention patterns and detects overlooked interface elements.
Facial expressions/ micro-expressions	Capture facial movements using facial tracking sensors available in some VR headsets.	Provides insight into emotional reactions such as confusion, frustration, or engagement.
Physiological responses (HR, EDA)	Use wearable sensors measuring heart rate or electrodermal activity during the VR session.	Indicates physiological arousal, stress level, or cognitive load.

CONCLUDING REMARKS

Usability testing is a key component in the development of VR applications, where immersive interaction introduces unique challenges related to spatial navigation, embodied interaction, and sensory feedback. This chapter presented selected methods and metrics used in VR usability evaluation, including subjective measures capturing users' perceptions and experiences, and objective metrics derived from behavioural and system data. When combined, these approaches provide complementary insights into how users interact with immersive environments and help identify usability issues that may not be visible through a single type of measurement. Integrating these evaluation methods into an iterative development process enables designers to refine VR applications based on empirical evidence, ultimately supporting the creation of more intuitive, efficient, and engaging immersive systems.

REFERENCES

- Ames, S. L., Wolffsohn, J. S., & McBrien, N. A. (2005). The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. *Optometry and Vision Science*, 82(3), 168–176.
- Brooke, J. (1996). SUS—A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Cacioppo, J. T., Tassinari, L. G., & Berntson, G. G. (2007). Psychophysiological science: Interdisciplinary approaches to classic questions about the mind. *Handbook of Psychophysiology*, 3, 1–16.
- Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.
- Hart, S. G. (2006). NASA-task load index (NASA-TLX): 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904–908. <https://doi.org/10.1177/154193120605000909>.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland.
- Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems*, 6(2), 156–166.
- Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.
- Kamińska, D., Zwoliński, G., Cuiñas, I., G. Castaño, I., & Merecz-Kot, D. (2025). VR for well-being: Reducing student stress through immersive experiences. In *Proceedings of the 2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Sydney, Australia (pp. 638–639). IEEE. <https://doi.org/10.1109/ISMAR-Adjunct68609.2025.00130>.
- Kamińska, D., Zwoliński, G., & Laska-Leśniewicz, A. (2022). Usability testing of virtual reality applications: The pilot study. *Sensors*, 22(4), 1342. <https://doi.org/10.3390/s22041342>.
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3), 203–220.
- Kim, Y. M., & Rhiu, I. (2024). Development of a virtual reality system usability questionnaire (VRSUQ). *Applied Ergonomics*, 119, 104319. <https://doi.org/10.1016/j.apergo.2024.104319>.
- Kober, S. E., & Neuper, C. (2012). Using auditory event-related EEG potentials to assess presence in virtual reality. *International Journal of Human-Computer Studies*, 70(9), 577–587.
- La Viola Jr J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: Theory and practice*. Addison-Wesley Professional.
- National Aeronautics and Space Administration. (n.d.). *NASA task load index (TLX) rating scale*. <https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf>.
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Raposo, R., Vairinhos, M., Laska-Leśniewicz, A., & Sztobryn-Giercuskiewicz, J. (2023). Increasing awareness and empathy among university students through immersive exercises—testing of the virtual reality application: A pilot study. *Medycyna Pracy. Workers' Health and Safety*, 74(3), 187–197. <https://doi.org/10.13075/mp.5893.01391>.
- Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan, design, and conduct effective tests*. John Wiley & Sons.
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (n.d.). *User experience questionnaire (UEQ)*. <https://www.ueq-online.org/>.
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6), 103–108. <https://www.ijimai.org/index.php/ijimai/article/view/6119/3537>.
- Voxel Research Lab. (n.d.). *Voxel Research Lab*. <https://voxellab.pl/>.
- Zwoliński, G., Kamińska, D., Laska-Leśniewicz, A., Haamer, R. E., Vairinhos, M., Raposo, R., Urem, F., & Reisinho, P. (2022). Extended reality in education and training: Case studies in management education. *Electronics*, 11(3), 336. <https://doi.org/10.3390/electronics11030336>.

Design Translation Patterns for Virtual Reality

16

Grzegorz Zwoliński

ABSTRACT

This chapter presents a conceptual and practical framework for translating screen-based (2D) user experiences into immersive virtual reality (VR). Rather than treating VR as a “screen floating in space,” we argue that effective translation requires redesigning user intentions, task logic, information structure, attention dynamics, and perceptual assumptions for an embodied and spatial medium. We introduce a consistent pattern description template—covering the original 2D element, the user’s underlying intention, constraints that emerge in VR, a spatial translation strategy, and typical risks and anti-patterns—to support systematic analysis and design reuse.

Building on this framework, this chapter surveys a broad landscape of potential translations across interface structure, interaction, navigation, content consumption, narrative, social communication, and productivity workflows. It then develops ten canonical Extended Reality (XR) translation patterns that illustrate fundamental shifts from framed observation to embodied presence, from cursor-mediated input to direct manipulation, from scrolling to spatial progression, from hyperlinks to portal-like transitions, from dashboards to navigable data environments, from window multitasking to spatial workspaces supported by spatial memory, from camera framing to attention choreography, from pop-up notifications to environmental events, from abstract timelines to spatialized time, and from toolbar commands to tool-based interaction. Finally, this chapter synthesizes these patterns into an overarching model of transformation from abstract interface conventions toward perceptually coherent, embodied, and continuous experiences—providing designers and researchers with a durable foundation for prototyping, evaluation, and future XR design practice.

INTRODUCTION

Designing for virtual reality (VR) requires a fundamental shift in how interaction, perception, and user experience are conceptualized. VR should not be treated as an extension of screen-based media, nor as a three-dimensional container for familiar interface elements. One of the most common mistakes in early immersive systems is the direct transfer of desktop or mobile layouts into virtual environments, resulting in interfaces that function technically but fail to leverage the perceptual and cognitive properties of immersive media.

Screen-based systems operate within a framed experience. Users interact with content through a bounded visual surface—a monitor, tablet, or smartphone screen—which determines what remains visible and what is excluded. Designers rely on this frame to guide attention through layout hierarchy, scrolling, modal windows, and sequential screen transitions.

VR fundamentally alters this relationship. Instead of observing content through a frame, the user becomes situated within a surrounding environment. Body orientation, head movement, gaze direction, and spatial awareness become integral to interaction. Actions that in 2D rely on symbolic mediation—such as clicking, dragging, or tapping—can be replaced by spatially grounded actions including reaching, grasping, turning, or approaching objects in VR. This reduces the distance between intention and execution, while simultaneously introducing new constraints related to comfort, fatigue, spatial orientation, and cognitive load (La Viola et al., 2017).

A particularly important difference concerns attention management. In screen-based systems, attention is controlled by limiting what is visible inside the frame. In immersive environments, users can look in any direction, which means critical information may fall outside their immediate focus. As a result, VR design requires active guidance through spatial composition, lighting, directional sound, movement, and environmental cues rather than conventional visual hierarchy (Jerald, 2015; Slater & Sanchez-Vives, 2016).

These differences explain why simple migration of 2D interfaces into VR is insufficient. Preserving appearance does not preserve user experience. A flat panel suspended in space may reproduce functionality, but it rarely supports embodied interaction or perceptual coherence.

For this reason, translating 2D systems into VR should begin not with interface elements, but with user intention. Users do not seek buttons or menus as such; they seek to complete tasks, understand information, communicate, or navigate complex processes. Effective VR design therefore asks not how an interface should look in immersive space, but how a goal can be achieved through natural action within that space.

This perspective allows VR to be treated as a distinct interaction medium and establishes the basis for systematic translation patterns between screen-based and immersive design.

WHAT IS EXPERIENCE TRANSLATION?

In the context of 2D → VR, experience translation does not mean transferring visual interface elements from a screen into three-dimensional space. It is not about “recreating” a layout or turning buttons into 3D objects. Translation operates at a much deeper level of design—the level of user experience.

To understand the essence of translation, we must assume that designing for VR does not begin with the question, “What should it look like?” but rather, “What is the user trying to achieve?” This implies that translation occurs across several key layers.

User Intention

Every interaction in a 2D system represents the realization of a specific intention: the user wants to find information, proceed to the next step, select an option, compare data, confirm an action, or connect with others.

In screen-based environments, these intentions are implemented through graphical conventions such as buttons, menus, scrolling, and forms. However, interface elements themselves are not the user’s goal—they are merely tools.

In VR, design begins by identifying that intention and asking how it can be realized naturally in space. For example:

- The intention to “move forward” in 2D may be executed by clicking a “Next” button.
- In VR, it may involve physically walking into the next area, opening a door, or activating an object.

Task Logic

Task logic in 2D systems is typically organized as a sequence of discrete actions such as clicking, navigating, completing a form, and confirming an operation. This structure is usually linear and reflects the screen-based model in which interaction unfolds through a succession of consecutive views. In VR, however, this logic changes because the user no longer moves between screens but exists within an environment. Translation therefore requires redesigning task logic from a screen-based sequence into a spatial sequence. Instead of progressing through a chain of screens, the experience may unfold through movement between places, interaction with objects, or exposure to environmental events. The essential challenge is to transform procedural structure into spatial experience while preserving the functional coherence of the task.

Information Structure

In screen-based systems, information structure is organized through visual hierarchy, layout grids, layering, and scrolling, with designers controlling visibility within the limited surface of the screen. In VR, by contrast, information may be distributed throughout space, which means that structure no longer needs to remain vertical or linear. It can instead be arranged radially, zonally, depth-wise, or according to spatial relationships between elements. Translating information structure into immersive space therefore requires reconsidering whether a previously linear arrangement should remain sequential, whether it could instead function as a spatial map, and whether the user is expected merely to observe information or to move through it as part of the experience.

Attention Dynamics

In 2D environments, attention is largely controlled through framing and composition, since designers determine which elements remain visible and in what sequence they appear. In VR, however, users are free to look in any direction, which fundamentally changes the conditions of attentional control. Translation therefore requires the use of spatial mechanisms that guide attention through lighting, movement, directional sound, scale changes, and interaction cues. Elements that attract attention in screen-based systems through color, contrast, or layout position may consequently require entirely different strategies when reinterpreted within immersive space.

Perceptual Model

Screen-based systems rely on the perception of a flat surface. Depth is an illusion, and interaction occurs within a single plane.

VR engages spatial perception, proprioception, and bodily awareness. Objects possess scale, distance, and symbolic and physical weight. Elements that are too small may be unreadable; too distant, difficult to interact with; too numerous, cognitively overwhelming.

Experience translation therefore requires acknowledging that VR fundamentally changes how users see, orient themselves, and respond within an environment.

Structure of a Translation Pattern Description

To organize the analysis of experience translation between screen-based and immersive environments, it is useful to adopt a consistent pattern description structure. Translation from 2D to VR is not a one-time interface transformation but a systematic reinterpretation of how users act within a different perceptual and interaction model.

Each translation pattern can be described according to the following structure:

- 1. **2D Element:** What solution functions in the screen-based environment? This may include a specific interface component (e.g., button, scroll bar) or a broader experience structure (e.g., dashboard, information feed, workflow).
- 2. **User Intention:** What does the user actually want to achieve? Interface elements are merely tools for accomplishing goals—exploring content, making decisions, communicating, progressing through a task, or orienting within information.
- 3. **Constraints in VR:** Why does direct transfer fail in immersive space? Problems may stem from the absence of a frame, changes in depth perception, physical fatigue, spatial disorientation, or cognitive overload.
- 4. **Spatial Translation:** How can the same intention be redesigned in VR? Translation may involve leveraging user movement, object manipulation, environmental cues, scale transformation, or spatial narrative.
- 5. **Risks and Anti-Patterns:** What should be avoided? What common mistakes arise when screen-based interfaces are copied directly, leading to reduced comfort, orientation loss, or diminished immersion?

Overview of 2D to VR Translation Domains

The scope of potential translation extends far beyond individual user interface (UI) components. It includes interface structures, interaction models, navigation strategies, content consumption formats, narrative techniques, social communication patterns, and productivity workflows.

The following overview presents examples of directions of transformation that illustrate the scale and diversity of possible design approaches (see Table 16.1). This list is intentionally exploratory and serves as a conceptual map of the territory. In the following sections of this chapter, selected patterns will be examined in depth.

The examples presented above demonstrate that translating 2D experiences into VR involves not only interface components but entire models of user action. In the following sections, selected patterns will be examined in detail, with particular attention to their perceptual foundations, design rationale, and implications for comfort and immersion.

TABLE 16.1 Selected examples of interface structure translation from framed screen layouts to immersive spatial organization in virtual reality

	2D ELEMENT	VR TRANSLATION
A. Interface structure	Application window	Spatial workspace
	Overlapping windows	Distributed around the user
	Tabs	Spatial zones
	Sidebar	Body-anchored menu
	Top bar	Environmental panel
	Dropdown	Cluster of selectable objects
	Modal window	Attention-focused environmental event
	Dashboard	Data landscape
	Card grid	Objects arranged in a physical grid
	Infinite canvas	Information field explored through movement
	Toolbar	Tools represented as physical objects
	HUD	Context-sensitive information layer

(Continued)

TABLE 16.1 (Continued) Selected examples of interface structure translation from framed screen layouts to immersive spatial organization in virtual reality

	2D ELEMENT	VR TRANSLATION
B. Interaction	Cursor	Raycast or gaze-based interaction
	Hover	Focus through proximity or visual attention
	Click	Physical press or touch
	Double click	Confirmation gesture
	Drag and drop	Grasping and placing objects
	Scroll	User movement through space
	Zoom	Physical approach or scale transformation
	Slider	Lever or rotary control
	Checkbox	Physical toggle switch
	Form	Sequence of spatial actions
	Copy-paste	Object duplication
	Undo	Time or environment state reversal
	Loading bar	Environmental transformation
C. Navigation	Page scroll	Movement between spatial zones
	Pagination	Teleportation to the next node
	Breadcrumb	Spatial trail
	Link	Portal or passage
	Category tree	Branching path
	2D map	Map room
	Infinite feed	Stream of objects
	Map zoom	Environmental scale transformation
D. Content consumption	Article	Knowledge environment
	Slide presentation	Presentation stage
	Video	Spatial event
	Image gallery	Exhibition space
	Timeline	Time corridor
	Table	Interactive data grid
	Chart	3D data sculpture
	Survey	Journey of questions
E. Narrative	Camera frame	Attention choreography
	Editing cut	Environmental transition
	Close-up	Attention attraction through light or sound
	Chapter	Spatial transformation
	Subtitles	Spatial text layer
	Highlight	Lighting or motion emphasis
F. Social interaction	Progress bar	Physical path or distance
	Chat window	Spatial communication point
	Emoji	Avatar micro-gesture
	User profile	Embodied representation
	Online status	Presence in space
	Friends list	Social map
	"Like" reaction	Gesture of approval
	Comments	Layered conversations in space

(Continued)

TABLE 16.1 (Continued) Selected examples of interface structure translation from framed screen layouts to immersive spatial organization in virtual reality

	2D ELEMENT	VR TRANSLATION
G. Workflow and productivity	Task list	Spatial task board
	Kanban	Project wall
	Calendar	Time room
	Folder	Spatial archive
	Desktop	Immersive studio
	Keyboard shortcuts	Gestures or rapid interactions
	Industrial dashboard	Control room
	Spreadsheet	Data represented as manipulable objects
	LMS	VR campus
	PROBLEM	WHY IT IS PROBLEMATIC
H. Anti-patterns (to avoid)	Large flat panel in space	Replicates a screen instead of using spatial affordances
	Very small distant text	Loss of readability and visual fatigue
	Scrolling floating panels	Ignores spatial navigation potential
	Excessive HUD elements	Cognitive overload and reduced immersion
	1:1 desktop replication	Lack of adaptation to 3D perception
	Interactions requiring continuous arm lifting	Rapid physical fatigue
	Ignoring neck comfort and field of view	Discomfort and reduced usability
	Too many elements at the same depth plane	Spatial clutter

CANONICAL PATTERNS OF 2D-TO-VR EXPERIENCE TRANSLATION

Having defined the fundamental principles of translating user experience between screen-based and immersive environments, we can now move to the analysis of specific design patterns. The following set does not constitute a comprehensive catalog of all possible transformations. Rather, it presents a selection of examples considered fundamental from the perspective of Extended Reality (XR) experience design.

What unites these patterns is a shared characteristic: each illustrates a shift not merely in interface form, but in the underlying model of perception, action, or spatial understanding. They may be regarded as canonical because they stem directly from the differences between framed and embodied experience discussed earlier in this chapter.

From Framed View to Embodied Presence

In screen-based environments, the user remains an observer of the represented world. The monitor or mobile screen acts as a boundary separating the viewer from the content. Design largely revolves around controlling the frame and its contents.

In VR, this boundary dissolves (see Figure 16.1). The user no longer watches the world—they inhabit it. Translation therefore does not involve enlarging the image or expanding the field of view. Instead, it requires redefining the relationship between the user and the environment.



FIGURE 16.1 From observing a world through a frame to inhabiting it.

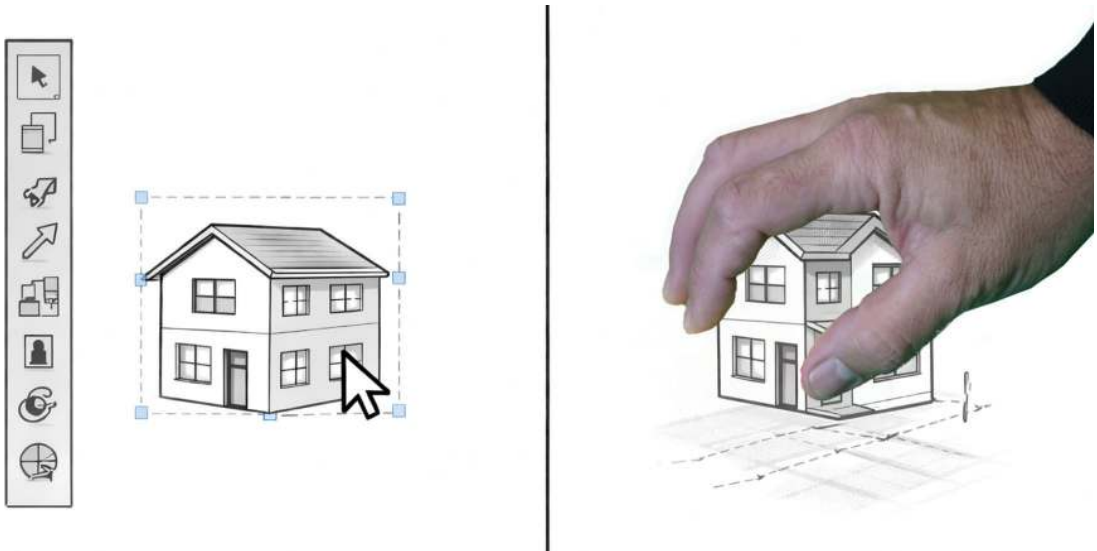


FIGURE 16.2 Symbolic command becomes physical action.

For example, a map presented in a 2D application may become, in VR, an explorable space in which the user moves between objects or locations. Similarly, an architectural model ceases to be a rendered object viewed from the outside and instead becomes a place where scale, proportion, and spatial relationships are directly experienced.

A common mistake is attempting to preserve the observer position by placing a large virtual screen in front of the user. Such a solution reduces VR to a cinematic projection and fails to leverage its perceptual and experiential potential.

Cursor Mediation and Direct Manipulation

Interaction in 2D systems is inherently indirect. The cursor functions as a symbolic intermediary between the user and digital objects. A mouse click does not correspond to physical contact with an object but instead represents an abstract command translated by the interface.

In VR, interaction can shift toward direct manipulation (see Figure 16.2). Users are able to reach for objects, move them, rotate them, or bring them closer through embodied action. The process of 2D-to-VR translation therefore involves moving away from pointer-based metaphors toward interaction grounded in bodily engagement.

A clear example can be found in 3D modeling workflows. Instead of selecting tools with a cursor and adjusting parameters through interface panels, users may deform or reshape a form directly using their hands or gestures within space.

A common anti-pattern is the simulation of a traditional cursor as the dominant interaction mechanism in VR environments. Such solutions reintroduce indirect control and often reduce the sense of natural action, limiting the experiential advantages offered by immersive interaction.

From Scroll to Spatial Progression

Scrolling is one of the fundamental mechanisms for exploring content in screen-based systems. It allows information to be revealed in a controlled manner within the limited boundaries of the frame.

In VR, scrolling loses much of its perceptual logic because the user can change their position relative to the surrounding environment (see Figure 16.3). The core of 2D-to-VR translation in this case involves replacing the movement of content with the movement of the user.

An article may become a sequence of knowledge stations. An exhibition may transform into a series of rooms. An educational course may take the form of a path guiding the user through successive tasks. Instead of advancing by moving a scroll bar, users progress by walking, turning, or navigating through space.

A common mistake is the introduction of large, scrollable panels floating in mid-air. Such solutions often require continuous arm elevation and quickly lead to physical fatigue, while also failing to take advantage of spatial navigation as a primary interaction paradigm.

From Hyperlink to Portal-Based Transition

In 2D environments, a hyperlink functions as an abstract shortcut leading to different content. A click results in an immediate change of context, often without any sense of physical continuity between the previous and subsequent states. The transition is logical rather than experiential.

In VR, a change of context can take the form of a spatial transition (see Figure 16.4). A portal, doorway, or transformation of the surrounding environment can preserve a sense of continuity within the experience. Instead of abruptly replacing one screen with another, the user moves through a threshold that connects spaces in a perceptually coherent way.

This approach is particularly effective in educational and immersive narrative contexts, where a change of topic may be experienced as entering a new world, historical period, or conceptual domain. By spatializing transitions, designers can maintain presence while still supporting shifts in content and meaning.

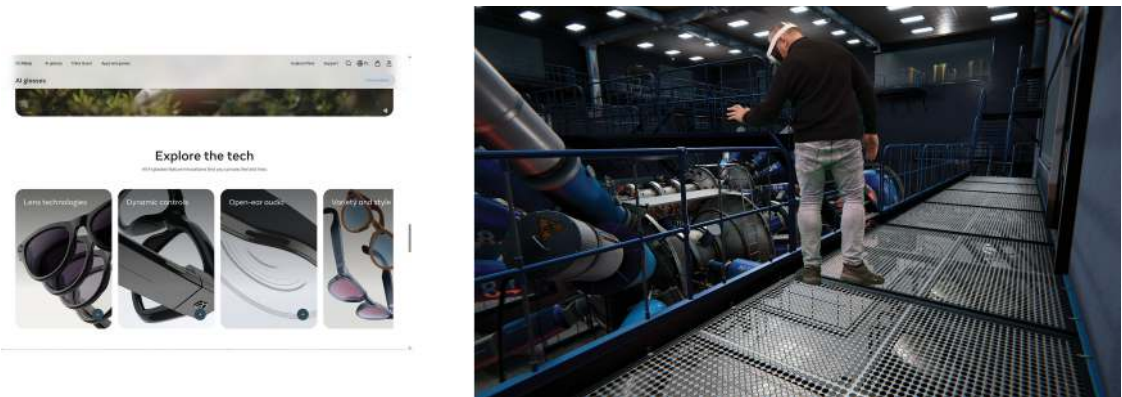


FIGURE 16.3 Evolution of digital interaction: from linear content scrolling to spatial navigation and exploration in immersive environments.

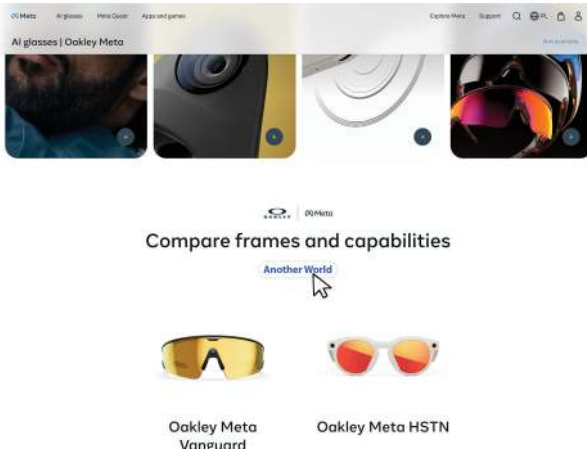


FIGURE 16.4 Context change becomes spatial transition.

From Dashboard to Environmental Data Representation

Screen-based dashboards compress large amounts of information into a limited visual area. Charts, indicators, and numerical summaries are organized to maximize visibility within the boundaries of a single display. In VR, however, data can be distributed throughout space rather than condensed into a flat surface (see Figure 16.5).

Instead of viewing charts from a fixed position, users may move between data representations, examine relationships from multiple perspectives, or manipulate scale to explore patterns at different levels of detail. Information becomes something that can be navigated and experienced spatially rather than simply observed.

Examples include visualizations of industrial processes or climate data represented as dynamic landscapes of interdependent variables. Such environments allow users to perceive trends, correlations, and anomalies through movement and spatial comparison.

A common mistake is retaining the dashboard as a large flat panel suspended in space, effectively enlarging the monitor without changing the interaction model. This approach fails to take advantage of the spatial and perceptual affordances offered by immersive environments.



FIGURE 16.5 Transformation of data visualization from screen-based dashboards into navigable spatial environments.



FIGURE 16.6 Transformation from window-based multitasking to spatially organized workspaces in immersive environments.

From Window Multitasking to Spatial Workspace

Multitasking in desktop systems relies on switching between windows and managing their overlap. Users navigate tasks by bringing one window to the foreground while minimizing or hiding others, often depending on visual stacking and tab switching to maintain workflow continuity.

In VR, designers can leverage the user's spatial memory as an organizing principle (see Figure 16.6). Information can be distributed around the user in stable spatial locations, allowing tasks, documents, or tools to be remembered based on where they are positioned. Rather than stacking windows in a single visual plane, content can occupy distinct regions in the surrounding environment.

This approach takes advantage of humans' natural ability to orient themselves in space and recall locations relative to their body. Spatial arrangement becomes a cognitive aid rather than a constraint.

A common anti-pattern is simulating a stack of overlapping windows in a single area of the user's field of view. Such solutions replicate desktop conventions without utilizing the spatial affordances that immersive environments make available.

From Camera Framing to Attention Choreography

Film and screen-based interfaces guide attention through framing. The camera determines what is visible and what remains outside the field of view, allowing designers to control composition, focus, and sequencing with precision.

In VR, users are free to look in any direction at any time. The absence of a fixed frame means that attention can no longer be enforced through cropping or viewpoint control (see Figure 16.7). Instead, design must rely on subtler mechanisms to guide perception.

Lighting, movement, directional sound, environmental contrast, or character behavior can all be used to suggest where attention should be directed. Rather than forcing the user to look at a predetermined point, the environment invites attention through perceptual cues.

This shift requires thinking less in terms of composition within a frame and more in terms of choreographing attention across space.



FIGURE 16.7 Attention is guided, not framed.



FIGURE 16.8 Notification becomes event.

From Notification to Environmental Event

In screen-based systems, notifications typically interrupt the experience through dialogue boxes, pop-up messages, or system alerts. These elements temporarily override the current interface state, drawing attention by breaking continuity.

In VR, notifications can be reinterpreted as environmental events (see Figure 16.8). A change in lighting, a directional sound cue, the appearance of an object, or a reaction from an avatar can communicate information without introducing foreign interface layers. Instead of overlaying the experience, the system responds within it.

This approach helps preserve continuity and maintain immersion. Rather than interrupting the user with a detached interface element, the environment itself becomes the medium through which information is conveyed.

From Timeline to Spatialized Time

In 2D systems, a timeline functions as an abstract axis used to organize events, processes, or historical sequences. Users interpret temporal progression through symbols such as markers, labels, or linear diagrams presented on a flat surface.

In VR, time can be represented spatially (see Figure 16.9). Instead of observing progression as a graphic abstraction, users may move through successive stages of a process, experience historical transformations as environmental changes, or explore the evolution of a project by navigating between spatially arranged moments.

Such approaches are particularly effective in educational applications, museum experiences, and process analysis, where understanding change over time benefits from embodied exploration rather than symbolic representation alone.

From Interface Controls to Tool-Based Interaction

In 2D applications, toolbars require users to select a function before performing an action. Interaction typically follows a sequence: choose a tool, apply it, and adjust parameters. The control interface precedes the action itself.

In VR, tools can exist as objects with physical affordances (see Figure 16.10). A flashlight enables exploration, a marker supports annotation, and a scalpel allows structural manipulation. Instead of selecting commands from a panel, users engage with instruments that embody specific capabilities.



FIGURE 16.9 Time becomes a place.

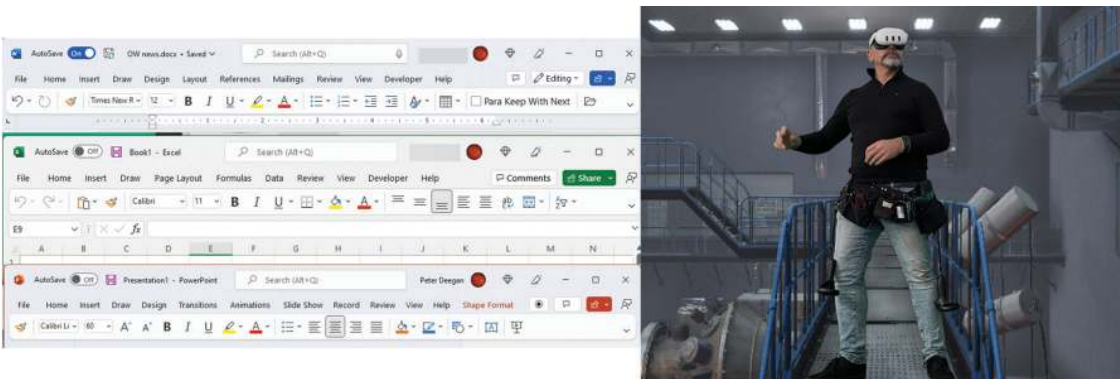


FIGURE 16.10 Tools replace commands

This approach reduces interface abstraction and builds upon users' existing experience with handling physical objects. By embedding functionality within manipulable tools, interaction becomes more intuitive and aligned with embodied action rather than symbolic command selection.

The examples presented above demonstrate that 2D-to-VR experience translation fundamentally involves redefining the relationship between the user, information, and the space of action. In the following sections of this chapter, selected patterns will be examined in greater detail, including practical application examples as well as their implications for comfort, immersion, and cognitive effectiveness.

A MODEL OF EXPERIENCE TRANSFORMATION: FROM ABSTRACTION TO EMBODIMENT

At first glance, the previously described 2D-to-VR translation patterns may appear to be a collection of independent examples. Some address interface structures, others focus on narrative, interaction techniques, or information organization. In reality, however, they all emerge from the same fundamental shift: the transition from screen-based experience to immersive experience (see Figure 16.11).

2D systems were designed around the constraints of a limited display surface. Content is observed from the outside, actions are performed indirectly, and user attention is controlled through composition and visual hierarchy. VR challenges these assumptions. The user is no longer a spectator looking at the world through a screen frame but becomes a participant situated within the environment itself.

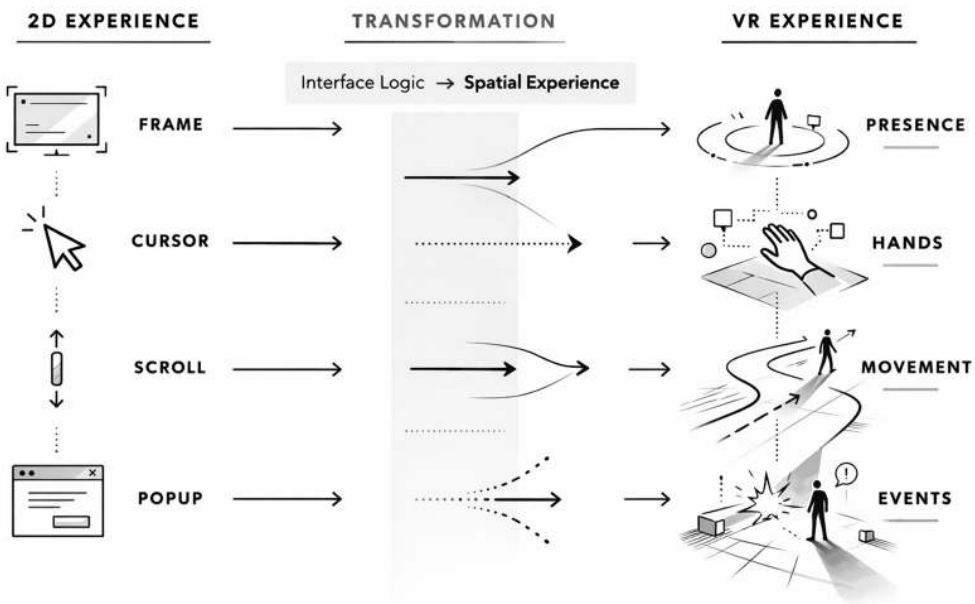


FIGURE 16.11 Core transformation from screen-based interaction logic to immersive spatial experience.

This transformation can be understood as a shift in design across several interconnected dimensions. These dimensions are not rigid categories but rather complementary perspectives for understanding the problem of 2D-to-VR experience translation.

- **From Observation to Presence:** The most visible difference between screen-based media and VR concerns the relationship between the user and content. In 2D systems, the user observes the represented world. Even in interactive applications, they remain outside it—controlling it from a safe distance.
- In VR, this relationship is reversed. The user is situated within the experience itself. The scale of space, the distance between objects, and the direction of gaze cease to be elements of representation and instead become elements of lived experience. For this reason, experience translation does not consist of enlarging the image or adding stereoscopic effects. An architectural model previously viewed on a screen may become a place that can be walked through. A film ceases to be a framed sequence and becomes a situation in which the user participates. The designer no longer composes an image but designs presence.
- **From Symbols to Physical Action:** A second shift concerns the nature of user action. Screen-based interfaces rely on symbols. A mouse click has no physical equivalent—it is a learned convention within digital systems. VR allows a return to more intuitive forms of interaction. Users can reach for objects, rotate them, open doors, or use tools. Instead of selecting a function from a list, they perform an action that resembles real-world experience. For example, a “Next” button in an educational application may be replaced by moving into a new space. A selection tool in a graphic application may take the form of a physical marker. This does not imply that all actions must mimic reality. Rather, the essential point is that VR reduces the distance between intention and execution, grounding interaction in embodied action.
- **From Visual Hierarchy to Spatial Organization:** Screen-based systems manage large amounts of information through visual hierarchy. Layout, typography, and color are used to structure content within a limited surface. In VR, information can be distributed throughout space. Distance, height, orientation, and scale begin to function as organizing principles. An analytical dashboard no longer needs to consist of multiple charts compressed into a single view. Data can surround the user as a spatial field of relationships. Similarly, multiple application windows can be replaced by stable spatial anchors positioned around the user, supporting recall through spatial orientation. Design shifts from compressing information into a frame to leveraging humans’ natural capacity for spatial navigation and environmental awareness.
- **From Context Switching to Experiential Continuity:** In screen-based environments, context changes occur instantaneously. Clicking a link opens a new page, and a new window replaces the previous one. VR makes it possible to preserve experiential continuity. A change of topic may take the form of passing through a portal, opening a door, or witnessing a transformation of the environment. A notification need not interrupt the experience through a pop-up window; it may be communicated through light, sound, or the appearance of an object. Such strategies prevent abrupt disruptions and help maintain a sustained sense of presence.
- **One Transformation, Multiple Patterns:** Although the previously described patterns differ in application, they can all be understood as expressions of the same underlying shift: a movement away from abstract interface conventions toward experiences grounded in space, action, and perception. 2D-to-VR translation is therefore not a process of copying interfaces or rendering them in three dimensions. It is a reinterpretation of how user intentions are realized within an environment where body, movement, and attention are as central as graphics and information structure. Understanding this transformation allows VR to be treated not as another display platform, but as a distinct design medium with its own principles and its own experiential language.

CONCLUSION

The aim of this chapter was to structure a way of thinking about 2D-to-VR experience translation so that it is not reduced to the conversion of interface elements, but understood as a design process grounded in a fundamentally different model of perception and action. We have shown that screen-based environments construct a framed experience, in which designers control attention through the frame and users act indirectly through symbolic interface conventions. VR, by contrast, introduces an embodied experience, where the body, movement, spatial orientation, and attention dynamics become primary design materials.

In defining translation, we identified five analytical levels: user intention, task logic, information structure, attention dynamics, and perceptual model. This shift in perspective allows designers to move from asking “How should the interface be transferred?” to asking “How should the goal and meaning of action be reinterpreted?” The proposed structure for describing translation patterns provides a systematic framework for comparing solutions across application domains.

We then outlined a broad spectrum of potential translations to demonstrate that the challenge extends beyond UI components to include narrative structures, social interaction, workflow organization, and modes of information exploration. From this landscape, 10 canonical patterns were selected to illustrate the fundamental transformation of the medium: from frame to presence, from cursor mediation to direct manipulation, from scrolling to spatial progression, from hyperlink to portal-based transition, from dashboard to environmental data representation, from window switching to spatial workspace, from camera framing to attention choreography, from notification to environmental event, from timeline to spatialized time, and from interface controls to tool-based interaction.

The central conclusion is that all these patterns can be understood as expressions of a single transformation: a reduction of interface abstraction in favor of spatial and embodied solutions that better align with human capabilities and constraints in immersive environments. Importantly, effective translation does not require literal simulation of the physical world. Instead, it should leverage what VR uniquely offers as a medium: presence, embodied action, spatial organization, and experiential continuity.

In further development of this chapter—or in subsequent design work—these patterns can serve as a starting point for informed prototyping and usability testing. Ultimately, the success of 2D-to-VR translation is not determined by visual sophistication, but by whether users can realize their intentions in a manner that feels natural, comfortable, and perceptually coherent within immersive space.

REFERENCES

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in psychology* (Vol. 52, pp. 139–183). North-Holland.
- Hassenzahl, M. (2010). *Experience design: Technology for all the right reasons* (Vol. 8). Morgan & Claypool Publishers.
- Kiltner, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4), 373–387.
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Nielsen, J., & Molich, R. (1990, March). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seattle, WA, United States (pp. 249–256).
- Norman, D. (2013). *The design of everyday things*. MIT Press.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. (3rd ed.). MacGraw-Hill.

- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549–3557.
- Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3, 74.
- Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997, October). Cybersickness is not simulator sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41(2), 1138–1142.
- Stanney, K. M., Lawson, B. D., & Oman, C. M. (Eds.). (2021). *Cybersickness in virtual reality versus augmented reality*. Frontiers Media SA.

Index

Note: **Bold** page numbers refer to tables; *italic* page numbers refer to figures.

- ability spectrum 164, 166–168, 171
- abstract interface conventions 230–231
- acceleration profiles 87, 92
- accessibility 73, 83, 98, 156, 199–200
 - definition of 164
 - in immersive environments 164–165, 171
 - inclusivity 152–153, 156
 - mapping of barriers 168
 - in 2D interfaces 164–165
 - visual and auditory 169
- accommodation 32
 - ciliary muscles 32
 - crystalline lens 22, 32
 - refractive power 32
 - short range 32
- acoustics 54, 55
 - human audio perception 55
 - physical properties 55
- adaptive interfaces 123–124
 - behavioural signals 124
 - bio-responsive systems 124
- adaptive speed control 85
- addiction 148; *see also* psychological impacts
- affective metrics 203, 210–213
 - facial expression tracking 210
 - physiological measurements 210–213
- affordances 46, 48, 50, 119, 177, 182
 - definition 46
 - interactive 48
 - perceived 46, 50
 - spatial 46, 136
- agency 116, 118–120, 131–132, 178, 182, 193–194, 198, 200
 - degrees of control 131
 - influence on narrative 131–132
 - user-scene interaction 131
 - user-technology interaction 131–132 (*see also* interactivity; ludonarrativism; player role)
- Akenine-Möller, T. 49–52, 104, 107
- allocentric frame of reference 20
 - environmental landmarks 20
 - memory 19, 20, 24
 - object location 20, 28, 30
 - stable map 19, 20
- alternative input methods 73; *see also* eye-tracking;
gaze-based interaction; multimodal interaction;
voice commands
- alternative inputs 168; *see also* interaction remapping
- ambient lighting 39, 42
- ambient occlusion (AO) 26, 39, 50
- ambient sound *see* sound design
- ambisonics 56, 57, 59
 - B-format 57
 - computational load 59
 - First-Order Ambisonics (FOA) 59
 - Higher-Order Ambisonics (HOA) 54, 59
- Ames Room 23
 - environmental geometry 23, 24, 31
 - monocular peephole 23
 - perspective consistency 23
 - size-distance relationships 23, 24
 - trapezoidal room 23 (*see also* optical illusion)
- amplitude 55
 - loudness 55
- anticipation principle 108, 110, 112
- anti-patterns 222–228
- appeal principle 113
- arcs principle 109–111
- arm-swing locomotion 87
- artifacts (cultural) 133
 - cultural references 133–134
 - definition 133
 - VR as cultural product 133
- art (visual) 132, 136
 - environmental storytelling 135–136
 - expressive value in VR 136
 - scene composition 132
 - symbolic use of colour 136
- artificial intelligence (AI) 54, 64–65, 74, 159–160
 - audio deepfakes 54, 65, 66
 - authorship 159
 - generated content 159–160
 - procedurally generated soundscapes 64
- artificial reality 5, **12**
- artistic resonance 112
- asset management 192, 197, 200
- atmospheric gradients 26
 - aerial perspective 26
 - bluish tint 26
 - desaturated 26
 - horizon 22, 26, 28, 29
 - Rayleigh scattering 26
 - shorter wavelengths 26
- atmospheric perspective 38
- attention 34–37, 39, 43
- attention choreography 226–227, 231
- attention distribution 210
- attention dynamics 218–219, 231
- attention guidance 183, 196, 200
- audio description 164, 169
- audiovisual integration 60–61
 - latency constraints 54, 60, 61
 - temporal synchronization 60, 61
- augmented accessibility *see* accessibility

- auricularization 129, 134
 - attention guidance 132, 134
 - immersion 134
 - narrative meaning 134
 - sound perspective 134
- autonomy 39, 43, 151
 - emotional manipulation 151
 - informed consent 151–152
 - user control 152–153
- avatars 82, 135, 140–141, 145, 148
 - customization 141
 - identity expression 141
 - movement tracking 141, 143 (*see also* embodiment; player role; Proteus effect)
- background 36–37, 41
- Banakou, D. 46
- barrier classification frameworks 170
- battery dependence 79; *see also* controllers: limitations; sustainability
- behavioural change 156–158
 - sustainability education 156–157
- behavioural metrics 203, 207–210
 - eye-tracking 209–210
 - interaction logs 209
 - navigation behaviour 208–209
- Bernini, G. L. 37, 41
- Biederman, I. 48
- binaural audio 54, 56–58
 - dummy head models 57
 - ego-centric perspective 54, 56–57
 - Head-Related Transfer Functions (HRTFs) 54, 56–58
- binocular disparity 19, 20, 23, 31
 - cross-correlation 20
 - horizontal shift 20
 - spatial layout 20, 24 (*see also* retinal disparity; stereopsis)
 - three-dimensional layout 30
- biometric data 147–148, 152, 160
 - data protection 160
 - privacy 152
- body-fixed reference frame 93
- body-integrated interfaces 118–120
 - peripersonal space 120
 - wrist-mounted menus 118–119
- body orientation 218
- body ownership 183
- breaks in presence (BIPs) 46, 48, 51
- Bruner, J. 129
 - cognitive significance 129
 - narrative construction 129
- buffer zone 37, 43; *see also* volumetric negative space
- calibration flows 169, 171
 - comfort 169
 - height 169
 - reach 169
- camera framing 221, 226–227, 231
- camouflage 37
- canonical patterns 222, 231
- carbon footprint 155–156
 - energy consumption 155
 - travel reduction 155–156
- Carson, D. 47
- cast shadows 39, 43
- category tree 221
- Cave Automatic Virtual Environment (CAVE) 7, 8, 12
- central vision 37
- Chang, D. 34
- characters 131, 136
 - animation as narrative meaning 136
 - NPC functions 136
 - protagonist vs avatar 135
 - roles 131
 - uncanny valley 136
- checklists 186–187
- choice mechanics *see* decisions
- chromatic key 36
- chromatic optimisation 41; *see also* typography
- chromatic tones 37
- Cifuentes-Albeza, R. 35
- cinematics 135; *see also* narrative tools
- circular economy 155
 - hardware lifecycle 155
 - recycling and reuse 155
- climate change 156–158
 - sustainability education 156–157
- climbing-based locomotion 87
- closure
 - law of 38
 - principle of 35, 35, 38
- co-design 143, 145
 - CakeVR 143, 145
- cognitive capacity 34
- cognitive load 35, 37–39, 43, 47, 49, 72, 116, 126, 179, 183, 202, 205, 210–213
 - EEG 211–212
 - intrinsic vs. extraneous 118
 - multilayered UIs 118, 124
 - NASA Task Load Index (NASA-TLX) 204–205 (*see also* Sweller, J.)
- cognitive maps 116, 118, 121
 - navigation cues 119–120
 - world-locked landmarks 124
- coherence 17, 18, 24
 - spatial immersion 18 (*see also* place illusion (PI))
- colour perception 37–38
- comfort 85–100, 183, 193–194, 198–200
- comfort zone 40
- common fate
 - comparison with continuity 38
 - definition 38
 - principle of 35, 38–39, 43
- communication modalities 142
 - multimodal interaction 142
 - nonverbal communication 142
 - voice interaction 142
- community participation 153
 - cultural sensitivity 153
- computational budgets 192, 196
- conflict 130; *see also* narrative structure; plot
- context switching 230
- continuity
 - comparison with common fate 38
 - comparison with similarity 36

- law of 40, 43
- principle of 35, 36, 38–39
- continuous locomotion 87
- continuous turn 92, 94
- contrast
 - chromatic 40, 43
 - principle of 35–38, 40–41
- controllers 36
 - characteristics 72, 78
 - limitations 79
- convergence 20, 32
 - binocular image 20, 29, 30, 32
 - extraocular muscles 32
 - inward rotation 20, 32
 - muscle tension 32
- CPU 196, 200
- CPU-GPU balance 192, 196
- Cronbach's alpha 190
- cross-view 29
 - cross-eyed viewing 29
 - fixation point 28, 30
 - stereo pairs 29
 - three-dimensional composite 29
- cultural context 133–134
 - artifacts 133
 - impact on interpretation 133–134
 - intertextuality 133–134
 - mindfacts 133
 - sociofacts 133
- cultural sensitivity 151, 153
 - inclusivity 152–153
 - social justice 157
- cursor 221, 223–224, 231
- cybersickness 13, 28, 54, 59–61, 64, 83, 85, 89, 104, 152, 161, 167, 169, 178, 183, 193–194, 199, 201, 207
 - contradictory data 28
 - health risks 161
 - motion sickness 104, 111
 - perceptual conflicts 64
 - sensory conflict 28
 - sensory mismatch 104
 - user comfort 28
 - user wellbeing 152
 - vestibular-ocular conflict 61
 - vestibular system 20, 28
 - vignetting 104
 - visual-vestibular discrepancy 104 (*see also* motion sickness; motion sickness questionnaires)
- dashboard 220, 225–226, 231
- DataGlove 6–7, 12
- data protection 152, 160
 - GDPR 160
- data security 147–148
- decisions 136
 - binary, multiple, open 136
 - emotional engagement 136
 - narrative consequences 136
 - personalization 136
- depth 219
- depth cues 17–21, 23, 24, 26, 29, 31, 34, 37, 39
 - atmospheric perspective 38
 - cast shadows 39
 - distance judgment 20, 24, 26–28, 32
 - motion parallax 38
 - occlusion 38–39, 43 (*see also* spatial situation model (SSM); pictorial cues)
- depth perception 38–39
- design justice 153
 - community participation 153
 - social sustainability 156
- dialogues (interactive) 136
 - branching options 136
 - character relationship impact 136
 - consequences 136
- dialogues (non-interactive) 136
 - context delivery 136
 - exposition 136
 - guided information 136
- diegetic interfaces 116–122, 124, 183
 - in-world existence 119
 - narrative justification 119, 121
 - shared perceptibility 119
- Dietrich, O. 40
- diffusion gradients 39
- digital divide *see* accessibility
- digital storytelling (DST) 130
 - educational uses 130
 - multimedia integration 130
 - VR adaptation 130
- directional vectors 38, 40
- direction memory 94
- direct manipulation 223–224, 231
- diversity 152–153, 157
 - representation 152
- dynamic framing 39
- dynamic reading 42
- dynamic speed reduction 92
- economic sustainability 157–158
 - financial viability 157
 - SROI 157
- education 140, 143–145
 - distance education 144
 - experiential learning 144
 - higher education 144
 - Mozilla Hubs 144
- egocentric frame of reference 17–20
 - motor-action planning 20
 - observer origin 20
 - primary reference frame 18
 - (*see also* depth cues)
- egocentric space 40
- electrodermal activity (EDA) 211–212
- electroencephalography (EEG) 211–212
- electronic waste (e-waste) 155
 - circular economy 155
- embodied action 223, 228, 230–231
- embodied cognition 178
- embodied interaction 85, 135, 176–179; *see also* presence; agency; mechanics
- embodied presence 222–223, 231

- embodiment 13, 80–82, 104, 105, 115–120, 123, 125, 135, 141, 143, 144, 148, 179, 182–183, 193–194, 198–199
 - avatar association 135
 - body-centred reference frames 119–120
 - body ownership 116, 118
 - full-body avatars 104
 - immersive identification 135
 - narrative influence 135
 - self-location 116
 - self-representation 105
 - synchrony 135
 - tracked controllers 104
 - user role 135
- emergent narrative 135
 - co-creation 135
 - system-driven storytelling 135
 - unpredictability 135
 - VR suitability 135
- emotional-cognitive engagement 81, 83
- emotional engagement 130, 136
 - empathy 136
 - immersive amplification 136
 - meaning-making 130
- empathy 156–157
 - global awareness 156–157
 - immersive storytelling 154
- energy consumption 155
 - green energy 155
- energy efficiency 199
- environmental context 167–168
- environmental cues 177, 181, 218, 220, 227
- environmental events 227–228, 230–231
- environmental projections 118–119
 - holographic displays 119
 - large-scale information layout 119
 - spatial annotations 119
- environmental storytelling 45, 47, 135–136
 - diegetic world-building 135–136
 - non-linear information discovery 135
 - spatial cues 136 (*see also* space (VR); world-building)
- environmental sustainability 155–156
 - carbon footprint 155–156
- ergonomics 42–43, 73
- ethical evaluation 157–158
- ethical principles 151–153
 - autonomy 151
 - justice 151
- ethics 54, 65, 66
 - audio deepfakes 54, 65, 66
 - intellectual property (IP) 65
- evaluation procedure 185
- experience
 - principle of 35, 35, 42
 - user 36; VR 38
- experience translation 218–220, 231
- experience types in VR 133
 - applications 133
 - interactivity levels 133
 - multimedia content 133
 - video games 133
- experiential continuity 224–225, 230–231
- experiential quality 192–194
- expert evaluation 177, 189
- extended reality (XR) 13, 176–177
- eye-tracking 37, 209–210
 - cognitive load assessment 210
 - gaze heat maps 209
 - visual attention 209–210
- face-to-face interaction 140, 143–144, 148
- facial expression tracking 210
 - affective responses 210
 - Meta Quest Pro 210
- familiarity effects 75
- familiar size 19, 24
 - cognitive heuristics 24
 - perceptual anchors 17, 24
 - size constancy 19, 24
 - spatial ambiguities 24
 - stored knowledge 24
- fatigue 73, 79–80; *see also* ergonomics
- feedback 103, 110, 111, 113, 182–186
- field of view 9, 13
- figure-ground 36, 37
 - contrast 37
 - definition 36
 - depth cues and lighting as guidance 37
 - volumetric negative space 43
- focalization 129, 134
 - external 134
 - fixed, multiple, variable 134
 - internal 134
 - POV construction 134
 - VR adaptations 134
 - zero 134
- focal points 39
- follow-through 108–110
- forms 221
- forward kinematics (FK) 104; *see also* kinematics
- fovea centralis 37
- foveated rendering 37, 77; *see also* eye-tracking
- fragmented platforms 147
- frame budget 51, 194–196
- framed experience 218, 222, 231
- frame rate 193–195, 200
- frame time 194–196
- framing 34, 39
 - the designer as architect of attention and hierarchy 43
- frequency 55
 - acoustic shadowing 55
 - pitch 55
- frontal field 37
- full-body motion capture 80
- full-body tracking 72
- Gaggioli, A. 39
- gaming and entertainment 143, 146
- gaze 37, 39, 40, 43
- gaze-based interaction 72
- gaze-based selection 77
- gaze behaviour 215
 - eye-tracking 209–210
 - visual attention 209–210
- gaze heat maps 209
 - eye-tracking 209

- navigation analysis 208
- Genette, G. 134
 - focalization theory 134
- Gestalt Psychology
 - overview 34
 - principles of 35, 35
- gesture-based interaction 72, 79–80
- gesture recognition 77, 143
- Gibson, J. J. 46, 52
- gigantism 41
- global illumination (GI) 50
- gorilla arm 80
- Grau, O. 45
- guardian systems 98
- guiding lines 40

- handheld devices (augmented reality) 77
- hand tracking 72
- haptic feedback 72, 73, 78
- hardware-assisted locomotion 88
- hardware lifespan 199
- head-locked HUD 117
 - camera-fixed overlays 117
 - persistent screen layer 117
 - viewport attachment 117
- head-locked interfaces 117, 121–122
 - depth conflicts 117, 122
 - HUD overlays 118
- head-mounted displays (HMDs) 1, 4–7, 71, 75–76, 140, 143–145
- Head-Related Transfer Functions (HRTFs) 54, 56–58
 - personalized measurements 58
 - pinna filtering 58
- health and safety 161–162
 - cybersickness 161
- healthcare 140, 145
- heart rate variability (HRV) 211
- hero object 50, 52
- heuristics 175–180
- hierarchy
 - primary 40
 - secondary 40
 - spatial 34, 36–37, 39–40, 43
- Hillier, B. 47
- Hoffman, D. M. 49
- Holmes stereoscope 2, 12
- horizon space 40
- horopter 30, 31
 - imaginary spatial curve 30
 - lateral separation 30
 - zero retinal disparity 30
- Howard, P. 45
- human capabilities 164, 166
- human-computer interaction (HCI) 205
- human perception 34–35, 43
- human rights 153–154
- hybrid interfaces 123
 - combined diegetic and world-locked layers 115, 123
 - task-critical panels 121, 123, 125
- hybrid user role 132, 134; *see also* player role; prosumer
- hyperlink 221, 224–225, 231

- iconicity 36
- immersion 1–2, 6, 10, 13, 17, 18, 23, 24, 34, 38, 42–43, 45, 47, 48, 50, 51, 54–57, 60–61, 64–66, 116–117, 122–123, 125, 141–142, 144, 146, 148–149, 176, 182, 193, 220, 227, 229
 - cognitive dissonance 56, 59, 61, 66
 - environmental immersion 142
 - interface consistency with world logic 116, 119–122
 - perception of presence 61
 - plausibility 116–117, 119, 126
 - presence 115–120, 125–126
 - sensory stimuli 18
 - tactile immersion 142
 - technological properties 18
 - visual and auditory immersion 142
- immersive environments 73, 202–204, 207, 210, 213; *see also* presence; space (VR); VR narrative
- immersive interaction 201, 203, 207
- immersive storytelling 154
- impact evaluation 157–158
- inclusive design 54, 62, 63, 66, 164–166, 171, 199
 - adaptable systems 62
 - visual alternatives 62, 66
- inclusivity 73, 83, 146, 152–153, 156
 - accessibility 156
 - diversity 152
- indirect locomotion 87–88
- information density 124
 - clutter avoidance 120, 122, 124
 - layered content 120–121, 124
 - progressive disclosure 124
- information structure 219, 231
- inside-out tracking 11, 80
- interaction 34
 - space 40
 - zones 40
- interaction complexity 196
- interaction design 13, 72–73
- interaction fidelity 115–117, 126
 - action similarity to physical world 115–117
 - spatial and temporal realism 115–117
- interaction latency 100
- interaction naturalness 181–182
- interaction remapping 168, 171
- interactivity 131–133
 - agency relationship 131–132
 - narrative impact 131–133
 - VR specificity 131–133
- Interaural Level Difference (ILD) 54, 56, 57
 - head shadow effect 57
- Interaural Time Difference (ITD) 54, 56, 57
 - arrival time gap 57
- interface controls 228–229
- interoperability 147, 149
- interposition 21
 - depth ordering 21
 - overlapping 26
 - spatial hierarchy 21 (*see also* occlusion)
- interpupillary distance (IPD) 29
 - horizontal offsets 29
 - physical separation 29
 - VR HMDs 28, 30

- inter-rater reliability 190
- intertextuality 133–134
 - cultural references 133–134
 - shared expectations 134
 - user recognition 134
- Invariance
 - depth cues and lighting as guidance 37
 - principle of 34, 35, 37, 42
 - readability, distance, scale and typography in VR 41–42
- inventory tools 36
- inverse kinematics (IK) 104; *see also* kinematics
- iterative design 202–203
- iterative development 202–203, 205, 207, 213
- Jenkins, H. 49
- Jerald, J. 35, 38–40, 46–49
- Kanizsa's triangle 21
 - depth stratification 21
 - illusory contours 21
 - modal completion 21 (*see also* optical illusion)
- kinematics 102, 104
 - procedural IK 105 (*see also* forward kinematics (FK); inverse kinematics (IK))
- kinetic property 39
- kinetic realism 102, 110
 - cognitive overload 110
 - mass 111
 - mechanical consistency 111
 - physical laws 110
- Kooi, F. L. 49, 52
- latency 72–73, 146–147, 193–194, 198, 200
- LaViola Jr, J. J. 46
- learning curve 72, 74–75
- learning support 184
- legal responsibility 161
 - content regulation 161
- legibility 34, 41–42
- level of detail (LOD) 50, 51, 197, 200
- Liang, H. 36, 39
- lighthouse base stations 9
- lighthouse tracking system 11
- lighting 34, 37, 39, 40, 42–43, 45, 49–51, 193, 196–197, 200
 - cast shadows 39
 - directional 49
- linear perspective 20–24, 38
 - conical perspective 21, 22
 - converging lens system 21
 - geometric projection 21, 22, 24
 - vanishing point 22
- lived-in aesthetic 45, 47
- locomotion 40, 42, 103, 104, 111, 178, 183
- locomotion taxonomy 86–88
- ludic 18
 - engagement 17, 18
 - game mechanics 18
 - immersive layers 18 (*see also* immersion)
- ludofictional worlds 132
 - interactive world-building 132
 - narrative-mechanics link 132
- ludonarrative coherence 136; *see also* ludonarrativism
- ludonarrativism 131–132
 - dissonance 132
 - harmony 132
 - narrative-mechanic alignment 132
 - player agency 132
- luminance 39–40
- luminous contrast 37, 40
- MacNamara, W. 35, 38, 41–42
- maps 223, 225
- Mazhar, A. A. 34
- mechanics 132, 136
 - implicit storytelling 136
 - interactive meaning 136
 - movement 136
 - object manipulation 136
- memory 41–42
- memory management 192, 197
- mental effort 115–116, 118, 123
 - divided attention 118, 120–122
 - multitasking 120, 125
- mental map 87, 92
- mindfacts 133
 - beliefs 133
 - ideological context 133
 - myths 133
 - values 133
- Mine, M. R. 46
- mixed reality 76
- modelling 39
- moderation 147, 161
 - user safety 161
- Moloney, J. 39
- monocular cues 19–21
 - distance landscapes 20
 - static images 20, 24 (*see also* pictorial cues)
- morphological similarity 36, 38
- motion capture 105
 - AI-driven computer vision 105
 - biological motion 105
 - virtual rig 105
- motion parallax 20, 23, 27, 28, 38
 - dynamic depth cue 20, 26, 27, 29
 - head movement 18, 28
 - relative velocity 20, 27
 - retinal displacement 28
 - 6DoF tracking 28
- motion sickness 29, 38, 73, 74, 83; *see also* cybersickness
- motion sickness questionnaires 207
 - Simulator Sickness Questionnaire (SSQ) 207
 - Virtual Reality Sickness Questionnaire (VRSQ) 207
- motion-to-photon latency 193
- motion tracking 1, 5–7, 80, 140–143
- multimedia storytelling *see* digital storytelling (DST)
- multimodal interaction 72, 77, 142, 148
- multiple resource theory 125
 - attentional interference 125
 - parallel channels 125
 - task allocation 125
- multisensory consistency 176
- multisensory design 34
- multisensory integration 54, 60–62, 65

- audio-haptic feedback 54, 60–61, 65, 66
- multimedia systems 60, 62
- multisensory stimulation 3
- multitasking 226
- Murray, J. H. 47
- narration 129
 - cognitive function 129
 - cultural function 129
 - definition 129
- narrative competence 129–130
 - coherence 130
 - understanding vs production 129–130
- narrative elements in VR 132
 - images 132
 - movement 132
 - narrative space 132
 - sound 132
 - typography 132
- narrative models 135
 - amusement-park 135
 - branching 135
 - emergent 135
 - funnel/critical path 135
 - linear 135
 - multi-level 135
 - open systems 135
 - pearl-necklace 135
- narrative space 132
 - audiovisual convergence 132
 - expressive meaning 132
- narrative structure 130–131
 - external 130
 - internal 130–131 (*see also* plot)
- narrative tools 130, 136
- narrator 131, 134
 - user-driven perspectives 134
 - VR hybridization 134
- NASA Task Load Index (NASA-TLX) 170, 189–190, 204–205
 - Raw TLX (RTLX) 205
 - workload assessment 204–205
- natural interaction 72–73
- natural locomotion 86–88
- navigation 177, 181, 221, 224
- navigation behaviour 202, 208–209
 - navigation efficiency 208
 - spatial orientation 202, 209
- navigation efficiency 208
- negative space *see* volumetric negative space
- non-discrimination 153–154
- nonverbal communication 141–144
- non-wearable systems 76
- normal map 26
 - material design 26
 - surface irregularities 26
- Norman, D. A. 46
- notifications 227–228, 230
- NPCs 136
 - behaviour systems 136
 - empathy 136
 - life-simulation 136
 - world believability 136
- object-fixed reference frame 93
- objective metrics 203, 207–210
 - behavioural metrics 207–210
 - performance-based metrics 207–208
- occlusion 20, 21, 26, 38–39, 43
 - contours 21
 - spatial layout 20, 24 (*see also* interposition)
- occlusion culling 50
- ocularization 134
 - internal 134
 - visual POV 134
 - VR predominance 134
 - zero 134
- oculomotor 19, 20, 24, 31, 32
 - accommodation 32
 - eye muscles 20, 31, 32
 - kinesthetic feedback 20 (*see also* convergence)
- Oculus Rift 9, 10, 11, **12**
- ontological security 42
- open worlds *see* narrative models: open systems; environmental storytelling
- optical illusion 21, 23
 - geometric ambiguity 23
 - visual cortex 20, 23, 26 (*see also* Ames Room; Kanizsa's triangle)
- optimization 106, 107, 192, 198–200
 - animation level of detail 107
 - culling 107
 - GPU-based skinning 107
 - performance 102, 104, 106
- orientation preview 86
- outside-in tracking 80
- overlapping action 108, 109, 111
- pairwise comparison 189
- panel widgets 121, 124
 - modular UI components 121
 - reusable world-locked elements 121, 123
 - spatial configuration 121, 124
- panoramic installation 2, **12**
- parallel viewing 29
 - depth relationships 29
 - infinity focus 29
 - stereo pairs 29
 - wall-eyed viewing 29
- participatory design 170
- pattern description 219–220
- pattern recognition
 - distance, scale, and readability 41
 - dynamic framing: architectural framing and 360° composition 39
 - gestalt in 3D: grouping 35
 - predictive movement patterns 38
 - staging: points of interest, guiding lines, and comfort zones 40
- perceptual consistency 180–181
- perceptual continuity 117, 193–195, 197
 - motion parallax 118
 - scene coherence 117
 - stable depth cues 117
- perceptual efficiency 196–197
- perceptual grouping 35–36

- perceptual model 219, 231
- perceptual psychology 192
- performance 192–200
- performance-aware art direction 45, 51–52
- performance-based metrics 207–208
 - errors 208
 - task completion time 208
 - task success rate 208
- peripersonal space 120
 - body-centred reference frame 119–220
 - near-body interaction zone 120, 124
 - reachable area 120
- peripheral vision 37–38, 40, 43
- personal space 40
- personas 166–167
- perspective 21–24, 26
 - linear perspective 21–24
 - vanishing points 22, 23
- photoreceptor cells 37
- physical depth 37, 39
- physics-based animation 105
 - collision shapes 106
 - mass/friction/elasticity 106
 - real-time 106
 - rigid body 105, 106
- physiological measurements 203, 210–213
 - EEG 211–212
 - electrodermal activity (EDA) 211–212
 - heart rate variability (HRV) 211
- pictorial 19–21, 24, 26, 31
- pictorial cues 20, 24, 26, 31; *see also* monocular cues
- place illusion (PI) 18, 45, 46, 49, 52, 82, 193; *see also* presence; coherence
- plausibility illusion (Psi) 45, 47, 116–117, 193, 200
 - causal consistency 116–117, 119–120, 122
 - diegetic *vs.* non-diegetic cues 116–117
 - internal world logic 116–117, 119–120, 122
- player role 132, 134–135
 - actor/spectator hybrid 132, 134
 - co-author 132, 135
 - prosumer 132, 135
- plot 130–131
 - causes 131
 - consequences 131
 - motives 131
 - open/closed structure 131
- portal 221, 224–225, 230
- Prägnanz
 - law of 35, 35
 - the prevalence of invariance structure 42
- precomputation 197, 200
- presence 1–3, 7, 13, 18, 26, 31, 34–35, 38, 43, 54, 56, 61, 64, 66, 81–82, 115–120, 125–126, 131, 136, 141, 176, 206, 211, 222–223, 230–231
 - “being there” 116–117
 - environmental context 61
 - immersion 131, 136
 - interactive feedback 61
 - perceptual engagement 17
 - place illusion 117, 126
 - self-presence 141
 - sensorimotor contingencies 116–117
 - sensory integration 136
 - social presence 141, 143–144, 146, 148
 - spatial presence 141
 - spatial situating 61
 - user position 131 (*see also* spatial situation model (SSM))
- privacy 147–148, 152, 160
- procedural integrity 111
- profiling 198–200
- projection-based augmented reality 77
- proprioception 46, 178, 219
- prosumer 129, 132, 135
 - decision relevance 135
 - exploration-based interpretation 135
 - user as consumer and creator 129, 135
- Proteus effect 82
- Proximity
 - figure-ground segregation by depth 36
 - law of 35
 - practical application in 3D design 36
 - principle of 34, 35, 36
 - proximity and symmetry 35
- psychoacoustics 56, 57
 - spatial hearing 54, 57
- psychological impacts 148–149, 152–153, 161
 - addiction 148
 - depersonalization 148
 - escapism 148
 - identity and self-perception 148
 - social anxiety 148
- QA gates 199
- quality assurance 198–199
- quality normalization 189
- rail movement 87, 90
- rating scales 186–188
- Raw TLX (RTLX) 205; *see also* NASA Task Load Index (NASA-TLX)
- readability 41, 43
 - reading conditions 41–42
- real-time 102, 104, 107, 109–111
- real-time systems 192, 194
- recall, principle of 42
- redirected walking 87, 90
- reference frames 92–93
- refresh rate 194–195
- remote collaboration 143–144, 148
- remote learning 155–156
- rendering 192, 194–197, 200
- reorientation design 93–94
- repetition 36
- representation 152–153
 - diversity 152
- resource allocation 196
- responsibility 153–154
- responsible design 198–200
- responsivity 102, 110
 - immediate feedback 110, 111
 - perceptual fusion 110
 - temporal precision 110
- retinal disparity 29, 30; *see also* binocular disparity
- rigging 105
- room-scale 86, 96–98

- room-scale VR 9, 11
- Rosli, M. H. W. 35
- Ruddle, R. A. 47
- Ryan, M. L. 48
- safety standards 162
- Sanchez-Vives, M. V. 46, 47
- scale 34, 40–41
- scenarios 166–168
- scene complexity 193, 197
- scenography 45–47, 51, 197–198, 200
- Schmidt, M. A. R. 37, 39
- seated mode 97
- secondary action 109, 110, 112
- segregation 35–37, 43
- self-reported measures *see* subjective metrics
- semantic coherence 47, 48
- semantic dimensions 36
- semantic expectancy 45, 47
- Sensorama 3, 4, **12**
- sensory conflict 85, 89, 193, 199
- sensory feedback 34
- serif typefaces 41
- shading 21, 26, 27, 39
 - attached shadows 26
 - convex 26
 - 3D volume 26
- shadows 26, 27; *see also* shading; ambient occlusion (AO)
- shared virtual space 141, 144, 148
- signal processing 57, 60, 64
 - GPU acceleration 60, 64
 - latency thresholds 60, 61
- Similarity
 - morphology 36
 - principle of 34–36, 35
- Simulator Sickness Questionnaire (SSQ) 170, 207
- Slater-Usch-Steed (SUS) presence scale 170
- smooth locomotion 87
- snap turn 85, 90, 91, 94, 97, 100
- social anxiety 148
- social interaction 221, 231
- social isolation 148
- social presence 141, 143–144, 146, 148
- social sustainability 156–157
- Social VR 140–149
- sound design 54, 56, 62, 65–66, 132, 136
 - adaptive audio 56, 65
 - attention guidance 132
 - emotional rhythm 132, 136
 - interactive audio feedback 56, 61
 - procedural generation 65, 66
 - spatial audio 136
 - voice function 132, 136
- space (VR) 131–132, 136
 - central narrative function 132
 - environmental meaning 136
 - exploration 131–2 (*see also* environmental storytelling; world-building)
- spatial anchoring 115–117, 119, 121
 - head-locked 117
 - reference frames 117, 119
 - world-locked 117, 119, 121, 123, 125
- spatial cognition 115–116, 118, 176–177
 - environmental cues 124
 - orientation 121
 - wayfinding 125
- spatial distance 36
- spatialized audio 140, 142
- spatial organization 219, 226, 230
- spatial orientation 40, 42, 92–94, 165, 167, 181, 183, 202, 209
- spatial positioning 58, 59, 61
 - acoustic localization 55–59
 - 3D sound field 58
- spatial progression 224, 231
- spatial situation model (SSM) 18
- spatial workspace 220, 226, 231
- specular highlights 39
- squash and stretch 108, 110, 111
- SROI (Social Return on Investment) 157
- staging 34, 36, 39, 40, 108, 112
- standalone VR 11, 12
- standing mode 98
- stereogram 29
- stereopsis 18, 20, 23, 29, 31
 - action space 31 (*see also* binocular disparity)
- stereoscope 2, **12**
- stereoscopic rendering 195–196
- stereoscopic vision 2, 41
- St. George, M. 48
- storyline 130–131
 - key events 131
 - summary 131
- storytelling 130
 - emotional language 130
 - expressive resources 130
 - intrigue 130
- streaming 192, 197–198
- structural coherence 34, 41–42
- subjective metrics 203–207
 - NASA Task Load Index (NASA-TLX) 204–205
 - System Usability Scale (SUS) 203–204
 - User Experience Questionnaire (UEQ) 204–205
 - Virtual Reality System Usability Questionnaire (VRSUQ) 205–207
- surface Gestalt 39
 - similarity and continuity 36
- suspension of disbelief 48
- sustainability 83, 199; *see also* battery dependence
- sustainability education 156–158
- Sustainable Development Goals (SDGs) 150–151, 154
- Sweller, J. 47
- Sword of Damocles 4, 5, **12**
- symbolic interaction 218, 223
- symmetry
 - distance, scale and readability 41
 - law of 35–36, 35, 41, 43
 - the prevalence of invariant structure 42
- system responsiveness 72, 82
- system transparency 183–185
- System Usability Scale (SUS) 203–204
- task completion time 203, 208
- task logic 219, 231
- task success rate 208

- telemetry 187
- teleportation 86–87
- telepresence 5, 7, 144
- temporal design 131, 135
 - cycles 135
 - pacing 135
 - real-time vs internal time 131, 135
- temporal stability 193–194
- testing protocols 170–171
 - functional validation 170
 - qualitative barrier analysis 170
- texture gradients 20, 24–25, 38; *see also* normal map
- texture resolution 196–197
- therapy and rehabilitation 82–83
- three-dimensional content design 34–36, 41
 - depth cues and lighting as guidance 37–39
 - Gestalt in 3D: grouping 35–37
 - readability, distance, scale, and typography in VR 41–43
- timbre 55
- time 228, 230
- tolerance windows 168, 171
- tool-based interaction 228–229, 231
- tracking latency 122
 - interaction responsiveness 122
 - motion-to-photon delay 122
 - temporal fidelity 122
- tracking stability 192, 198–200
- tracking systems 1, 5–7, 11, 141–143, 148
 - gesture recognition 143
 - hand and body tracking 143
 - head tracking 143
- translation domains 220–222
- transparency 152–153
- travel reduction 155–156
- tunneling 85, 91
- 2D interfaces 217–218
- 2D-to-VR translation 218–231
- typography 34, 41, 43
 - sans-serif typefaces 41, 43
 - serif typefaces 41
- UEQ-S 204
- uncanny valley 46, 47, 51, 105, 113, 136
 - character acceptance 136
 - realism limits 136
- UNESCO principles 153–154
- unification 35
- usability 13, 177
- usability evaluation 202–203, 207
- usability testing 201–203, 207–213
 - iterative design 202–203
 - iterative development 202–203
- user behavior tracking 148
- user comfort 73, 83; *see also* accessibility; ergonomics;
fatigue; motion sickness
- user experience (UX) 36, 43, 177
- User Experience Questionnaire (UEQ) 204–205
- user-informed validation 187
- user intention 218–220, 231
- user interface (UI) 43
- user perspective 134
 - auricularization 134
 - focalization 134
 - narrative knowledge 134
 - ocularization 134
 - sensory access 134
- user profile 85, 88–90, 122, 123, 165, 168, 221
- user safety 147
- user testing 63–64
 - physiological measurement 64
 - Quality of Experience (QoE) 63
- user wellbeing 152–153, 161, 199–200
- vection 95
- vestibular 103, 104
- vestibular system 85
- Videoplace 5, **12**
- vignetting 85, 91, 104
- Villafañe, J. 38, 40
- virtual environment 1, 4–7, 8, 9, 11, 13
- virtual environment (as narrative agent) 132
 - character-like function 132
 - meaning-bearing spaces 132
- virtual events and performances 145–146, 148
- virtual meetings 143–144, 147, 155
- virtual reality (VR) 1–13, 115, 150–151, 164–165, 169, 171,
175, 192–213, 217–231
 - ethical design 151–153
 - immersive systems 115, 117, 119, 122
 - real-time interaction 115
 - spatial UIs 115
 - sustainability 154–158
- Virtual Reality Sickness Questionnaire (VRSQ) 207
- Virtual Reality System Usability Questionnaire (VRSUQ)
205–207
- visual accents 40
- visual attention 209–210
- visual comfort 197
- visual ergonomics 124
 - comfortable viewing distances 124
 - eye strain reduction 124
 - gaze angles 121, 124
- visual hierarchy 45, 49, 218–219, 230
- visual impairments 73
- visual interaction 75–77
- visual language 34
- visual noise 37, 39
- voice commands 72
- voice interaction 142
- volumetric negative space 37, 43
- volumetric staging 34, 36–37
- Voxel Research Lab 213
- VR-Heuristics Severity Rating Scale 187–188
- VR-Heuristic Weighted Index (VR-HWI) 189–190
- VR narrative 129, 131–132, 136
 - active user role 129, 131
 - immersive expression 136
 - transformation of traditional model 129, 131
- wearable sensors 211–212, 215
 - EEG devices 211
 - Empatica wristband 212
- Web Accessibility Initiative (WAI) 165
- Web Content Accessibility Guidelines (WCAG) 165

-
- weighted scoring 175, 189–190
 - Wheatstone 29
 - stereoscope 29 (*see also* stereopsis)
 - Wheatstone stereoscope 2, **12**
 - Widjojo, E. A. 34
 - window multitasking 226, 231
 - Wong, W. 34
 - workflows 192, 198–199
 - workload 202, 204–205, 210–213; *see also* cognitive load
 - world-building 135–136
 - coherence 135
 - historical framework 135
 - internal rules 135
 - spatial lore 135
 - world-fixed reference frame 93
 - World Health Organization (WHO) 164
 - world scale 94–95
 - World Wide Web Consortium (W3C) 164–165