The VR Book

Perception and Interaction Design for Virtual Reality

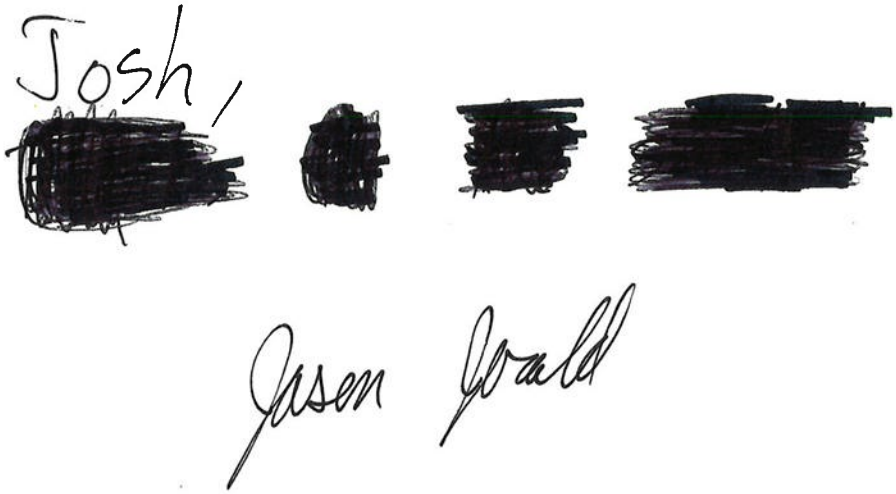
Jason Jerald





Meta Platforms, Inc.
Exhibit 1063
Page 001

The VR Book



Meta Platforms, Inc.
Exhibit 1063
Page 002

ACM Books

Editor in Chief
M. Tamer Özsu, *University of Waterloo*

ACM Books is a new series of high-quality books for the computer science community, published by ACM in collaboration with Morgan & Claypool Publishers. ACM Books publications are widely distributed in both print and digital formats through booksellers and to libraries (and library consortia) and individual ACM members via the ACM Digital Library platform.

The VR Book: Perception and Interaction Design for Virtual Reality
Jason Jerald, *NextGen Interactions*
2015

Edmund Berkeley and the Social Responsibility of Computer Professionals
Bernadette Longo, *New Jersey Institute of Technology*
2015

Candidate Multilinear Maps
Sanjam Garg, *University of California, Berkeley*
2015

Smarter than Their Machines: Oral Histories of Pioneers in Interactive Computing
John Cullinane, *Northeastern University; Mossavar-Rahmani Center for Business and Government, John F. Kennedy School of Government, Harvard University*
2015

A Framework for Scientific Discovery through Video Games
Seth Cooper, *University of Washington*
2014

Trust Extension as a Mechanism for Secure Code Execution on Commodity Computers
Bryan Jeffrey Parno, *Microsoft Research*
2014

Embracing Interference in Wireless Systems
Shyamnath Gollakota, *University of Washington*
2014

Meta Platforms, Inc.
Exhibit 1063
Page 003

The VR Book

Perception and Interaction Design for Virtual Reality

Jason Jerald

NextGen Interactions

ACM Books #7



Meta Platforms, Inc.
Exhibit 1063
Page 004

Copyright © 2015 by the Association for Computing Machinery
and Morgan & Claypool Publishers

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews—without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks or registered trademarks. In all instances in which Morgan & Claypool is aware of a claim, the product names appear in initial capital or all capital letters. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

The VR Book: Perception and Interaction Design for Virtual Reality

Jason Jerald

books.acm.org
www.morganclaypool.com

ISBN: 978-1-97000-112-9 paperback
ISBN: 978-1-97000-113-6 ebook
ISBN: 978-1-62705-114-3 ePub
ISBN: 978-1-97000-115-0 hardcover
Series ISSN: 2374-6769 print 2374-6777 electronic

DOIs

10.1145/2792790 Book
10.1145/2792790.2792791 Preface/Intro
10.1145/2792790.2792792 Part I
10.1145/2792790.2792793 Chap. 1
10.1145/2792790.2792794 Chap. 2
10.1145/2792790.2792795 Chap. 3
10.1145/2792790.2792796 Chap. 4
10.1145/2792790.2792797 Chap. 5
10.1145/2792790.2792798 Part II
10.1145/2792790.2792799 Chap. 6
10.1145/2792790.2792800 Chap. 7
10.1145/2792790.2792801 Chap. 8
10.1145/2792790.2792802 Chap. 9
10.1145/2792790.2792803 Chap. 10
10.1145/2792790.2792804 Chap. 11
10.1145/2792790.2792805 Part III
10.1145/2792790.2792806 Chap. 12
10.1145/2792790.2792807 Chap. 13
10.1145/2792790.2792808 Chap. 14
10.1145/2792790.2792809 Chap. 15
10.1145/2792790.2792810 Chap. 16
10.1145/2792790.2792811 Chap. 17
10.1145/2792790.2792812 Chap. 18
10.1145/2792790.2792813 Chap. 19
10.1145/2792790.2792814 Part IV
10.1145/2792790.2792815 Chap. 20
10.1145/2792790.2792816 Chap. 21
10.1145/2792790.2792817 Chap. 22
10.1145/2792790.2792818 Chap. 23
10.1145/2792790.2792819 Chap. 24
10.1145/2792790.2792821 Chap. 25
10.1145/2792790.2792820 Part V
10.1145/2792790.2792822 Chap. 26
10.1145/2792790.2792823 Chap. 27
10.1145/2792790.2792824 Chap. 28
10.1145/2792790.2792825 Chap. 29
10.1145/2792790.2792826 Part VI
10.1145/2792790.2792827 Chap. 30
10.1145/2792790.2792828 Chap. 31
10.1145/2792790.2792829 Chap. 32
10.1145/2792790.2792830 Chap. 33
10.1145/2792790.2792831 Chap. 34
10.1145/2792790.2792832 Part VII
10.1145/2792790.2792833 Chap. 35
10.1145/2792790.2792834 Chap. 36
10.1145/2792790.2792835 Appendix A
10.1145/2792790.2792836 Appendix B
10.1145/2792790.2792837 Glossary/Refs

A publication in the ACM Books series, #7
Editor in Chief: M. Tamer Özsu, *University of Waterloo*
Area Editor: John C. Hart, University of Illinois

First Edition

10 9 8 7 6 5 4 3 2 1

Meta Platforms, Inc.
Exhibit 1063
Page 005

This book is dedicated to the entire community of VR researchers, developers, designers, entrepreneurs, managers, marketers, and users. It is their passion for, and contributions to, VR that makes this all possible. Without this community, working in isolation would make VR an interesting niche research project that could neither be shared nor improved upon by others. If you choose to join this community, your pursuit of VR experiences may very well be the most intense years of your life, but you will find the rewards well worth the effort. Perhaps the greatest rewards will come from the users of your experiences—for if you do VR well then your users will tell you how you have changed their lives—and that is how we change the world.

There are many facets to VR creation, ranging from getting the technology right, sometimes during exhausting overnight sessions, to the fascinating and abundant collaboration with others in the VR community. At times, what we are embarking on can feel overwhelming. When that happens, I look to a quote posted on my wall and am reminded about the joy of being a part of the VR revolution, as stated by George Bernard Shaw.

> This is the true joy in life, the being used for a purpose recognized by yourself as a mighty one; the being a force of nature . . . I am of the opinion that my life belongs to the whole community and as long as I live it is my privilege to do for it whatever I can. I want to be thoroughly used up when I die, for the harder I work, the more I live. I rejoice in life for its own sake. Life is no 'brief candle' to me. It is sort of a splendid torch which I have a hold of for the moment, and I want to make it burn as brightly as possible before handing it over to future generations.
>
> —George Bernard Shaw

This book is thus dedicated to the VR community and the future generations that will create many virtual worlds as well as change the real world. My purpose in writing this book is to welcome others into this VR community, to help fuel a VR revolution that changes the world and the way we interact with it and each other, in ways that have never before been possible—until now.

Meta Platforms, Inc.
Exhibit 1063
Page 006

Contents

Preface **xvii**
Figure Credits **xxv**
Overview **1**

PART I **INTRODUCTION AND BACKGROUND 7**

Chapter 1 **What Is Virtual Reality? 9**
1.1 The Definition of Virtual Reality **9**
1.2 VR Is Communication **10**
1.3 What Is VR Good For? **12**

Chapter 2 **A History of VR 15**
2.1 The 1800s **15**
2.2 The 1900s **18**
2.3 The 2000s **27**

Chapter 3 **An Overview of Various Realities 29**
3.1 Forms of Reality **29**
3.2 Reality Systems **30**

Chapter 4 **Immersion, Presence, and Reality Trade-Offs 45**
4.1 Immersion **45**
4.2 Presence **46**
4.3 Illusions of Presence **47**
4.4 Reality Trade-Offs **49**

* Practitioner chapters are marked with a star next to the chapter number. See page 5 for an explanation.

Meta Platforms, Inc.
Exhibit 1063
Page 007

★ Chapter 5 The Basics: Design Guidelines 53
Introduction and Background (Part I) 53
VR Is Communication (Section 1.2) 53
An Overview of Various Realities (Chapter 3) 54
Immersion, Presence, and Reality Trade-Offs (Chapter 4) 54

PART II PERCEPTION 55

Chapter 6 Objective and Subjective Reality 59
6.1 Reality Is Subjective 59
6.2 Perceptual Illusions 61

Chapter 7 Perceptual Models and Processes 71
7.1 Distal and Proximal Stimuli 71
7.2 Sensation vs. Perception 72
7.3 Bottom-Up and Top-Down Processing 73
7.4 Afference and Efference 73
7.5 Iterative Perceptual Processing 74
7.6 The Subconscious and Conscious 76
7.7 Visceral, Behavioral, Reflective, and Emotional Processes 77
7.8 Mental Models 79
7.9 Neuro-Linguistic Programming 80

Chapter 8 Perceptual Modalities 85
8.1 Sight 85
8.2 Hearing 99
8.3 Touch 103
8.4 Proprioception 105
8.5 Balance and Physical Motion 106
8.6 Smell and Taste 107
8.7 Multimodal Perceptions 108

Chapter 9 Perception of Space and Time 111
9.1 Space Perception 111
9.2 Time Perception 124
9.3 Motion Perception 129

Meta Platforms, Inc.
Exhibit 1063
Page 008

Chapter 10 Perceptual Stability, Attention, and Action 139
10.1 Perceptual Constancies 139
10.2 Adaptation 143
10.3 Attention 146
10.4 Action 151

★ **Chapter 11** Perception: Design Guidelines 155
Objective and Subjective Reality (Chapter 6) 155
Perceptual Models and Processes (Chapter 7) 155
Perceptual Modalities (Chapter 8) 156
Perception of Space and Time (Chapter 9) 156
Perceptual Stability, Attention, and Action (Chapter 10) 157

PART III ADVERSE HEALTH EFFECTS 159

Chapter 12 Motion Sickness 163
12.1 Scene Motion 163
12.2 Motion Sickness and Vection 164
12.3 Theories of Motion Sickness 165
12.4 A Unified Model of Motion Sickness 169

Chapter 13 Eye Strain, Seizures, and Aftereffects 173
13.1 Accommodation-Vergence Conflict 173
13.2 Binocular-Occlusion Conflict 173
13.3 Flicker 174
13.4 Aftereffects 174

Chapter 14 Hardware Challenges 177
14.1 Physical Fatigue 177
14.2 Headset Fit 178
14.3 Injury 178
14.4 Hygiene 179

Chapter 15 Latency 183
15.1 Negative Effects of Latency 183
15.2 Latency Thresholds 184
15.3 Delayed Perception as a Function of Dark Adaptation 185

Meta Platforms, Inc.
Exhibit 1063
Page 009

15.4 Sources of Delay 187
15.5 Timing Analysis 193

Chapter 16 Measuring Sickness 195
16.1 The Kennedy Simulator Sickness Questionnaire 195
16.2 Postural Stability 196
16.3 Physiological Measures 196

Chapter 17 Summary of Factors That Contribute to Adverse Effects 197
17.1 System Factors 198
17.2 Individual User Factors 200
17.3 Application Design Factors 203
17.4 Presence vs. Motion Sickness 205

★ Chapter 18 Examples of Reducing Adverse Effects 207
18.1 Optimize Adaptation 207
18.2 Real-World Stabilized Cues 207
18.3 Manipulate the World as an Object 209
18.4 Leading Indicators 210
18.5 Minimize Visual Accelerations and Rotations 210
18.6 Ratcheting 211
18.7 Delay Compensation 211
18.8 Motion Platforms 212
18.9 Reducing Gorilla Arm 213
18.10 Warning Grids and Fade-Outs 213
18.11 Medication 213

★ Chapter 19 Adverse Health Effects: Design Guidelines 215
Hardware 215
System Calibration 216
Latency Reduction 216
General Design 217
Motion Design 218
Interaction Design 219
Usage 220
Measuring Sickness 221

Meta Platforms, Inc.
Exhibit 1063
Page 010

PART IV CONTENT CREATION 223

Chapter 20 High-Level Concepts of Content Creation 225

20.1 Experiencing the Story 225
20.2 The Core Experience 228
20.3 Conceptual Integrity 229
20.4 Gestalt Perceptual Organization 230

Chapter 21 Environmental Design 237

21.1 The Scene 237
21.2 Color and Lighting 238
21.3 Audio 239
21.4 Sampling and Aliasing 240
21.5 Environmental Wayfinding Aids 242
21.6 Real-World Data 246

Chapter 22 Affecting Behavior 251

22.1 Personal Wayfinding Aids 251
22.2 Center of Action 254
22.3 Field of View 255
22.4 Casual vs. High-End VR 255
22.5 Characters, Avatars, and Social Networking 257

★ **Chapter 23 Transitioning to VR Content Creation 261**

23.1 Paradigm Shifts from Traditional Development to VR Development 261
23.2 Reusing Existing Content 262

★ **Chapter 24 Content Creation: Design Guidelines 267**

High-Level Concepts of Content Creation (Chapter 20) 267
Environmental Design (Chapter 21) 269
Affecting Behavior (Chapter 22) 271
Transitioning to VR Content Creation (Chapter 23) 272

PART V INTERACTION 275

Chapter 25 Human-Centered Interaction 277

25.1 Intuitiveness 277

Meta Platforms, Inc.
Exhibit 1063
Page 011

25.2 Norman's Principles of Interaction Design 278
25.3 Direct vs. Indirect Interaction 284
25.4 The Cycle of Interaction 285
25.5 The Human Hands 287

★ Chapter 26 VR Interaction Concepts 289
26.1 Interaction Fidelity 289
26.2 Proprioceptive and Egocentric Interaction 291
26.3 Reference Frames 291
26.4 Speech and Gestures 297
26.5 Modes and Flow 301
26.6 Multimodal Interaction 302
26.7 Beware of Sickness and Fatigue 303
26.8 Visual-Physical Conflict and Sensory Substitution 304

★ Chapter 27 Input Devices 307
27.1 Input Device Characteristics 307
27.2 Classes of Hand Input Devices 311
27.3 Classes of Non-hand Input Devices 317

★ Chapter 28 Interaction Patterns and Techniques 323
28.1 Selection Patterns 325
28.2 Manipulation Patterns 332
28.3 Viewpoint Control Patterns 335
28.4 Indirect Control Patterns 344
28.5 Compound Patterns 350

★ Chapter 29 Interaction: Design Guidelines 355
Human-Centered Interaction (Chapter 25) 355
VR Interaction Concepts (Chapter 26) 358
Input Devices (Chapter 27) 361
Interaction Patterns and Techniques (Chapter 28) 363
Viewpoint Control Patterns (Section 28.3) 364

PART VI ITERATIVE DESIGN 369

Chapter 30 Philosophy of Iterative Design 373
30.1 VR Is Both an Art and a Science 373

Meta Platforms, Inc.
Exhibit 1063
Page 012

30.2 Human-Centered Design 373
30.3 Continuous Discovery through Iteration 374
30.4 There Is No One Way—Processes are Project Dependent 375
30.5 Teams 376

★ Chapter 31 The Define Stage 379
31.1 The Vision 380
31.2 Questions 380
31.3 Assessment and Feasibility 382
31.4 High-Level Design Considerations 383
31.5 Objectives 383
31.6 Key Players 384
31.7 Time and Costs 385
31.8 Risks 387
31.9 Assumptions 388
31.10 Project Constraints 388
31.11 Personas 391
31.12 User Stories 392
31.13 Storyboards 393
31.14 Scope 393
31.15 Requirements 395

★ Chapter 32 The Make Stage 401
32.1 Task Analysis 402
32.2 Design Specification 405
32.3 System Considerations 410
32.4 Simulation 413
32.5 Networked Environments 415
32.6 Prototypes 421
32.7 Final Production 423
32.8 Delivery 424

★ Chapter 33 The Learn Stage 427
33.1 Communication and Attitude 428
33.2 Research Concepts 429
33.3 Constructivist Approaches 436
33.4 The Scientific Method 443
33.5 Data Analysis 447

Meta Platforms, Inc.
Exhibit 1063
Page 013

★ Chapter 34 Iterative Design: Design Guidelines 453
Philosophy of Iterative Design (Chapter 30) 453
The Define Stage (Chapter 31) 454
The Make Stage (Chapter 32) 458
The Learn Stage (Chapter 33) 464

PART VII THE FUTURE STARTS NOW 471

Chapter 35 The Present and Future State of VR 473
35.1 Selling VR to the Masses 473
35.2 Culture of the VR Community 474
35.3 Communication 475
35.4 Standards and Open Source 480
35.5 Hardware 483
35.6 The Convergence of AR and VR 484

★ Chapter 36 Getting Started 485

Appendix A Example Questionnaire 489

Appendix B Example Interview Guidelines 495

Glossary 497
References 499
Index 525
Author's Biography 527

Meta Platforms, Inc.
Exhibit 1063
Page 014

Preface

I've known for some time that I wanted to write a book on VR. However, I wanted to bring a unique perspective as opposed to simply writing a book for the sake of doing so. Then insight hit me during Oculus Connect in the fall of 2014. After experiencing the Oculus Crescent Bay demo, I realized the hardware is becoming really good. Not by accident, but as a result of some of the world's leading engineers diligently working on the technical challenges with great success. What the community now desperately needs is for content developers to understand human perception as it applies to VR, design experiences that are comfortable (i.e., do not make you sick), and to create intuitive interactions within their immersive creations. That insight led me to the realization that I need to stop focusing primarily on technical implementation and start focusing on higher-level challenges in VR and its design. Focusing on these challenges offers the most value to the VR community of the present day, and is why a new VR book is necessary.

I had originally planned to self-publish instead of spending valuable time on pitching the idea of a new VR book to publishers which I know can be a very long, arduous, and disappointing path. Then one of the most serendipitous things occurred. A few days after having the idea for a new VR book and committing to make it happen, my former undergraduate advisor Dr. John Hart, now Professor of Computer Science at the University of Illinois Urbana-Champaign and, more germanely, the Editor of the Computer Graphics Series of ACM Books, contacted me out of the blue. He informed me that Michael Morgan of Morgan & Claypool Publishers wanted to publish a book on VR content creation and John thought I was the person to make it happen. It didn't take much time to enthusiastically accept their proposal, as their vision for the book was the same as mine.

I've condensed approximately 20 years and 30,000 hours of personal study, application, notes, and VR sickness into this book. Those two decades of my VR career have somehow been summarized with six months of intense writing and editing from January to July of 2015. I knew, as others working hard on VR know, the time is now, and

Meta Platforms, Inc.
Exhibit 1063
Page 015

after finishing up a couple of contracts I was finally able to put other aspects of my life on hold (clients have been very understanding!), sometimes writing more than 75 hours a week. I hope this rush to get these timely concepts into one book does not sacrifice the overall quality, and I am always seeking feedback. Please contact me at book@nextgeninteractions.com to let me know of any errors, lack of clarity, or important points missed so that an improved second edition can emerge at some point.

It all started in 1980

I owe much of my pursuit of developing VR systems and applications to my parents. It started at the age of six with an Atari game system that my family couldn't afford, but I wanted so badly. For Christmas 1980, I somehow received it. Then in 1986 my mother took away the games in hopes of curing my addiction, which naturally forced me to make my own games (with advanced moving 2D sprites!) on the family's Commodore 64. Here, I taught myself programming and essential software design concepts, such as simple computer graphics and collision detection, which turns out to be quite important for VR development. At that point, she thankfully gave up on trying to cure me. I also owe my father who connected me with my first internship in the Summer of 1992 after my junior year of high school at a design firm where he worked. My job was to deliver plots from the printers to the designers. This did not completely fill my time and I somehow managed to gain access to AutoCad and 3D Studio R2 (long before today's 3ds Max!). I was soon extruding 2D architectural plans into 3D worlds and animating horrible-looking flat-shaded polygons in the evenings and weekends when nobody was the wiser.

SIGGRAPH 1995 and 1996

After regrettably missing SIGGRAPH 1994, I somehow acquired the conference course notes on creating virtual worlds and other related concepts. Soon afterwards, I discovered the SIGGRAPH Student Volunteer Program and realized that was my ticket to the conference. After realizing what I missed, I was not going to leave the opportunity to chance. I sought out a referral from Dr. John Hart, a professor very heavily involved with SIGGRAPH. He must have given me a solid referral based on my initiative and passion for computer graphics, as I had not yet taken a class from him. John would later become my undergraduate advisor and eventually editor of this book. Both his advice through the years and SIGGRAPH had an enormous influence on my career. In fact, I ended up leading the SIGGRAPH Student Volunteer Program more than a decade later, out of appreciation and respect for its ability to inspire careers in computer graphics.

Meta Platforms, Inc.
Exhibit 1063
Page 016

I was quite fortunate to be accepted to the 1995 SIGGRAPH Student Volunteer Program in Los Angeles which is where I first experienced VR and have been fully hooked ever since. Having never attended a conference up to that point that I was this passionate about, I was completely blown away by its people and its magnitude. I was no longer alone in this world—I had finally found "my people." Soon afterwards, I found my first VR mentor Richard May, now Director of the National Visualization and Analytics Center. I enthusiastically jumped in on helping him build one of the world's first immersive VR medical applications at Battelle Pacific Northwest National Laboratories. Richard and I returned to SIGGRAPH 1996 in New Orleans, where I more intentionally sought out VR. VR was now bigger than ever and I remember two things from the conference that defined my future. One was a course on VR interactions by the legendary Dr. Frederick P. Brooks, Jr. and one of his students Mark Miné. The other event was a VR demo that I still remember vividly to this day—and still one of the most compelling demos I've seen yet. It was a virtual Legoland where the user built an entire world around himself by snapping Legos and Lego neighborhoods together in an intuitive, almost effortless manner.

Post 1996

SIGGRAPH 1996 was a turning point for me in knowing what I wanted to do with my life. Since that time, I have been fortunate to know and work with many individuals that have inspired me. This book is largely a result of their work and their mentorship.

Since 1996, Dr. Brooks unknowingly inspired me to move on from a full-time dream VR job in the late 1990s at HRL Laboratories to pursue VR at the next level at UNC-Chapel Hill. Once there, I managed to persuade Dr. Brooks to become my Ph.D. advisor in studying VR latency. In addition to being a major influence on my career, two of his books—*The Mythical Man Month* and more recently *The Design of Design*—are heavily referenced throughout Part VI, Iterative Design. He also had significant input for Part II, Perception, especially Chapter 16, Latency, as some of that writing originally came from my dissertation that was largely a result of significant suggestions from him. His serving as advisor for Mark Miné's seminal work on VR interaction also indirectly affected Part V, Interaction. It is still quite difficult for me to fathom that before he came along, bytes—the molecules of all digital technology—were not always eight bits in size like they are defined to be today. That 8-bit design decision he made in the 1960s, which most all of us computer scientists now just assume to be an inherent truth, is just one of his many contributions to computers in general, along with his more specific contributions to VR research. It boggles my mind even more how a small-town kid like myself growing up in a tiny town of 1200 people on the opposite

Meta Platforms, Inc.
Exhibit 1063
Page 017

side of the country, somehow came to study under an ACM Turing Award recipient (the equivalent of the Nobel Prize in Computer Science). For that, I will forever be grateful to Dr. Brooks and the UNC-Chapel Hill Department of Computer Science for admitting me.

In 2009, I interviewed with Paul Mlyniec, President of Digital ArtForms. At some point during the interview, I came to the realization that this was the man that led the VR Lego work from SIGGRAPH that had inspired and driven me for over a decade. The interface in that Lego demo is one of the first implementations of what I refer to as the 3D Multi-Touch Pattern in Section 29.3.3, a pattern two decades ahead of its time. I've now worked closely with Paul and Digital ArtForms for six years on various projects, including ones that have improved upon 3D Multi-Touch (Sixense's MakeVR also uses this same viewpoint control implementation). We are currently working together (along with Sixense and Wake Forest School of Medicine) on immersive games for neuroscience education funded by the National Institute of Health. In addition to these games, several other examples of Digital ArtForms' work (along with work from its sister company Sixense that Paul was essential in helping to form) are featured throughout the book.

VR Today

After a long VR drought after the 1990s, VR is bigger than ever at SIGGRAPH. An entire venue, the VR Village, is dedicated specifically to VR and I have the privelege of leading the Immersive Realities Contest. After so many years of heavy involvement with my SIGGRAPH family, it feels quite fitting that this book is launched at the SIGGRAPH bookstore on the 20th anniversary of my discovery of SIGGRAPH and VR.

I joke that when I did my Ph.D. Dissertation on latency perception for head-mounted displays, perhaps ten people in the world cared about VR latency—and five of those people were on my committee (in three different time zones!). Then in 2011, people started wanting to know more. I remember having lunch with Amir Rubin, CEO of Sixense. He was inquiring about consumer HMDs for games. I thought the idea was crazy; we could barely get VR to work well in a lab and he wanted to put it in people's living rooms! Other inquires led to work with companies such as Valve and Oculus. All three of these companies are doing spectacular jobs of making high-quality VR accessible, and now suddenly everyone wants to experience VR. Largely due to these companies' efforts, VR has turned a corner, transitioning from a specialized laboratory instrument available only to the technically elite, to a mainstream mode of content consumption available to any consumer. Now, most everyone that is involved in VR technology understands at least the basics of la-

Meta Platforms, Inc.
Exhibit 1063
Page 018

tency and its challenges/dangers, most notably motion sickness—the greatest risk for VR. Even better, ultra-low latency hardware technologies (e.g., low persistence OLED displays) that I once unsuccessfully searched the world for (I ended up having to build/simulate my own; the best I did was 7.4 ms of end-to-end latency with tracking and rendering performed at ~1500 frames per second) are being developed in mass quantities by giants such as Samsung, Sony, Valve, and Oculus. Times have certainly changed!

The last 20 years of pursuing VR have truly been a dream. During that time, I imagined and even seriously considered starting companies devoted to VR, but it was never feasible until recently. Today it is more of a real fantasy than a simple dream, now that VR technology is delivering upon its promise of the 1990s. Describing the feeling is like trying to describe a VR experience. Words cannot do justice concerning what it is like to be a part of this VR community and contributing to the VR revolution. Through my VR consulting and contracting firm, NextGen Interactions, I have the privilege of working with some of the best companies in the world that are able to do things that could only previously be imagined. Virtual reality is unlike any technology devised to date and has the potential to not only change the fictional synthetic worlds we make up, but to change people's real lives. I very much look forward to seeing what this community discovers and creates over the next 20 years!

Acknowledgments

It would be untrue if I claimed that this book was completely based on my own work. The book was certainly not written in isolation and is a result of contributions from a vast number mentors, colleagues, and friends in the VR community that have been working just as hard as I have been over the years to make VR something real. I couldn't possibly say everything there is about VR, and I encourage readers to further investigate the more than 300 references discussed throughout the book, along with other writings. I've learned so much from those who came before me as well as those newer to VR. Unfortunately, it is not possible to list everyone that has influenced this book, but the following gives thanks to those that have given the most support.

First of all, this book would have never been anything beyond a simple self-published work without the team at Morgan & Claypool Publishers and ACM Books. On the Morgan & Claypool Publishers side I'd like to thank Executive Editor Diane Cerra, Editorial Assistant Samantha Draper, Copyeditor Sara Kreisman, Producer Paul Anagnostopoulos, Artist Laurel Muller, and President Michael Morgan. On the ACM Books side, I'd like to thank Computer Graphics Series Editor John Hart and Editor-in-Chief M. Tamer Özsu.

Meta Platforms, Inc.
Exhibit 1063
Page 019

This book you are reading is very different and much better than initial drafts due to suggestions from great reviewers. Reviewers of the book include Paul Mlyniec (President of Digital ArtForms), Arun Yoganandan (formerly of Digital ArtForms and Disney, now a Research Engineer at Samsung), Russell Taylor (former professor of Computer Science at UNC-Chapel Hill, now Co-Founder of Rheomics and independent consultant at Relia Solve), Beau Cronin (an expert VR neuroscientist at Salesforce who is writing his own VR book), Mike McArdle (Co-Founder of the Virtual Reality Learning Experience), Ryan McMahan (Professor at University of Texas at Dallas), Eugene Nalivaiko (Professor at New Castle University), Kyle Yamamoto (Co-Founder and game developer at MochiBits), Ann McNamara (Professor at Texas A&M), Mark Fagiano (Professor at Emory University), Zach Wendt (computer games guru and independent software developer), Francisco Ortega (PostDoc Fellow at Florida International University and who is writing his own VR book), Matt Cook (Emerging Technologies Librarian at the University of Oklahoma), Sharif Razzaque (CTO of Inneroptic), Neeta Nahta (Co-Founder of NextGen Interactions), Kevin Rio (UX Researcher at Microsoft), Chris Pusczak (Project Manager at Burnout Game Ventures and SymbioVR), Daniel Ackley (psychologist), my mother Susan Jerald (behavioral psychologist), and my father Rick Jerald (CEO of Envirocept).

Additional specific input came from Mark Bolas (Professor at the University of Southern California), Neil Schneider (Founder of Meant to be Seen and Founder/Executive Director of the Immersive Technology Alliance), Eric Greenbaum (attorney and Founder of Jema VR), John Baker (COO of Chosen Realities), Denny Unger (Founder and President of Cloudhead Games), Andrew Robinson and Sigurdur Gunnarsson (developers at CCP Games), Max Rheiner (Founder and CTO of Somniacs), Chadwick Wingrave (Founder of Conquest Creations), Barry Downes (CEO of TSSG), Denise Quesnel (Research Associate at Emily Carr University of Art and Design, Co-Founder of the Canadian & Advanced Imaging Society, and Co-Chair of the SIGGRAPH VR Village), Nick England (Founder and CEO of 3rdTech), Jesse Joudrey (CEO of VRChat and CTO of Jespionage Entertainment), David Collodi (Co-Founder and CTO of CI Dynamics), David Beaver (Co-Founder of the Overview Institute), and Bill Howe (CEO of the Growth Engine Group).

I've been very fortunate to be employed by some amazing organizations and even more amazing bosses that are world-leading experts in VR. These individuals gave me the opportunity to make a career out of what I love doing and work with them led to much of what is contained within these pages: Richard May of Battelle Pacific Northwest National Laboratories (now Director of the National Visualization and Analytics Center), Michael Papka of Argonne National Laboratories, Mike Daily of HRL Laboratories, Greg Schmidt of the Naval Research Laboratory, Steve Ellis and

Meta Platforms, Inc.
Exhibit 1063
Page 020

Dov Adelstein of NASA Ames, Paul Mlyniec of Digital ArtForms, and Michael Abrash at Valve (now at Oculus).

In addition to those listed above, the following individuals have been invaluable as VR advisors/mentors throughout the years: Howard Neely, Research Project Manager at HRL Laboratories and now CEO at Three Birds Systems; Ron Azuma, Senior Research Staff Member and now at Intel; Fred Brooks, Henry Fuchs, Mary Whitton, and Anselmo Lastra—all Professors at UNC Chapel Hill; Jeff Bellinghausen, formerly lead Engineer at Digital ArtForms and CTO at Sixense, now at Valve; and Amir Rubin, CEO of Sixense.

I also very much appreciate working with the teams at Sixense, Valve, and Oculus. I've been incredibly impressed with the people and their commitment in creating the highest quality software and hardware, and not just resurrecting VR but raising awareness to levels beyond anything that previously existed.

Others I've had the pleasure of working with that have influenced this book include Rupert Meghnot and his team at Burnout Game Ventures, especially Chris Pusczak of the VRMor team; Jan Goetgeluk of Virtuix; Simon Solotko, the Kickstarter guru who marketed some of the biggest VR Kickstarters such as the Virtuix Omni, the Cyberith Virtualizer, and the Sixense STEM; the UNC Effective Virtual Environments team; Karl Krantz of Silicon Valley Virtual Reality who was the first to pioneer local meetups where his events are duplicated throughout the world; and Henry Velez, Founder of the RTP Virtual Reality Enthusiasts, that I now have the privilege of leading.

I've also enjoyed working with literally thousands of volunteers through ACM SIGGRAPH, IEEE VR, and IEEE 3DUI, ranging from student volunteers to the legends of computer graphics and VR. In particular, I'd like to thank the SIGGRAPH Conference Committees of the last 20 years. Of course, I've also enjoyed working with those outside of conferences, ranging from those in academics to those doing startups. Then there are those who have given me invaluable feedback on VR projects, ranging from middle schoolers to VR experts.

I'd especially like to thank Neeta Nahta, my partner in business and life, who comes from a completely different world of sales, financial analysis, and corporate boardrooms. Her input enables me to see VR from a different perspective than most anyone else in the VR community. Seeing VR as a sustainable and profitable business that adds real value to the world will ensure that this time VR is here for good.

I would also like to thank [Jerald2]the Link Foundation for supporting me through the Link Foundation Advanced Simulation and Training Fellowship which helped to fund graduate work in studying and reducing latency. It is certainly an honor to be funded by the foundation started by Edwin A Link, who built the world's first mechanical flight simulator in 1928 (Figure.3.5) which some consider to be one of

Meta Platforms, Inc.
Exhibit 1063
Page 021

the first VR systems, as well as to be among a list of world-leading VR experts that also received the Fellowship. I highly encourage any graduate student readers who have already selected a dissertation topic to apply for the Link Foundation Fellowship ($29,000 to be awarded for 2016). Details are available at www.linksim.org.

Meta Platforms, Inc.
Exhibit 1063
Page 022

Figure Credits

Figure 2.1 Based on: Ellis, S. R. (2014). Where are all the Head Mounted Displays? Retrieved April 14, 2015, from http://humansystems.arc.nasa.gov/groups/acd/projects/hmd_dev.php.

Figure 2.3 Courtesy of The National Media Museum / Science & Society Picture Library, United Kingdom.

Figure 2.4 From: Pratt, A. B. (1916). Weapon. US.

Figure 2.5 Courtesy of Edwin A. Link and Marion Clayton Link Collections, Binghamton University Libraries' Special Collections and University Archives, Binghamton University.

Figure 2.6 From: Weinbaum, S. G. (1935, June). Pygmalion's Spectacles. *Wonder Stories.*

Figure 2.7 From: Heilig, M. (1957). Stereoscopic-television apparatus for individual use.

Figure 2.9 From: Comeau, C.P. & Brian, J.S. "Headsight television system provides remote surveil-lance," *Electronics*, November 10, 1961, pp. 86-90.

Figure 2.10 From: Rochester, N., & Seibel, R. (1962). Communication Device. US.

Figure 2.11 (left) Courtesy of Tom Furness.

Figure 2.11 (right) From: Sutherland, I. E. (1968). A Head-Mounted Three Dimensional Display. In *Proceedings of the 1968 Fall Joint Computer Conference AFIPS* (Vol. 33, part 1, pp. 757–764). Copyright © ACM 1968. Used with permission. DOI: 10.1145/1476589.1476686

Figure 2.12 From: Brooks, F. P., Ouh-Young, M., Batter, J. J., & Jerome Kilpatrick, P. (1990). Project GROPE Haptic displays for scientific visualization. *ACM SIGGRAPH Computer Graphics*, *24*(4), 177–185. Copyright © ACM 1990. Used with permission. DOI: 10.1145/97880.97899

Figure 2.13 Courtesy of NASA/S.S. Fisher, W. Sisler, 1988

Figure 3.1 Based on: Milgram, P., & Kishino, F. (1994). Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, *E77-D* (12), 1321–1329. DOI: 10.1.1.102.4646

Figure 3.2 Based on: Jerald, J. (2009). *Scene-Motion-and Latency-Perception Thresholds for Head-Mounted Displays*. Department of Computer Science, University of North Carolina at Chapel Hill.

Figure 3.3 (top left) Courtesy of Oculus VR.

Meta Platforms, Inc.
Exhibit 1063
Page 023

Figure 3.3 (top right) Courtesy of CastAR.

Figure 3.3 (lower left) Courtesy of Marines Magazine.

Figure 3.3 (lower right) From: Jerald, J., Fuller, A. M., Lastra, A., Whitton, M., Kohli, L., & Brooks, F. (2007). Latency compensation by horizontal scanline selection for head-mounted displays. *Proceedings of SPIE, 6490*, 64901Q–64901Q–11. Copyright © 2007 Society of Photo Optical Instrumentation Engineers. Used with permission.

Figure 3.4 (left) From: 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.* Copyright © ACM 1992. Used with permission. DOI: 10.1145/129888.129892

Figure 3.4 (right) From: Daily, M., Sarfaty, R., Jerald, J., McInnes, D., & Tinker, P. (1999). The CABANA: A Re-configurable Spatially Immersive Display. In *Projection Technology Workshop* (pp. 123–132). Used with permission.

Figure 3.5 From: Jerald, J., Daily, M. J., Neely, H. E., & Tinker, P. (2001). Interacting with 2D Applications in Immersive Environments. In *EUROIMAGE International Conference on Augmented Virtual Environments and 3d Imaging* (pp. 267–270). Used with permission.

Figure 3.6 From: Krum, D. M., Suma, E. A., & Bolas, M. (2012). Augmented reality using personal projection and retroreflection. *Personal and Ubiquitous Computing, 16*(1), 17–26. Copyright © 2012 Springer. Used with Kind permission of Springer Science + Business Media. DOI: 10.1007/s00779-011-0374-4

Figure 3.7 Courtesy of USC Institute for Creative Technologies.

Figure 3.9 Courtesy of Tactical Haptics.

Figure 3.10 Courtesy of Dexta Robotics.

Figure 3.11 Courtesy of INITION.

Figure 3.12 Courtesy of Haptic Workstation with HMD at VRLab in EPFL, Losanne, 2005.

Figure 3.13 Courtesy of Shifz, Syntharturalist Art Association.

Figure 3.14 Courtesy of Virtuix.

Figure 3.15 Courtesy of Swissnex San Francisco and Myleen Hollero.

Figure 4.1 Courtesy of Gail Drinnan/Shutterstock.com.

Figure 4.2 Courtesy of NextGen Interactions.

Figure 4.3 Based on: Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518. DOI: 10.1016/j.chb.2010.05.015

Figure II.1 Courtesy of Sleeping cat/Shutterstock.com

Figure 6.2 From: Harmon, Leon D. (1973). The Recognition of Faces. *Scientific American*, *229* (5).© 2015 Scientific American, A Division of Nature America, Inc. Used with permission.

Figure 6.4 Courtesy of Fibonacci.

Meta Platforms, Inc.
Exhibit 1063
Page 024

Figure 6.5b From: Pietro Guardini and Luciano Gamberini, (2007) The Illusory Contoured Tilting Pyramid. Best illusion of the year contest 2007. Sarasota, Florida. Available on line at http://illusionoftheyear.com/cat/top-10-finalists/2007/. Used with permission.

Figure 6.5c From: Lehar, S. (2007). The Constructive Aspect of Visual Perception: A Gestalt Field Theory Principle of Visual Reification Suggests a Phase Conjugate Mirror Principle of Perceptual Computation. Used with permission.

Figure 6.9 Courtesy of PETER ENDIG/AFP/Getty Images.

Figure 6.12 Image © 2015 The Association for Research in Vision and Ophthalmology

Figure 7.1 From: Razzaque, S. (2005). *Redirected Walking.* Department of Computer Science, University of North Carolina at Chapel Hill. Used with permission.
Adapted from: Gregory, R. L. (1973). *Eye and Brain: The Psychology of Seeing* (2nd ed.). London: Weidenfeld and Nicolson. Copyright © 1973 The Orion Publishing Group. Used with permission.

Figure 7.2 Based on: Goldstein, E. B. (2007). *Sensation and Perception* (7th ed.). Belmont, CA: Wadsworth Publishing.

Figure 7.3 Based on: James, T., & Woodsmall, W. (1988). *Time Line Therapy and the Basis of Personality.* Meta Pubns.

Figure 8.2 Based on: Badcock, D. R., Palmisano, S., & May, J. G. (2014). Vision and Virtual Environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of Virtual Environments* (2nd ed.). CRC Press.

Figure 8.3 Based on: Coren, S., Ward, L. M., & Enns, J. T. (1999). *Sensation and Perception* (5th ed.). Harcourt Brace College Publishers.

Figure 8.4 Based on: Clulow, F. W. (1972). *Color: Its principle and their applications.* New York: Morgan & Morgan.

Figure 8.5 Based on: Coren, S., Ward, L. M., & Enns, J. T. (1999). *Sensation and Perception* (5th ed.). Harcourt Brace College Publishers.

Figure 8.6 Based on: Coren, S., Ward, L. M., & Enns, J. T. (1999). *Sensation and Perception* (5th ed.). Harcourt Brace College Publishers.

Figure 8.7 Based on: Goldstein, E. B. (2014). *Sensation and Perception* (9th ed.). Cengage Learning.

Figure 8.9 From: Jerald, J. (2009). *Scene-Motion- and Latency-Perception Thresholds for Head-Mounted Displays.* Department of Computer Science, University of North Carolina at Chapel Hill. Used with permission.

Figure 9.1 Courtesy of Paolo Gianti/Shutterstock.com

Figure 9.5 Courtesy of Galkin Grigory/Shutterstock.com

Figure 9.7 Courtesy of NextGen Interactions.

Meta Platforms, Inc.
Exhibit 1063
Page 025

Figure 9.9 Based on: Coren, S., Ward, L. M., & Enns, J. T. (1999). *Sensation and Perception* (5th ed.). Harcourt Brace College Publishers.

Figure 10.1 From: Pfeiffer, T., & Memili, C. (2015). GPU-Accelerated Attention Map Generation for Dynamic 3D Scenes. In *IIEE Virtual Reality*. Used with permission. Courtesy of "BMW 3 Series Coupe" by mikepan (Creative Commons Attribution, Share Alike 3.0 at http://www.blendswap.com).

Figure 12.1 Based on: Razzaque, S. (2005). *Redirected Walking*. Department of Computer Science, University of North Carolina at Chapel Hill.

Figure 14.1 Courtesy of Jema VR.

Figure 15.1 Based on: Gregory, R. L. (1973). *Eye and Brain: The Psychology of Seeing* (2nd ed.). London: Weidenfeld and Nicolson.

Figure 15.4 From: Jerald, J. (2009). *Scene-Motion- and Latency-Perception Thresholds for Head-Mounted Displays*. Department of Computer Science, University of North Carolina at Chapel Hill. Used with permission.

Figure 18.1 Courtesy of CCP Games.

Figure 18.3 Courtesy of Sixense.

Figure 20.1 From: Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*, 243–259. DOI: 10.2307/1416950

Figure 20.2 From: Lehar, S. (n.d.). The Constructive Aspect of Visual Perception: A Gestalt Field Theory Principle of Visual Reification Suggests a Phase Conjugate Mirror Principle of Perceptual Computation. Used with permission.

Figure 20.3 Courtesy of Sixense.

Figure 20.5 From: Yoganandan, A., Jerald, J., & Mlyniec, P. (2014). Bimanual Selection and Interaction with Volumetric Regions of Interest. In *IEEE Virtual Reality Workshop on Immersive Volumetric Interaction*. Used with permission.

Figure 20.7 Courtesy of Sixense.

Figure 20.9 Courtesy of Sixense.

Figure 21.3 Courtesy of NextGen Interactions.

Figure 21.4 Courtesy of Digital ArtForms.

Figure 21.5 Courtesy of Digital ArtForms.

Figure 21.6 Image © Strangers with Patrick Watson / Félix and Paul Studios

Figure 21.7 Courtesy of 3rdTech.

Figure 21.8 Courtesy of Digital ArtForms.

Figure 21.9 Courtesy of UNC CISMM NIH Resource 5-P41-EB002025 from data collected in Alisa S. Wolberg's laboratory under NIH award HL094740.

Figure 22.1 Courtesy of Digital ArtForms.

Meta Platforms, Inc.
Exhibit 1063
Page 026

Figure 22.2 Courtesy of NextGen Interactions.

Figure 22.4 From: Daily, M., Howard, M., Jerald, J., Lee, C., Martin, K., McInnes, D., & Tinker, P. (2000). Distributed design review in virtual environments. In *Proceedings of the third international conference on Collaborative virtual environments* (pp. 57–63). ACM. Copyright © ACM 2000. Used with permission. DOI: 10.1145/351006.351013

Figure 22.5 Courtesy of NextGen Interactions.

Figure 22.6 Courtesy of VRChat.

Figure 25.1 Based on: Norman, D. A. (2013). *The Design of Everyday Things, Expanded and Revised Edition. Human Factors and Ergonomics in Manufacturing*. New York, NY: Basic Books. DOI: 10.1002/hfm.20127

Figure 26.1 Courtesy of Digital ArtForms.

Figure 26.3 Courtesy of NextGen Interactions.

Figure 26.4 Courtesy of NextGen Interactions.

Figure 26.5 Courtesy of Cloudhead Games.

Figure 27.1 From: Pausch, R., Snoddy, J., Taylor, R., Watson, S., & Haseltine, E. (1996). Disney's Aladdin: first steps toward storytelling in virtual reality. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (pp. 193–203). ACM Press. Copyright © ACM 1996. Used with permission. DOI: 10.1145/237170.237257

Figure 27.2 Courtesy of Stock photo © juniorbeep.

Figure 27.3 Courtesy of Sixense.

Figure 27.4 Courtesy of CyberGlove Systems LLC.

Figure 27.5 Courtesy of Leap Motion.

Figure 27.6 Courtesy of Dassault Systèmes, iV Lab.

Figure 28.1 Courtesy of NextGen Interactions.

Figure 28.2 From: Pierce, J., Forsberg, A., Conway, M., Hong, S., Zeleznik, R., & Miné, M. R. (1997). Image Plane Interaction Techniques in 3D Immersive Environments. In *ACM Symposium on Interactive 3D Graphics* (pp. 39–44). ACM Press. Copyright © ACM 1997. Used with permission. DOI: 10.1145/253284.253303

Figure 28.3 Courtesy of Digital ArtForms.

Figure 28.4 From: Hinckley, K., Pausch, R., Goble, J. C., & Kassell, N. F. (1994). Passive Real-world Interface Props for Neurosurgical Visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94* (Vol. 30, pp. 452–458). Copyright © ACM 1994. Used with permission. DOI: 10.1145/191666.191821

Figure 28.5 Courtesy of Digital ArtForms and Sixense.

Figure 28.6 From: Mlyniec, P., Jerald, J., Yoganandan, A., Seagull, J., Toledo, F., & Schultheis, U. (2011). iMedic: A two-handed immersive medical environment for distributed interactive

Meta Platforms, Inc.
Exhibit 1063
Page 027

consultation. In *Studies in Health Technology and Informatics* (Vol. 163, pp. 372–378). Copyright © 2011 IOS Press. Reprinted with permission.

Figure 28.7 From: Schultheis, U., Jerald, J., Toledo, F., Yoganandan, A., & Mlyniec, P. (2012). Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3D tasks. In *IEEE Symposium on 3D User Interfaces 2012, 3DUI 2012 - Proceedings* (pp. 117–124). Copyright © 2012 IEEE. Used with permission. DOI: 10.1109/3DUI.2012.6184195

Figure 28.8 Courtesy of Sixense and Digital ArtForms.

Figure 28.9 From: Bowman, D. A., & Wingrave, C. A. (2001). Design and evaluation of menu systems for immersive virtual environments. In *IEEE Virtual Reality* (pp. 149–156). Copyright © 2001 IEEE. Used with permission. DOI: 10.1109/VR.2001.913781

Figure 31.1 Based on: Rasmusson, J. (2010). The Agile Samurai–How Agile Masters Deliver Great Software. *Pragmatic Bookshelf, Año*, 280.

Figure 31.2 Based on: Rasmusson, J. (2010). The Agile Samurai–How Agile Masters Deliver Great Software. *Pragmatic Bookshelf, Año*, 280.

Figure 31.6 Courtesy of Geomedia.

Figure 32.1 Courtesy of Andrew Robinson of CCP Games.

Figure 32.2 Based on: Neely, H. E., Belvin, R. S., Fox, J. R., & Daily, M. J. (2004). Multimodal interaction techniques for situational awareness and command of robotic combat entities. In *IEEE Aerospace Conference Proceedings* (Vol. 5, pp. 3297–3305). DOI: 10.1109/AERO.2004.1368136

Figure 32.4 Based on: Daily, M., Howard, M., Jerald, J., Lee, C., Martin, K., McInnes, D., & Tinker, P. (2000). Distributed design review in virtual environments. In *Proceedings of the third international conference on Collaborative virtual environments* (pp. 57–63). ACM. DOI: 10.1145/351006.351013

Figure 33.2 Based on: Gabbard, J. L. (2014). Usability Engineering of Virtual Environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of Virtual Environments* (2nd ed., pp. 721–747). CRC Press.

Figure 33.3 Courtesy of NextGen Interactions.

Figure 35.1 Courtesy of USC Institute for Creative Technologies. Principal Investigators: Albert (Skip) Rizzo and Louis-Philippe Morency.

Figure 35.2 From: Grau, C., Ginhoux, R., Riera, A., Nguyen, T. L., Chauvat, H., Berg, M., . . . Ruffini, G. (2014). Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies. *PloS One, 9*(8), e105225. DOI: 10.1371/journal.pone.0105225

Meta Platforms, Inc.
Exhibit 1063
Page 028

Overview

Virtual reality (VR) can provide our minds with direct access to digital media in a way that seemingly has no limits. However, creating compelling VR experiences is an incredibly complex challenge. When VR is done well, the results are brilliant and pleasurable experiences that go beyond what we can do in the real world. When VR is done badly, not only do users get frustrated, but they can get sick. There are many causes of bad VR; some failures come from the limitations of technology, but many come from a lack of understanding perception, interaction, design principles, and real users. This book discusses these issues by emphasizing the human element of VR. The fact is, if we do not get the human element correct, then no amount of technology will make VR anything more than an interesting tool confined to research laboratories. Even when VR principles are fully understood, the first implementation is rarely novel and almost never ideal due to the complex nature of VR and the countless possibilities that can be created. The VR principles discussed in this book will enable readers to intelligently experiment with the rules and iteratively design toward innovative experiences.

Historically, most VR creators have been engineers (the author included) with expertise in technology and logic but limited in their understanding of humans. This is primarily due to the fact that VR has previously been so technically challenging that it was difficult to build VR experiences without an engineering background. Unfortunately, we engineers often believe "I am human, therefore I understand other humans and what works for them." However, the way humans perceive and interact with the world is incredibly complex and not typically based on logic, mathematics, or a user manual. If we stick solely to the expertise of knowing it all through engineering and logic, then VR will certainly be doomed. We have to accept human perception and behavior the way it is, not the way logic tells us it should be. Engineering will always be essential as it is the core of VR systems that everything else builds upon, but VR itself presents a fascinating interplay of technology and psychology, and we must understand both to do VR well.

Meta Platforms, Inc.
Exhibit 1063
Page 029

Technically minded people tend to dislike the word "experience" as it is less logical and more subjective in nature. But when asked about their favorite tool or game they often convey emotion as they discuss how, for example, the tool responds. They talk about how it makes them feel often without them consciously realizing they are speaking in the language of emotion. Observe how they act when attempting to navigate through a poorly designed voice-response system for technical support via phone. In their frustration, they very well may give up on the product they are calling about and never purchase from that company again. The experience is important for everything, even for engineers, for it determines our quality of life on a moment-by-moment basis. For VR, the experience is even more critical. To create quality VR, we need to continuously ask ourselves and others questions about how we perceive the VR worlds that we create. Is the experience understandable and enjoyable? Or is it sometimes confusing to use while at other times just all-out frustrating and sickness inducing? If the VR experience is intuitive to understand, easily controlled, and comfortable, then just like anything in life, there will be a sense of mastery and satisfaction with "logical" reasons why it works well. Emotion and cognition are tightly coupled, with cognition more often than not justifying decisions that were actually made emotionally. We must create VR experiences with both emotion and logic.

What This Book Is

This book focuses on **human-centered design**, a design philosophy that puts human needs, capabilities, and behavior first, then designs to accommodate those needs, capabilities, and ways of behaving (Norman, 2013). More specifically, it focuses on the human elements of VR—how users perceive and intuitively interact with various forms of reality, causes of VR sickness, creating content that is pleasing and useful, and how to design and iterate upon effective VR applications.

Good VR design starts with an understanding of both technology and perception. It requires good communication between human and machine, indicating what interactions are possible, what is currently occurring, and what is about to occur. Human-centered design also comes about through observation, for humans are often unaware of their perceptual processes and methods of interacting (at least for VR that works well). Getting the specification of a VR experience is difficult, and rarely does a VR creator do it well for his first few projects. In fact, even VR experts do not perfectly define the project from the start if they are creating a novel experience. A human-centered design principle, like lean methods, is to avoid completely defining the problem at the start and to iterate upon repeated approximations and modifications through rapid tests of ideas with real users.

Meta Platforms, Inc.
Exhibit 1063
Page 030

Developing intuitive VR should not be driven by software/hardware engineering considerations alone (e.g., we need to do much more than figure out how to efficiently display at the highest resolution available based on the most current hardware). A good portion of this book is devoted to how the human mind works in order to help readers create higher0quality VR applications. VR design, however, goes beyond just technology and psychology; VR is intensely multidisciplinary. VR is an incredibly complex challenge and the study, design, and implementation of high-quality VR requires an understanding of various disciplines, including the behavioral and social sciences, neuroscience, information and computer science, physics, communications, art, and even philosophy. When reflecting back upon VR design and implementation, Wingrave and LaViola (2010) point out: "Practitioners must be carpenters, electricians, engineers, artists, and masters of duct tape and Velcro." This book takes a broad perspective, applying insights from various disciplines to VR design.

In summary, this book provides basic theory, an overview of various concepts useful for VR, examples that put the theory and concepts into more understandable form, useful guidelines, and a foundation for further exploration and interaction design for virtual worlds that do not yet exist.

What This Book Is Not

There are more questions about VR than there are answers and the intent of this book is not to attempt to provide all of the answers. VR covers a broad range of imaginary spaces broader than the real world; nobody could possibly have all the answers to the real world, so it is unreasonable to expect to find all the answers for virtual worlds. Instead, this book attempts to help readers build and iterate upon creative answers and compelling experiences. Although this book can't possibly cover all aspects of VR in detail, it does provide an overview of various topics and delves more deeply into those that are most important. References are provided throughout for those wishing to further study concepts of interest.

In some cases, the concepts presented in this book follow well-understood principles or conclusive research (although even conclusive research rarely holds 100% of the time across all conditions). In other cases, concepts are not the "truth" but have been found to be useful in the way we think about design and interaction. Studying theory can be useful, but VR development should always follow pragmatism over theory.

Although there is a brief chapter on getting started (Chapter 36), and there are tips throughout on high- and mid-level implementation concepts, this book is not a step-by-step tutorial of how to implement an example VR system. In fact, this book intentionally contains no code or equations so that all concepts can be understood

Meta Platforms, Inc.
Exhibit 1063
Page 031

by anyone from any discipline (references are provided for more rigorous detail). Although researchers have been experimenting with VR for decades (Section 2.1), VR in no way comes close to reaching its full potential. Not only are there many unknowns, but VR implementation very much depends on the project. For example, a surgical training system is very different from an immersive film.

Who This Book Is For

This book is for the entire team that works on a VR project, not just for those who define themselves as designers. It is intended to act as a foundation for anyone and everyone involved with creating VR experiences. The book also is meant to serve as a bridge between academic research and practical advice to a wide range of individuals who wish to build compelling experiences. This includes designers, managers, programmers, artists, psychologists, engineers, students, educators, and user experience professionals so the entire team can have a common understanding and language of VR. Everyone involved with a VR project should understand at least the basics of perception, VR sickness, interaction, content creation, and iterative design. VR requires specialized experts in various disciplines to each contribute in their own unique way, but we each must also know at least a little about human-centered design in order to effectively communicate with teammates and to integrate the various components together into a seamless, quality experience.

How to Read This Book

Readers may wish to read and use this book differently depending on their background, particular interests, and how they would like to apply it.

Newcomers

Those completely new to VR who want a high-level understanding will most appreciate Part I, Introduction and Background. After reading Part I, the reader may want to skip ahead to Part VII, The Future Starts Now. Once these basics are understood, most of Part IV, Content Creation, should be able to be understood. As the reader learns more about VR, the other parts will become easier to digest.

Teachers

This book will be especially relevant to interdisciplinary VR courses. Teachers will want to choose chapters that are most relevant to the course requirements and student interests. It is highly suggested that any course take a project-centered approach. In such a case, the teacher should consider the suggestions below for both students and

Meta Platforms, Inc.
Exhibit 1063
Page 032

practitioners. An outline for starting a first project, albeit slightly different for a class project, is outlined in Chapter 36, Getting Started.

Students

Students will gain a core understanding of VR by first gaining a high-level overview of VR through Part I, Introduction and Background. For those wishing to understand theory, Part II, Perception, and Part III, Adverse Health Effects, will be invaluable. For students working on VR projects, they should also follow the advice below for practitioners.

Practitioners

Practitioners who want to immediately get the most important points that apply to their VR creations will want to start with the practitioner chapters that are highlighted in the table of contents (mostly Parts IV–VI). In particular, they may want to start with the design guidelines chapters at the end of each part, where each part contains multiple chapters on one of the primary topics. Most of the guidelines provide back references to the relevant sections for more detailed information.

VR Experts

VR experts will likely use this book more as a reference so they do not spend time reading material they are already familiar with. The references will also serve VR experts for further investigation. Those experts who primarily work with head-mounted displays may find Part I, Introduction and Background, useful to understand how head-mounted displays fit within the larger scheme of other implementations of VR and augmented reality (AR). Those interested in implementing straightforward applications that have already been built may not find Part II, Perception, useful. However, for those who want to innovate with new forms of VR and interaction, they may find this part useful to understand how we perceive the world in order to help invent novel creations.

Overview of the Seven Parts

Part I, Introduction and Background, provides a background of VR including a brief history of VR, different forms of VR and related technologies, and a broad overview of some of the most important concepts that will be further discussed in later parts.

Part II, Perception, provides a background in perception to educate VR creators on concepts and theories of how we perceive and interact with the world around us.

Meta Platforms, Inc.
Exhibit 1063
Page 033

This part serves as an intellectual framework that will enable the reader to not only implement the ideas discussed in later chapters but more thoroughly understand why some techniques do or do not work, to extend those techniques, to intelligently experiment with new concepts that have a better chance of working without causing human factors issues, and to know when it might be appropriate to break the rules.

Part III, Adverse Health Effects, describes one of the most difficult challenges of VR and helps to reduce the greatest risk to VR succeeding at a massive scale: VR sickness. Whereas it may be impossible to remove 100% of VR sickness for the entire population, there are several ways to dramatically reduce it if we understand the theories of why it occurs. Other adverse health effects such as risk of injury, seizures, and aftereffects are also discussed.

Part IV, Content Creation, discusses high-level concepts for designing/building assets and how subtle design choices can influence user behavior. Examples include story creation, the core experience, environmental design, wayfinding aids, social networking, and porting existing content to VR.

Part V, Interaction, focuses on how to design the way users interact within the scenes they find themselves in. For many applications, we want to engage the user by creating an active experience that consists of more than simply looking around; we want to empower users by enabling them to reach out, touch, and manipulate that world in a way that makes them feel they are a part of the world instead of just a passive observer.

Part VI, Iterative Design, provides an overview of several different methods for creating, experimenting, and improving upon VR designs. Whereas each project may not utilize all methods, it is still good to understand them all to be able to apply them when appropriate. For example, you may not wish to conduct a formal and rigorous scientific user study, but you do want to understand the concepts to minimize mistaken conclusions due to confounding factors.

Part VII, The Future Starts Now, summarizes the book, discusses the current and future state of VR, and provides a brief plan to get started.

Meta Platforms, Inc.
Exhibit 1063
Page 034

PART

INTRODUCTION AND BACKGROUND

What is virtual reality (VR)? What does VR consist of and for what situations is it useful? What is different about VR that gets people so excited? How do developers engage users so that they feel present in a virtual environment? This part of the book answers such questions, and provides a basic background that later chapters build upon. This introduction and background serves as a simple high-level toolbox of options to intelligently choose from, such as different forms of virtual and augmented reality (AR), different hardware options, various methods of presenting information to the senses, and ways to induce presence into the minds of users.

Part I consists of five chapters that cover the basics of VR.

Chapter 1, What Is Virtual Reality?, begins by describing what VR is at a high level and what it is suitable/effective for. This includes descriptions of different forms of communication that are at the heart of what VR is—communication between the user and a system created by the VR designer.

Chapter 2, A History of VR, provides a history of VR starting with stereoscopes created in the 1800s. The concept and implementation of VR is not new.

Chapter 3, An Overview of Various Realities, discusses forms of reality ranging from the real world to augmented reality (AR) to VR. Whereas the focus of this book is on fully immersive VR, this chapter provides context of where VR fits into the overall picture of related technologies. The chapter also gives a high-level description of various forms of input and output hardware options that can be used as part of AR and VR systems.

Meta Platforms, Inc.
Exhibit 1063
Page 035

Chapter 4, Immersion, Presence, and Reality Trade-Offs, discusses the often-used terms of immersion and presence. Readers may be surprised to learn that realism is not necessarily the goal of VR and there are trade-offs for attempting to perfectly simulate reality, even if reality could be perfectly simulated.

Chapter 5, The Basics: Design Guidelines, concludes this introductory part of the book and gives a small number of guidelines for those looking to create VR experiences.

Meta Platforms, Inc.
Exhibit 1063
Page 036

What Is Virtual Reality?

1.1 The Definition of Virtual Reality

The term virtual reality (VR) is commonly used by the popular media to describe imaginary worlds that only exist in computers and our minds. However, let us more precisely define the term. Sherman and Craig (2003) point out in their book *Understanding Virtual Reality* that *Webster's New Universal Unabridged Dictionary* (1989) defines virtual as "being in essence or effect, but not in fact" and reality as "the state or quality of being real. Something that exists independently of ideas concerning it. Something that constitutes a real or actual thing as distinguished from something that is merely apparent." Thus, virtual reality is a term that contradicts itself—an oxymoron! Fortunately, the website merriam-webster.com (Merriam-Webster, 2015) has more recently defined the full term virtual reality to be "an artificial environment which is experienced through sensory stimuli (as sights and sounds) provided by a computer and in which one's actions partially determine what happens in the environment." In this book, **virtual reality** is further defined to be a computer-generated digital environment that can be experienced and interacted with as if that environment were real.

An ideal VR system enables users to physically walk around objects and touch those objects as if they were real. Ivan Sutherland, the creator of one of the world's first VR systems in the 1960s, stated (Sutherland, 1965): "The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal." We haven't yet come anywhere near Ivan Sutherland's vision (nor do we necessarily want to!) and perhaps we never will. However, there are some quite engaging virtual realities today—many of which are featured throughout this book.

Meta Platforms, Inc.
Exhibit 1063
Page 037

1.2 VR Is Communication

Normally, communication is thought of as interaction between two or more people. This book defines **communication** more abstractly: the transfer of energy between two things, even if just the cause and effect of one object colliding with another object. Communication can also be between human and technology—an essential component and basis of VR. VR design is concerned with the communication of how the virtual world works, how that world and its objects are controlled, and the relationship between user and content—ideally where users are focused on the experience rather than the technology.

Well-designed VR experiences can be thought of as collaboration between human and machine where both software and hardware work harmoniously together to provide intuitive communication with the human. Developers write complex software to create, if designed well, seemingly simple transfer functions to provide effective interactions and engaging experiences. Communication can be broken down into direct communication and indirect communication as discussed below.

1.2.1 Direct Communication

Direct communication is the direct transfer of energy between two entities with no intermediary and no interpretation attached. In the real world, pure direct communication between entities doesn't represent anything as the purpose is not communication, but it is a side effect. However, in VR, developers insert an artificial intermediary (the VR system that is ideally unperceivable) between the user and carefully controlled sensory stimuli (e.g., shapes, motions, sounds). When the goal is direct communication, VR creators should focus on making the intermediary transparent so users feel like they have direct access to those entities. If that can be achieved, then users will perceive, interpret, and interact with stimuli as if they are directly communicating with the virtual world and its entities.

Direct communication consists of structural communication and visceral communication.

Structural Communication

Structural communication is the physics of the world, not the description or the mathematical representation but the thing-in-itself (Kant, 1781). An example of structural communication is the bouncing of a ball off of the hand. We are always in relationship to objects, which help to define our state—e.g., the shape of our hand around a controller. The world, as well as our own bodies, directly tells us what the structure is through our senses. Although thinking and feeling do not exist within structural

Meta Platforms, Inc.
Exhibit 1063
Page 038

communication, such communication does provide the starting point for perception, interpretation, thinking, and feeling.

In order for our ideas to persist through time, we must put those ideas into structural form, what Norman (2013) calls knowledge in the world. Recorded information and data is the obvious example of structural form, but sometimes less obvious structural forms are the signifiers and constraints (Section 25.1) of interaction. In order to induce experiences into others through VR, we present structural stimuli (e.g., pixels on a display, sound through headphones, or the rumble/vibration of a controller) so the users can sense and interact with our creations.

Visceral Communication

Visceral communication is the language of automatic emotion and primal behavior, not the rational representation of the emotions and behavior (Section 7.7). Visceral communication is always present for humans and is the in-between of structural communication and indirect communication. Presence (Chapter 4) is the act of being fully engaged via direct communication (albeit primarily one way). Examples of visceral communication are the feeling of awe while sitting on a mountaintop, looking down at the earth from space, or being with someone via solid eye contact (whether in the real world or via avatars in VR). The full experience of such visceral communication cannot be put into words, although we often attempt to do so at which point the experience is transformed into indirect communication (e.g., explaining VR to someone is not the same as experiencing VR).

1.2.2 Indirect Communication

Indirect communication connects two or more things through some intermediary. The intermediary need not be physical; in fact, the intermediary is often our mind's interpretation that sits between the world and behavior/action. Once we interpret and give something meaning, then we have transformed the direct communication into indirect communication. Indirect communication includes what we normally think of as language, such as spoken and written language, as well as sign languages and our internal thoughts (i.e., communicating with oneself). Indirect communication consists of talking, understanding, creating stories/histories, giving meaning, comparing, negating, fantasizing, lying, and romancing. These are not part of the objective real world but are what we humans describe and create with our minds. Indirect VR communication includes the user's internal mental model (Section 7.8) of how the VR world works (e.g., the interpretation of what is occurring in the VR world), and indirect interactions (Section 28.4) such as moving a slider that changes an object property,

Meta Platforms, Inc.
Exhibit 1063
Page 039

speech recognition that changes the system's state, and indirect gestures that act as a sign language to the computer.

1.3 What Is VR Good For?

The recent surge in media coverage has inspired the public to become quite excited about the potential for VR. This coverage has focused on the entertainment industry, specifically video games and immersive film. VR is a great fit for the entertainment industry and will certainly be the driving force behind VR in the short term. However, what is VR good for beyond entertainment? It turns out VR can have enormous benefit over a wide range of verticals. VR has been successfully deployed in various industries for many years now. Successful applications include oil and gas exploration, scientific visualization, architecture, flight simulation, therapy, military training, theme-park entertainment, engineering analysis, and design review. Using VR in such situations has successfully revealed costly design mistakes before manufacturing anything, reduced time to market by speeding up iterative processes, provided safe learning environments that would otherwise be dangerous, reduced PTSD by gradually increasing exposure to feared stimuli, and helped to visualize large datasets that would be difficult to comprehend with traditional systems.

Unfortunately, to date, VR has largely been limited to well-funded academic and corporate research labs to which few have access. That is all changing with consumer-priced systems now becoming widely available. The VR market is expected to first explode in the entertainment industry but will soon expand significantly in other industries. Education, telepresence, and professional training will likely be the next industries to take advantage of VR in a big way.

Whatever the industry, VR is largely about providing understanding—whether that is understanding an entertaining story, learning an abstract concept, or practicing a real skill. Actively using more of the human sensory capability and motor skills has been known to increase understanding/learning for some time (Dale, 1969). This is in part due to the increased sensory bandwidth between human and information, but there is much more to understanding. Actively participating in an action, making concepts intuitive, encouraging motivation through engaging experiences, and the thoughts inside one's head all contribute to understanding. This book focuses on how to design such concepts into VR experiences.

Figure 1.1 shows Edgar Dale's Cone of Experience (Dale, 1969). As can be seen in the figure, direct purposeful experiences provide the best basis for understanding. As Confucius stated, "I see and I forget. I hear and I remember. I do and I understand." Note this diagram does not suggest direct purposeful experiences should be the only

Meta Platforms, Inc.
Exhibit 1063
Page 040

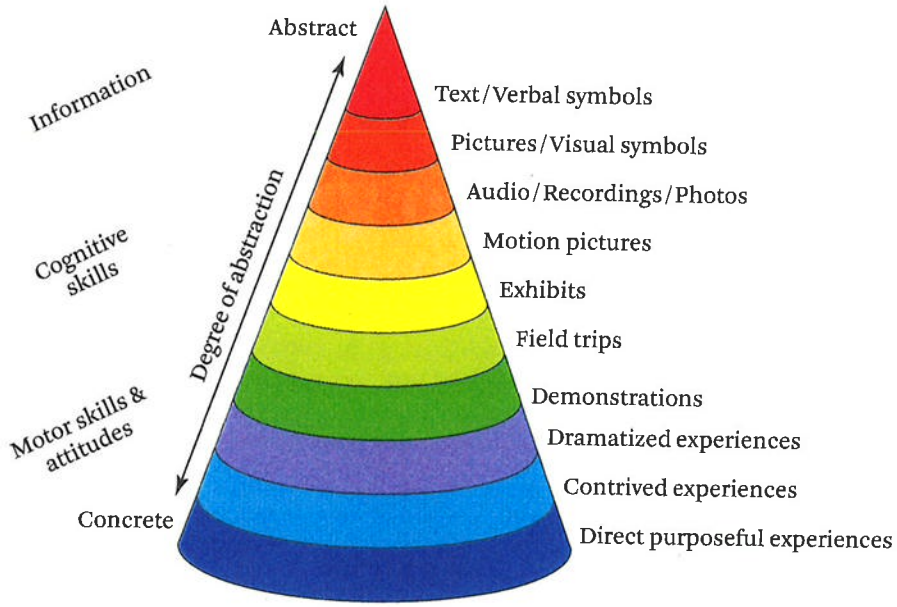


Figure 1.1 **The Cone of Experience. VR uses many levels of abstraction.**

or best method of learning, but instead describes the progression of learning experience. Adding other indirect information within direct purposeful VR experiences can further enhance understanding. For example, embedding abstract information such as text, symbols, and multimedia directly into the scene and onto virtual objects can lead to more efficient understanding than what can be achieved in the real world.

Meta Platforms, Inc.
Exhibit 1063
Page 041

A History of VR

When anything new comes along, everyone, like a child discovering the world, thinks that they've invented it, but you scratch a little and you find a caveman scratching on a wall is creating virtual reality in a sense. What is new here is that more sophisticated instruments give you the power to do it more easily.

—Morton Heilig (Hamit, 1993)

Precursors to what we think of today as VR go back as far as humans have had imaginations and the ability to communicate through the spoken word and cave drawings (what could be called analog VR). The Egyptians, Chaldeans, Jews, Romans, and Greeks used magical illusions to entertain and control the masses. In the Middle Ages, magicians used smoke and concave mirrors to produce faint ghost and demon illusions to gull naive apprentices as well as larger audiences (Hopkins, 2013). Although the words and implementation have changed over the centuries, the core goals of creating the illusion of conveying that which is not actually present and capturing our imaginations remain the same.

2.1 The 1800s

The static version of today's stereoscopic 3D TVs is called a stereoscope and was invented before photography in 1832 by Sir Charles Wheatstone (Gregory, 1997). As shown in Figure 2.2, the device used mirrors angled at 45° to reflect images into the eye from the left and right side.

David Brewster, who earlier invented the kaleidoscope, used lenses to make a smaller consumer-friendly hand-held stereoscope (Figure 2.3). His stereoscope was demonstrated at the 1851 Exhibition at the Crystal Palace where Queen Victoria found it quite pleasing. Later the poet Oliver Wendell Holmes stated ". . . is a surprise such as no painting ever produced. The mind feels its way into the very depths of the picture." (Zone, 2007). By 1856 Brewster estimated over a half million stereoscopes had been sold (Brewster, 1856). This first 3D craze included various forms of the stereoscope,

Meta Platforms, Inc.
Exhibit 1063
Page 042

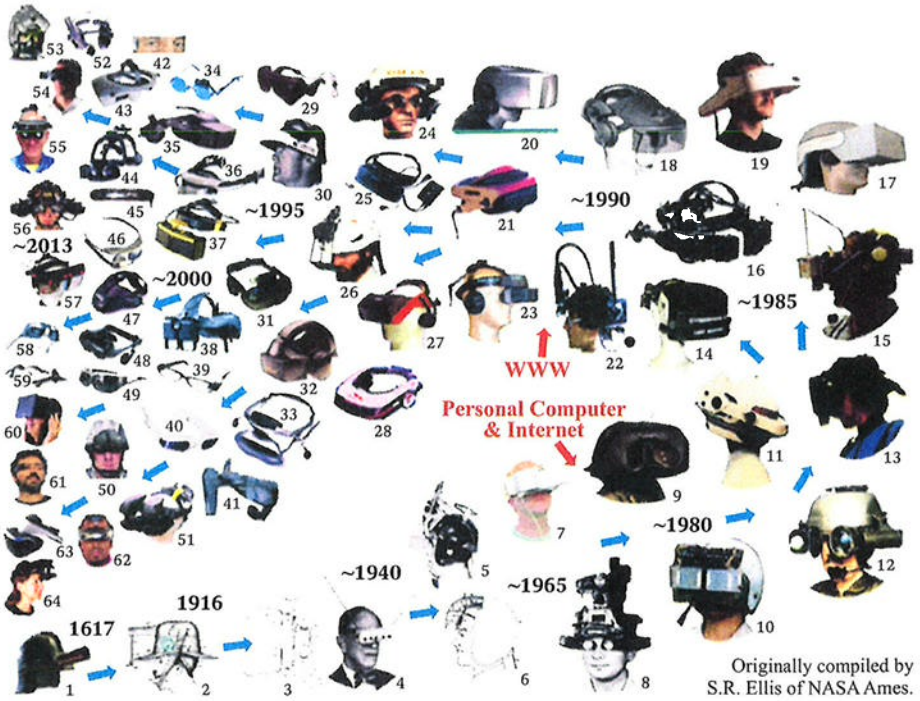


Figure 2.1 **Some head-mounted displays and viewers over time.** (Originally compiled from Ellis (2014))

including self-assembled cardboard versions with moving images controlled by the hand in 1860 (Zone, 2007). One company alone sold a million stereoscopic views in 1862. Brewster's design is conceptually the same as the 20th century View-Master and today's Google Cardboard. In the case of Google Cardboard and similar phone-based VR systems, a cellular phone is used to display the images in place of the actual physical images themselves.

Many years later, a 360° VR-type display known as the Haunted Swing was shown at the '95 Midwinter Fair in San Francisco, and is still one of the most compelling technical demonstrations of an illusion to this day. The demo consisted of a room and large swing that held approximately 40 people. After the audience seated themselves, the swing was put in motion and as the swing oscillated, users felt motion similar to being in an elevator while they involuntarily clutched their seats. In fact, the swing hardly moved at all, but the surrounding room moved substantially, resulting in the sense of self-motion (Section 9.3.10) and motion sickness (Chapter 12). The date was not 1995, but was 1895 (Wood, 1895).

Meta Platforms, Inc.
Exhibit 1063
Page 043



Figure 2.2 **Charles Wheatstone's stereoscope.** (From http://www.boiledbeans.net/2011/05/19/before-james-cameron; accessed April 10, 2015)



Figure 2.3 **A Brewster stereoscope from 1860.**

Meta Platforms, Inc.
Exhibit 1063
Page 044

It was also in 1895 that film began to go mainstream; and when the audience saw a virtual train coming at them through the screen in the short film "L'Arrivée d'un train en gare de La Ciotat," some people reportedly screamed and ran to the back of the room. Although the screaming and running away is more rumor than verified reports, there was certainly hype, excitement, and fear about the new artistic medium, perhaps similar to what is happening with VR today.

2.2 The 1900s

VR-related innovation continued in the 1900s that moved beyond simply presenting visual images. New interaction concepts started to emerge that might be considered novel for even today's VR systems. For example, Figure 2.4 shows what is a head-worn gun pointing and firing device patented by Albert Pratt in 1916 (Pratt, 1916). No hand tracking is required to fire this device as the interface consists of a tube the user blows through.

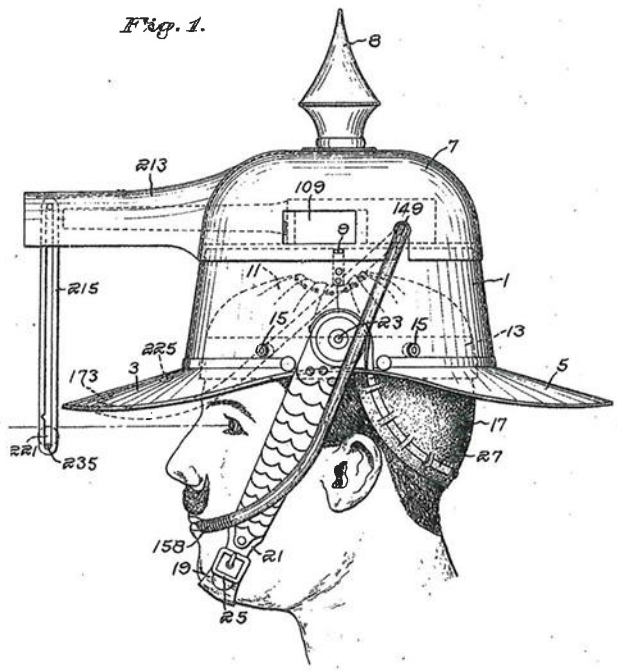


Figure 2.4 **Albert Pratt's head-mounted targeting and gun-firing interface.** (From Pratt (1916))

Meta Platforms, Inc.
Exhibit 1063
Page 045

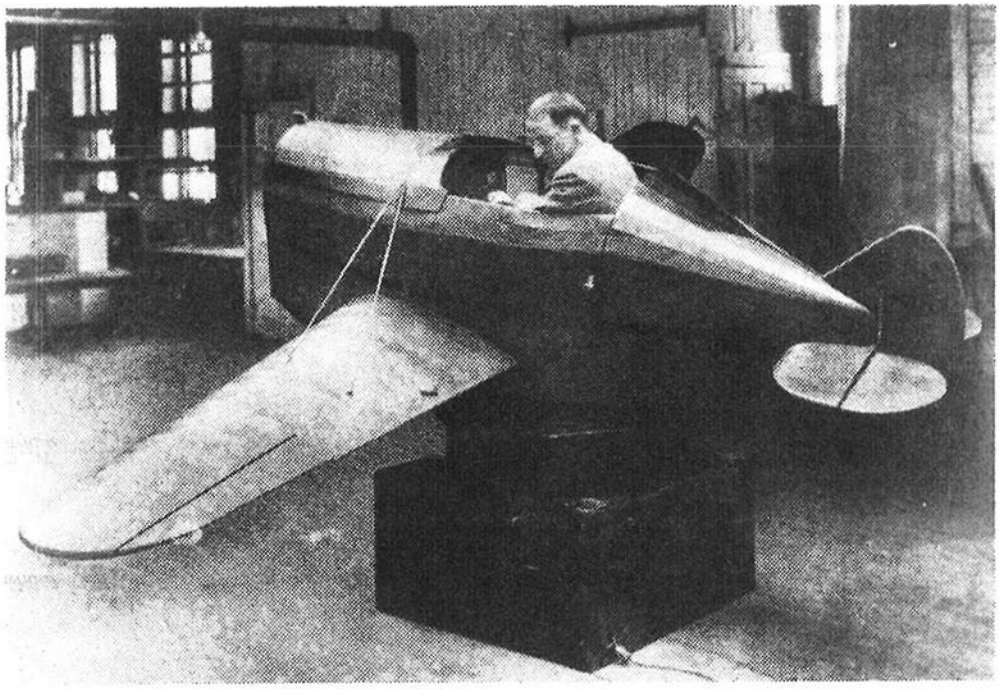


Figure 2.5 **Edwin A. Link and the first flight simulator in 1928.** (Image courtesy of Edwin A. Link and Marion Clayton Link Collections, Binghamton University Libraries' Special Collections and University Archives, Binghamton University)

A little over a decade after Pratt received his weapon patent, Edwin Link developed the first simple mechanical flight simulator, a fuselage-like device with a cockpit and controls that produced the motions and sensations of flying (Figure 2.5). Surprisingly, his intended client—the military—was not initially interested, so he pivoted to selling to amusement parks. By 1935, the Army Air Corps ordered six systems and by the end of World War II, Link had sold 10,000 systems. Link trainers eventually evolved into astronaut training systems and advanced flight simulators complete with motion platform and real-time computer-generated imagery, and today is Link Simulation & Training, a division of L-3 Communications. Since 1991, the Link Foundation Advanced Simulation and Training Fellowship Program has funded many graduate students in their pursuits of improving upon VR systems, including work in computer graphics, latency, spatialized audio, avatars, and haptics (Link, 2015).

As early 20th century technologies started to be built, science fiction and questions inquiring about what makes reality started to become popular. In 1935, for example,

Meta Platforms, Inc.
Exhibit 1063
Page 046


PYGMALION'S SPECTACLES

By STANLEY G. WEINBAUM

Author of "The Black Flame," "A Martian Odyssey," etc.

© 1935 by Continental Publications, Inc.

Figure 2.6 ***Pygmalion's Spectacles* is perhaps the first science fiction story written about an alternate world that is perceived through eyeglasses and other sensory equipment.** (From Weinbaum (1935))

science fiction readers got excited about a surprisingly similar future that we now aspire to with head-mounted displays and other equipment through the book *Pygmalion's Spectacles* (Figure 2.6). The story opens with the words "But what is reality?" written by a professor friend of George Berkeley, the Father of Idealism (the philosophy that reality is mentally constructed) and for whom the University of California, Berkeley, is named. The professor then explains a set of eyeglasses along with other equipment that replaces real-world stimuli with artificial stimuli. The demo consists of a quite compelling interactive and immersive world where "The story is all about you, and you are in it" through vision sound, taste, smell, and even touch. One of the virtual characters calls the world Paracosma—Greek for "land beyond-the-world." The demo is so good that the main character, although skeptical at first, becomes convinced that it is no longer illusion, but reality itself. Today there are countless books ranging from philosophy to science fiction that discuss the illusion of reality.

Perhaps inspired by *Pygmalion's Spectacles,* McCollum patented the first stereoscopic television glasses in 1945. Unfortunately, there is no record of the device ever actually having been built.

In the 1950s Morton Heilig designed both a head-mounted display and a world-fixed display. The head-mounted display (HMD) patent (Heilig, 1960) shown in Figure 2.7 claims lenses that enable a 140° horizontal and vertical field of view, stereo

Meta Platforms, Inc.
Exhibit 1063
Page 047

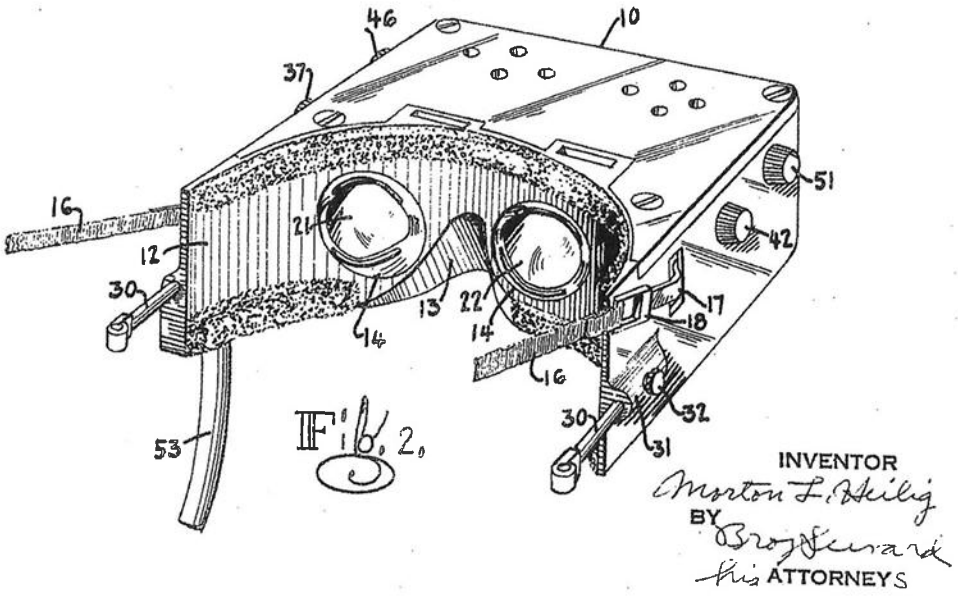


Figure 2.7 **A drawing from Heilig's 1960 Stereoscopic Television Apparatus patent.** (From Heilig (1960))

earphones, and air discharge nozzles that provide a sense of breezes at different temperatures as well as scent. He called his world-fixed display the Sensorama. As can be seen in Figure 2.8, the Sensorama was created for immersive film and it provided stereoscopic color views with a wide field of view, stereo sounds, seat tilting, vibrations, smell, and wind (Heilig, 1992).

In 1961, Philco Corporation engineers built the first actual working tracked HMD that included head tracking (Figure 2.9). As the user moved his head, a camera in a different room moved so the user could see as if he were at the other location. This was the world's first working telepresence system.

One year later IBM was awarded a patent for the first glove input device (Figure 2.10). This glove was designed as a comfortable alternative to keyboard entry, and a sensor for each finger could recognize multiple finger positions. Four possible positions for each finger with a glove in each hand resulted in 1,048,575 possible input combinations. Glove input, albeit with very different implementations, later became a common VR input device in the 1990s.

Starting in 1965, Tom Furness and others at the Wright-Patterson Air Force Base worked on visually coupled systems for pilots that consisted of head-mounted displays

Meta Platforms, Inc.
Exhibit 1063
Page 048



Figure 2.8 **Morton Heilig's Sensorama created the experience of being fully immersed in film.**

(Figure 2.11, left). While Furness was developing head-mounted displays at Wright-Patterson Air Force Base, Ivan Sutherland was doing similar work at Harvard and the University of Utah. Sutherland is known as the first to demonstrate a head-mounted display that used head tracking and computer-generated imagery (Oakes, 2007). The system was called the Sword of Damocles (Figure 2.11, right), named after the story of King Damocles who, with a sword hanging above his head by a single hair of a horse's tail, was in constant peril. The story is a metaphor that can be applied to VR technology: (1) with great power comes great responsibility; (2) precarious situations give a sense of foreboding; and (3) as stated by Shakespeare (1598) in *Henry IV*, "uneasy lies the head that wears a crown." All these seem very relevant for both VR developers and VR users even today.

Meta Platforms, Inc.
Exhibit 1063
Page 049

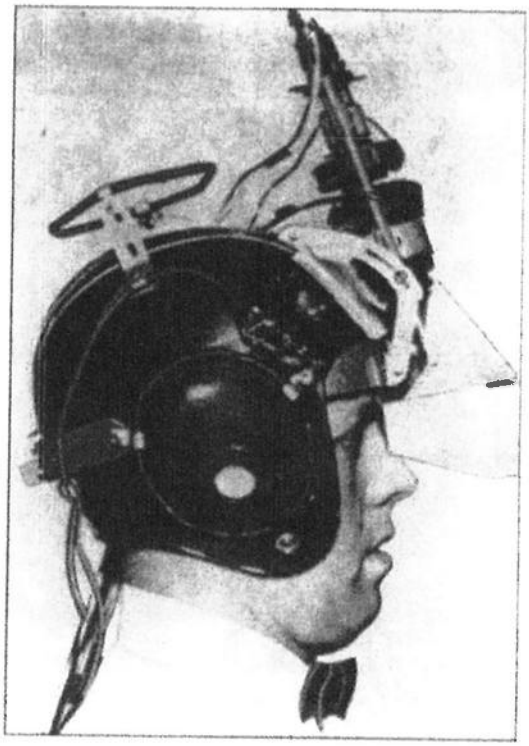


Figure 2.9 **The Philco Headsight from 1961.** (From Comeau and Brian (1961))

Dr. Frederick P. Brooks, Jr., inspired by Ivan Sutherland's vision of the Ultimate Display (Sutherland, 1965), established a new research program in interactive graphics at the University of North Carolina at Chapel Hill, with the initial focus being on molecular graphics. This not only resulted in a visual interaction with simulated molecules but also included force feedback where the docking of simulated molecules could be felt. Figure 2.12 shows the resulting Grope-III system that Dr. Brooks and his team built. UNC has since focused on building various VR systems and applications with the intent to help practitioners solve real problems ranging from architectural visualization to surgical simulation.

In 1982, Atari Research, led by legendary computer scientist Alan Kay, was formed to explore the future of entertainment. The Atari research team, which included Scott Fisher, Jaron Lanier, and Thomas Zimmerman, brainstormed novel ways of interacting with computers and designed technologies that would soon be essential for commercializing VR systems.

Meta Platforms, Inc.
Exhibit 1063
Page 050

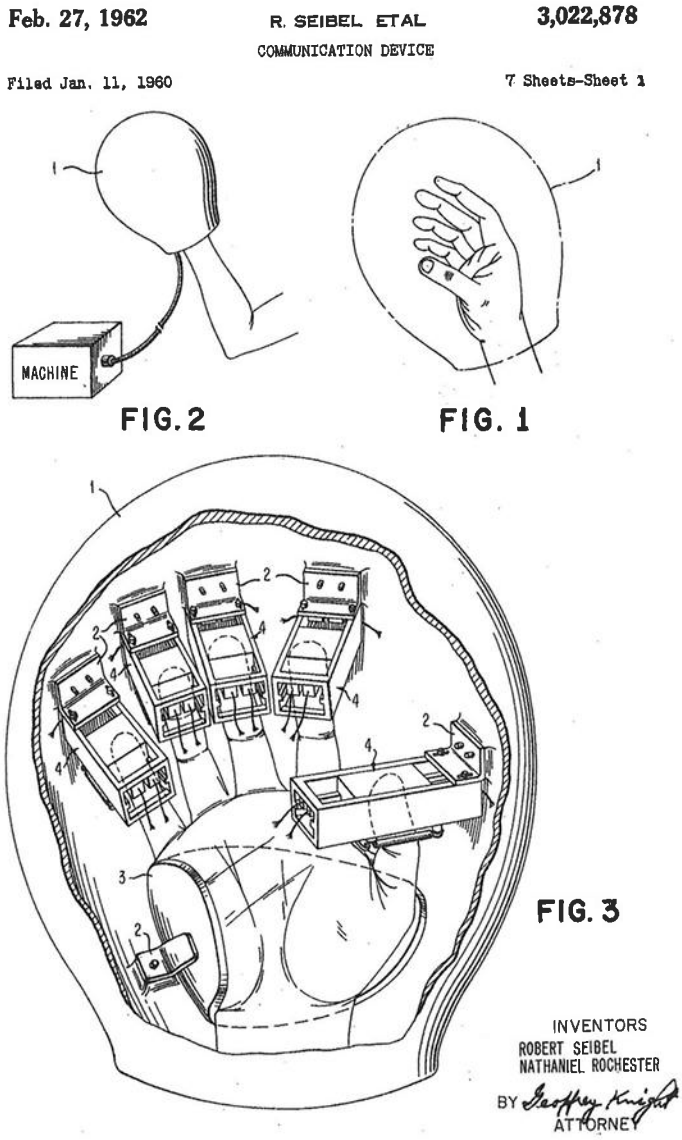


Figure 2.10 **An image from IBM's 1962 glove patent.** (From Rochester and Seibel (1962))

Meta Platforms, Inc.
Exhibit 1063
Page 051

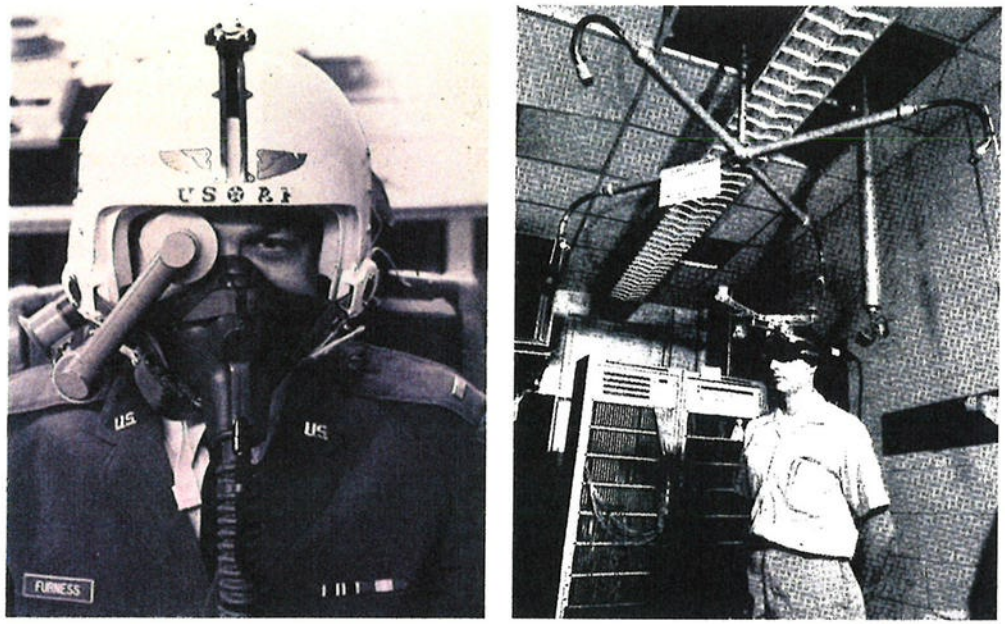


Figure 2.11 **The Wright-Patterson Air Force Base head-mounted display from 1967 (image courtesy of Tom Furness) and the Sword of Damocles (Sutherland, 1968).**

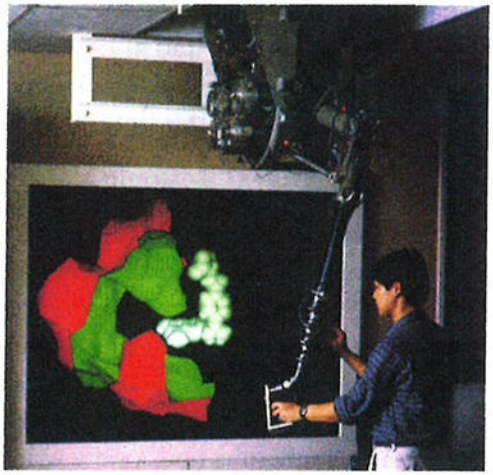


Figure 2.12 **The Grope-III haptic display for molecular docking.** (From Brooks et al. (1990))

Meta Platforms, Inc.
Exhibit 1063
Page 052



Figure 2.13 **The NASA VIEW System.** (Courtesy NASA/S.S. Fisher, W. Sisler, 1988)

In 1985, Scott Fisher, now at NASA Ames, along with other NASA researchers developed the first commercially viable, stereoscopic head-tracked HMD with a wide field of view, called the Virtual Visual Environment Display (VIVED). It was based on a scuba diver's face mask with the displays provided by two Sony Watchman handheld televisions (Costello, 1997). This system was unprecedented as the HMD could be produced at a relatively affordable price, and as a result the VR industry was born. Figure 2.13 shows a later system called the VIEW (Virtual Interface Environment Workstation) system.

Jaron Lanier and Thomas Zimmerman left Atari in 1985 to start VPL Research (VPL stands for Visual Programming Language) where they built commercial VR gloves, head-mounted displays, and software. During this time Jaron coined the term "virtual reality." In addition to building and selling head-mounted displays, VPL built the Dataglove—a VR glove with optical flex sensors to measure finger bending and tactile vibrator feedback (Zimmerman et al., 1987).

VR exploded in the 1990s with various companies focusing mostly on the professional research market and location-based entertainment. Examples of the more well-known newly formed VR companies were Virtuality, Division, and Fakespace. Existing companies such as Sega, Disney, and General Motors, as well as numerous universities and the military, also started to more extensively experiment with VR technologies. Movies were made, numerous books were written, journals emerged, and conferences formed—all focused exclusively on VR. In 1993, *Wired* magazine predicted that within five years more than one in ten people would wear HMDs while traveling in buses, trains, and planes (Negroponte, 1993). In 1995, the *New York Times*

Meta Platforms, Inc.
Exhibit 1063
Page 053

reported that Virtuality Managing Director Jonathan Waldern predicted the VR market to reach $4 billion by 1998 (Bailey, 1995). It seemed VR was about to change the world and there was nothing that could stop it. Unfortunately, technology could not support the promises of VR. In 1996, the VR industry peaked and then started to slowly contract with most VR companies, including Virtuality, going out of business by 1998.

2.3 The 2000s

The first decade of the 21st century is known as the "VR winter." Although there was little mainstream media attention given to VR from 2000 to 2012, VR research continued in depth at corporate, government, academic, and military research laboratories around the world. The VR community started to turn toward human-centered design with an emphasis on user studies, and it became difficult to get a VR paper accepted at a conference without including some form of formal evaluation. Thousands of VR-related research papers from this era contain a wealth of knowledge that today is unfortunately largely unknown and ignored by those new to VR.

A wide field of view was a major missing component of consumer HMDs in the 1990s, and without it users were just not getting the "magic" feeling of presence (Mark Bolas, personal communication, June 13, 2015). In 2006, Mark Bolas of USC's MxR Lab and Ian McDowall of Fakespace Labs created a 150° field of view HMD called the Wide5, which the lab later used to study the effects of field of view on the user experience and behavior. For example, users can more accurately judge distances when walking to a target when they have a larger field of view (Jones et al., 2012). The team's research led to the low-cost Field of View To Go (FOV2GO), which was shown at the IEEE VR 2012 conference in Orange County, California, where the device won the Best Demo Award and was part of the MxR Lab's Open-source project that is the precursor to most of today's consumer HMDs. Around that time, a member of that lab named Palmer Luckey started sharing his prototype on Meant to be Seen (mtbs3D.com) where he was a forum moderator and where he first met John Carmack (now CTO of Oculus VR) and formed Oculus VR. Shortly after that he left the lab and launched the Oculus Rift Kickstarter. The hacker community and media latched onto VR once again. Companies ranging from start-ups to the Fortune 500 began to see the value of VR and started providing resources for VR development, including Facebook, which acquired Oculus VR in 2014 for $2 billion. The new era of VR was born.

Meta Platforms, Inc.
Exhibit 1063
Page 054

An Overview of Various Realities

This chapter aims to provide a basic high-level overview of various forms of reality, as well as different hardware options to build systems supporting those forms of reality. Whereas most of the book focuses on fully immersive VR, this chapter takes a broader view; its aim is to put fully immersive VR in the context of the larger array of options.

3.1 Forms of Reality

Reality takes many forms and can be considered to range on a virtuality continuum from the real environment to virtual environments (Milgram and Kishino, 1994). Figure 3.1 shows various forms along that continuum. These forms, which are somewhere between virtual and augmented reality, are broadly defined as "mixed reality," which can be further broken down into "augmented reality" and "augmented virtuality." This book focuses on the right side of the continuum from augmented virtuality to virtual environments.

The **real environment** is the real world that we live in. Although creating real-world experiences is not always the goal of VR, it is still important to understand the real world and how we perceive and interact with it in order to replicate relevant functionality into VR experiences. What is relevant depends on the goals of the application. Section 4.4 further discusses trade-offs of realism vs. more abstract implementations of reality. Part II discusses how we perceive real environments in order to help build better fully immersive virtual environments.

Instead of replacing reality, **augmented reality** (AR) adds cues onto the already existing real world, and ideally the human mind would not be able to distinguish between computer-generated stimuli and the real world. This can take various forms, some of which are described in Section 3.2.

Meta Platforms, Inc.
Exhibit 1063
Page 055

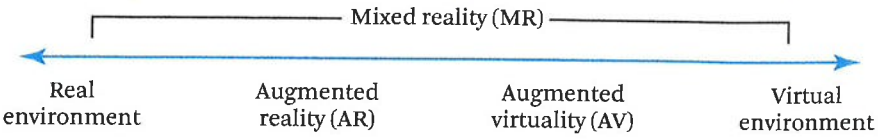


Figure 3.1 **The virtuality continuum.** (Adapted from Milgram and Kishino (1994))

Augmented virtuality (AV) is the result of capturing real-world data and bringing that data into VR. Immersive film is an example of augmented virtuality. In the simplest case, the capture is taken from a single viewpoint, but in other cases, real-world capture can consist of light fields or geometry, where users can freely move about the environment, perceiving it from any perspective. Section 21.6 provides some examples of augmented virtuality.

True **virtual environments** are artificially created without taking any data from the real world. The goal of virtual environments is to completely engage a user in an experience so that she feels as if she is present (Chapter 4) in another world such that the real world is temporarily forgotten, while minimizing any adverse effects (Part III).

3.2 Reality Systems

The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real.

—Sutherland (1965)

A **reality system** is the hardware and operating system that full sensory experiences are built upon. The reality system's job is to effectively communicate the application content to and from the user in an intuitive way as if the user is interacting with the real world. Humans and computers do not speak the same language so the reality system must act as a translator or intermediary between them (note the reality system also includes the computer). It is the VR creator's obligation to integrate content with the system so the intermediary is transparent and to ensure objects and system behaviors are consistent with the intended experience. Ideally, the technology will not be perceived so that users forget about the interface and experience the artificial reality as if it is real.

Communication between the human and system is achieved via hardware devices. These devices serve as input and/or output. A transfer function, as it relates to interaction, is conversion from human output to digital input and from digital output to

Meta Platforms, Inc.
Exhibit 1063
Page 056

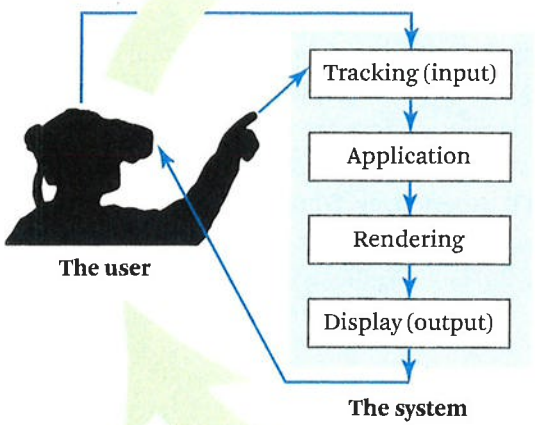


Figure 3.2 **A VR system consists of input from the user, the application, rendering, and output to the user.** (Adapted from Jerald (2009))

human input. What is output and what is input depends on whether it is from the point of view of the system or the human. For consistency, input is considered information traveling from the user into the system and output is feedback that goes from the system back to the user. This forms a cycle of input/output that continuously occurs for as long as the VR experience lasts. This loop can be thought of as occurring between the action and distal stimulus stages of the perceptual process (Figure 7.2) where the user is the perceptual process.

Figure 3.2 shows a user and a VR system divided into their primary components of input, application, rendering, and output. **Input** collects data from the user such as where the user's eyes are located, where the hands are located, button presses, etc. The **application** includes non-rendering aspects of the virtual world including updating dynamic geometry, user interaction, physics simulation, etc. **Rendering** is the transformation of a computer-friendly format to a user-friendly format that gives the illusion of some form of reality and includes visual rendering, auditory

Meta Platforms, Inc.
Exhibit 1063
Page 057

rendering (called auralization), and haptic (the sense of touch) rendering. An example of rendering is drawing a sphere. Rendering is already well defined (e.g., Foley et al., 1995) and other than high-level descriptions and elements that directly affect the user experience the technical details are not the focus of this book. **Output** is the physical representation directly perceived by the user (e.g., a display with pixels or headphones with sound waves).

The primary output devices used for VR are visual displays, speakers, haptics, and motion platforms. More exotic displays include olfactory (smell), wind, heat, and even taste displays. Input devices are only briefly mentioned in this chapter as they are described in detail in Chapter 27. Selecting appropriate hardware is an essential part of designing VR experiences. Some hardware may be more appropriate for some designs than others. For example, large screens are more appropriate than head-mounted displays for large audiences located at the same physical location. The following sections provide an overview of some commonly used VR hardware.

3.2.1 Visual Displays

Today's reality systems are implemented in one of three ways: head-mounted displays, world-fixed displays, and hand-held displays.

Head-Mounted Displays

A **head-mounted display** (HMD) is a visual display that more or less is rigidly attached to the head. Figure 3.3 shows some examples of different HMDs. Position and orientation tracking of HMDs is essential for VR because the display and earphones move with the head. For a virtual object to appear stable in space, the display must be appropriately updated as a function of the current pose of the head; for example, as the user rotates his head to the left, the computer-generated image on the display should move to the right so that the image of the virtual objects appears stable in space, just as real-world objects are stable in space as people turn their heads. Well-implemented HMDs typically provide the greatest amount of immersion. However, doing this well consists of many challenges such as accurate tracking, low latency, and careful calibration.

HMDs can be further broken down into three types: non-see-through HMDs, video-see-through HMDs, and optical-see-through HMDs. Non-see-through HMDs block out all cues from the real world and provide optimal full immersion conditions for VR. Optical-see-through HMDs enable computer-generated cues to be overlaid onto the visual field and provide the ideal augmented reality experience. Conveying the ideal augmented reality experience using optical-see-through head-mounted displays is extremely challenging due to various requirements (extremely low latency, extremely accurate tracking, optics, etc.). Due to these challenges, video-see-through displays are

Meta Platforms, Inc.
Exhibit 1063
Page 058

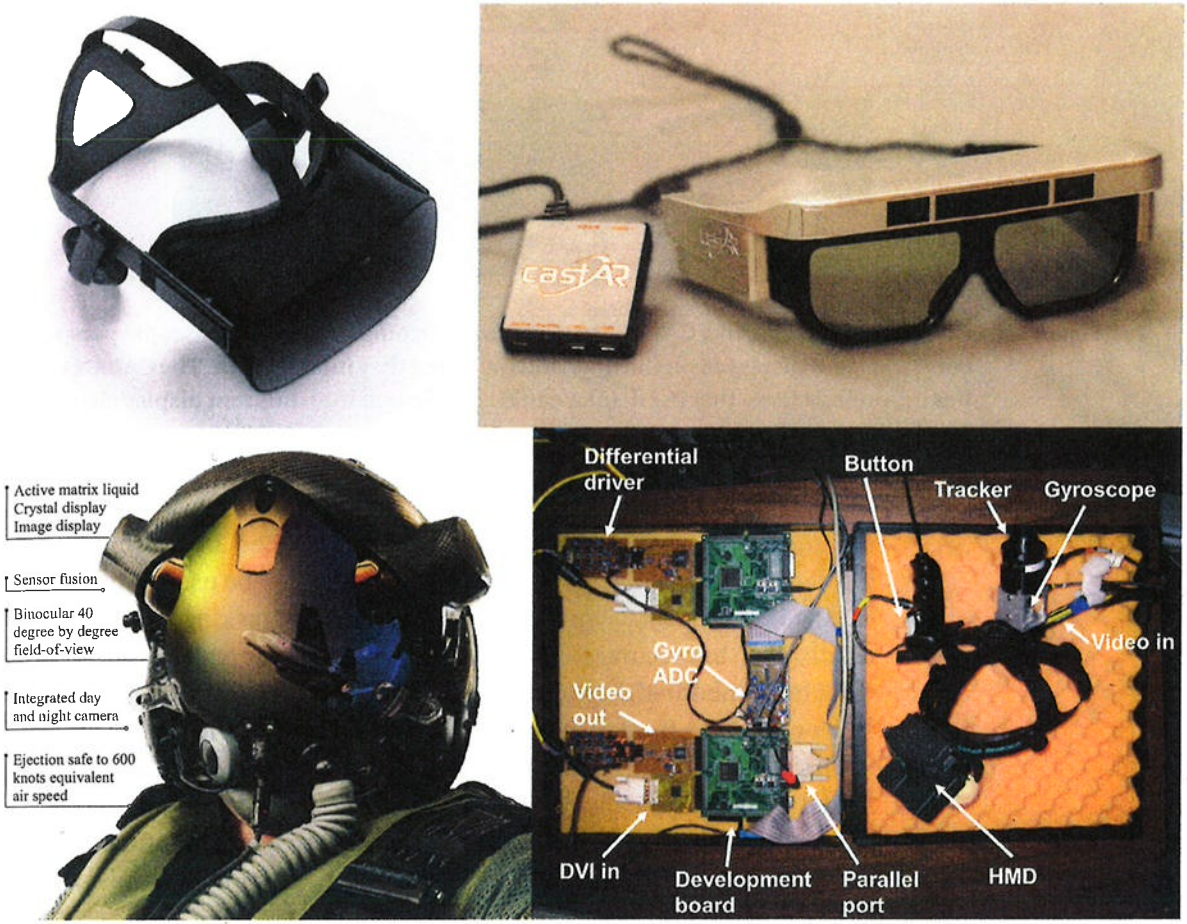


Figure 3.3 **The Oculus Rift (top left; courtesy of Oculus VR), CastAR (upper right; image courtesy of CastAR), the Joint-Force Fighter Helmet (lower left; Courtesy of Marines Magazine), and a custom built/modified HMD (lower right; from Jerald et al. (2007)).**

sometimes used. Video-see-through HMDs are often considered to be augmented virtuality (Section 3.1), and have some advantages and disadvantages of both augmented reality and virtual reality.

World-Fixed Displays

World-fixed displays render graphics onto surfaces and audio through speakers that do not move with the head. Displays take many forms, ranging from a standard monitor (also known as fish-tank VR) to displays that completely surround the user

Meta Platforms, Inc.
Exhibit 1063
Page 059

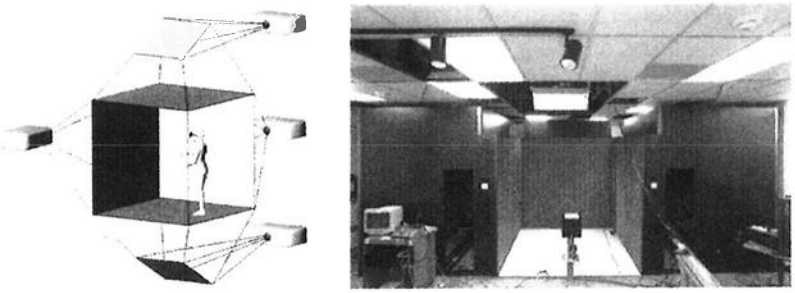


Figure 3.4 **Conceptual drawing of a CAVE (left). Users are surrounded with stereoscopic perspective-correct images displayed on the floor and walls that they interact with. The CABANA (right) has movable walls so that the display can be configured into different display shapes such as a wall or L-shape.** (From Cruz et al. (1992) and Daily et al. (1999))

(e.g., CAVEs and CAVE-like displays as shown in Figures 3.4 and 3.5). Display surfaces are typically flat surfaces, although more complex shapes can be used if those shapes are well defined or known, as shown in Figure 3.6. Head tracking is important for world-fixed displays, but accuracy and latency requirements are typically not as critical as they are for head-mounted displays because stimuli are not as dependent upon head motion. High-end world-fixed displays with multiple surfaces and projectors can be highly immersive but are more expensive in dollars and space.

World-fixed displays typically are considered to be part virtual reality and part augmented reality. This is because real-world objects are easily integrated into the experience, such as the physical chair shown in Figure 3.7. However, it is often the intent that the user's body is the only visible real-world cue.

Hand-Held Displays

Hand-held displays are tracked devices that can be held with the hand(s) and do not require precise alignment with the eyes (in fact the head is rarely tracked for hand-held displays). Hand-held augmented reality, also called indirect augmented reality, has recently become popular due to the ease of access and improvements in smartphones/tablets (Figure 3.8). In addition, system requirements are much less since viewing is indirect—rendering is independent of the user's head and eyes.

3.2.2 Audio

Spatialized audio provides a sense of where sounds are coming from in 3D space. Speakers can be located either in the world or over the ears. Headphones are preferred for a fully immersive system as they block out more of the real world. How the ears and

Meta Platforms, Inc.
Exhibit 1063
Page 060

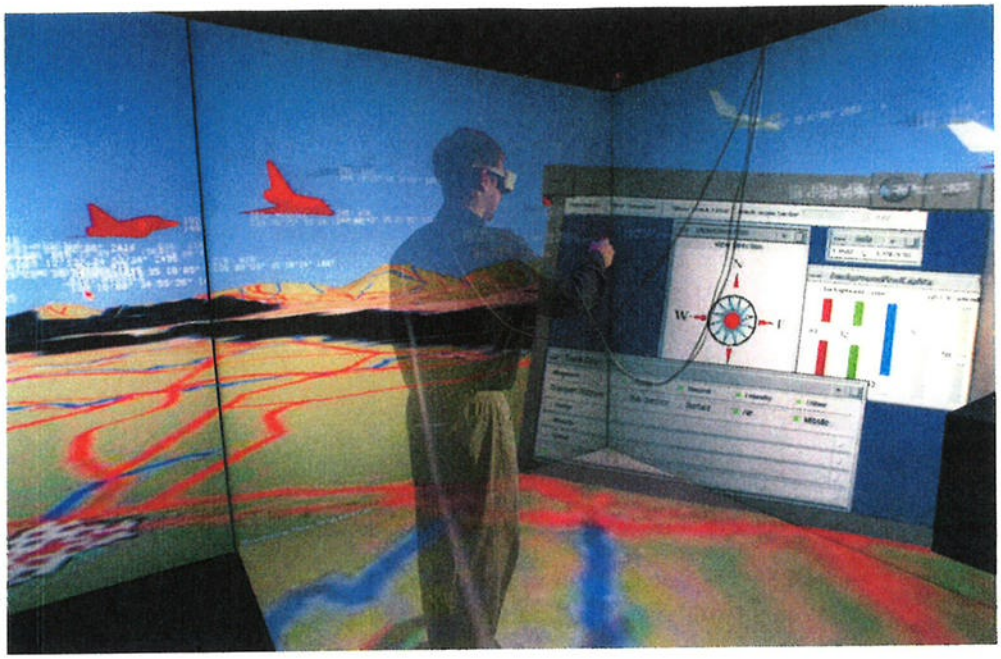


Figure 3.5 **The author interacting with desktop applications within the CABANA.** (From Jerald et al. (2001))

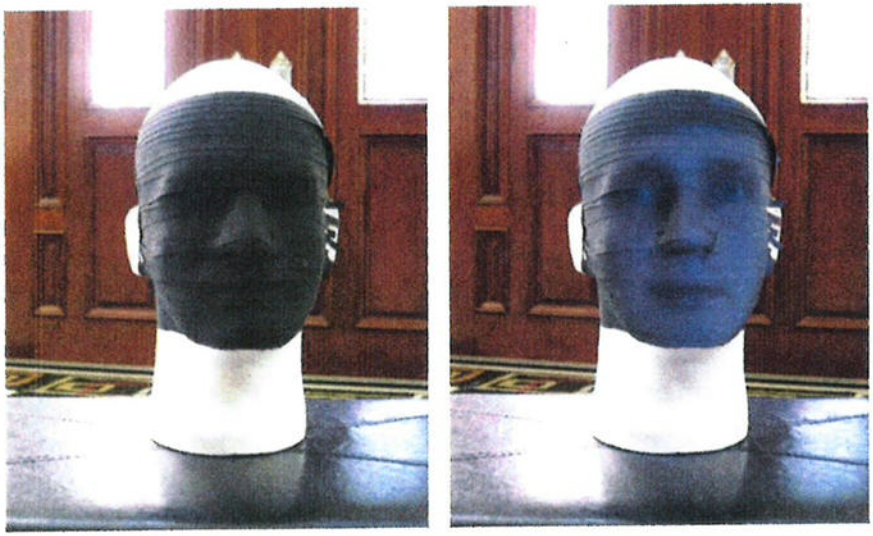


Figure 3.6 **Display surfaces do not necessarily need to be planar.** (From Krum et al. (2012))

Meta Platforms, Inc.
Exhibit 1063
Page 061



Figure 3.7 **The University of Southern California's Gunslinger uses a mix of the real world along with world-fixed displays.** (Courtesy of USC Institute for Creative Technologies)



Figure 3.8 **Zoo-AR from GeoMedia and a virtual assistant that appears on a business card from NextGen Interactions.**

brain perceive sound is discussed in Section 8.2. Audio more specific to how content is created for VR is discussed in Section 21.3.

3.2.3 Haptics

Haptics simulate physical forces between virtual objects and the user's body. Haptics can be classified as passive (static physical objects) or active (physical feedback

Meta Platforms, Inc.
Exhibit 1063
Page 062

controlled by the computer), tactile (through skin) or proprioceptive force (through joints/muscles), and self-grounded (worn) or world-grounded (attached to real world). Many haptic systems also serve as input devices.

Passive vs. Active Haptics

Passive haptics provide a sense of touch in VR at a low cost—one simply creates a real-world physical object and matches that object to the shape of a virtual object (Lindeman et al., 1999). These physical objects can be hand-held props or larger objects in the world that can be touched. Passive haptics increases presence, improves cognitive mapping of the environment, and improves training performance (Insko and Insko, 2001).

Touching a few objects with passive haptics can make everything else seem more real. Perhaps the most compelling VR experience to this day is the legendary UNC-Chapel Hill Pit demo (Meehan et al., 2002). Users first experience a virtual room that includes passive haptics made from Styrofoam blocks and other real-world material to match the visual VR environment. After touching different parts of the room, users walk into a second room and see a pit in the floor. The pit is quite compelling (in fact, heart rate increases) because everything else they have touched up to this point has physically felt real, thus users assume the pit is physically real as well. There is an even more startling response from many users when they put their toe over the virtual ledge and feel a real ledge. What they don't realize is the physical ledge is only a 1.5 inch drop-off compared to the visual pit that is 20 feet deep.

Active haptics are controlled by a computer and are the most common form of haptics. Active haptics have the advantage that forces can be dynamically controlled to provide a feeling of a wide range of simulated virtual objects. The remainder of this section focuses on active haptics.

Tactile vs. Proprioceptive Force Haptics

Tactile haptics provide a sense of touch through the skin. Vibrotactile stimulation evokes tactile sensations using mechanical vibration of the skin. Electrotactile stimulation evokes tactile sensation via an electrode passing current through the skin.

Figure 3.9 shows Tactical Haptics' Reactive Grip technology, which provides a sense of tactile feedback that is surprisingly compelling, especially when combined with fully immersive visual displays (Provancher, 2014). The system utilizes sliding skin-contact plates that can be added to any hand-held controller. Translational motions and forces are portrayed along the length of the grip by moving the plates in unison. Opposing motion and forces from different plates creates the feeling of a virtual object wrenching within the user's grasp.

Meta Platforms, Inc.
Exhibit 1063
Page 063

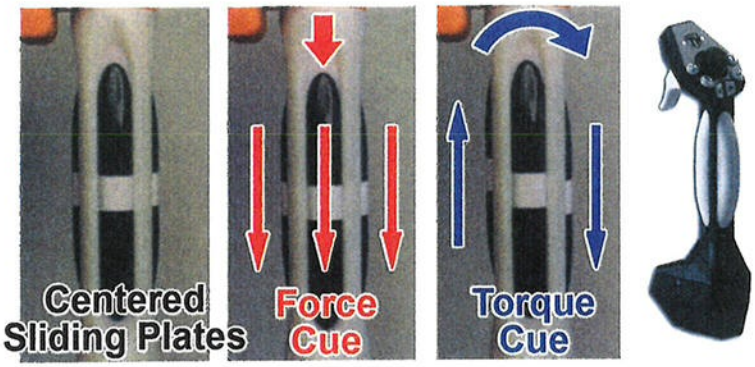


Figure 3.9 **Tactical Haptics technology uses sliding plates to provide a sense of up or down force as well as rotational torque. The rightmost image shows the sliding plates integrated into the latest controller design.** (Images courtesy of Tactical Haptics)

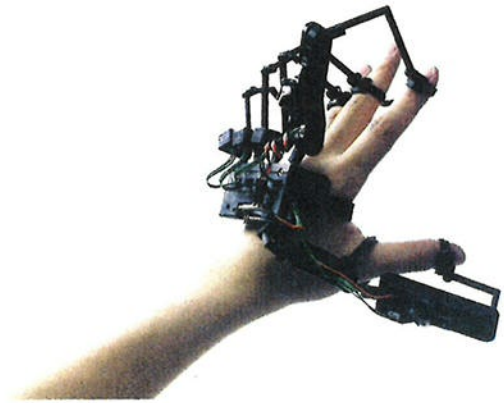


Figure 3.10 **The Dexta Robotics Dexmo F2 device provides both finger tracking and force feedback.** (Image courtesy of Dexta Robotics)

Proprioceptive force provides a sense of limb movement and muscular resistance. Proprioceptive haptics can be self-grounded or world-grounded.

Self-Grounded vs. World-Grounded Haptics

Self-grounded haptics are worn/held by and move with the user. The forces applied are relative to the user. Gloves with exoskeletons or buzzers are examples of self-grounded haptics. Figure 3.10 shows an exoskeleton glove. Hand-held controllers are also examples of self-grounded haptics. Such controllers might be simply a passive

Meta Platforms, Inc.
Exhibit 1063
Page 064



Figure 3.11 **Sensable's Phantom haptics system.** (Courtesy of INITION)

prop that acts as a handle to virtual objects or might be rumble controllers that vibrate to provide feedback to the user (e.g., to signify the user has put his hand through a virtual object).

World-grounded haptics are physically attached to the real world and can provide a true sense of fully solid objects that don't move because the position of the virtual object providing the force can remain stable relative to the world. The ease of which the object can be moved can also be felt, providing a sense of weight and friction (Craig et al., 2009).

Figure 3.11 shows Sensable's Phantom haptic device, which provides stable force feedback for a single point in space (the tip of the stylus). Figure 3.12 shows Cyberglove's CyberForce glove that provides the sense of touching real objects with the entire hand as if the objects were stationary in the world.

3.2.4 Motion Platforms

A **motion platform** is a hardware device that moves the entire body resulting in a sense of physical motion and gravity. Such motions can help to convey a sense of orientation, vibration, acceleration, and jerking. Common uses of platforms are for racing games, flight simulation, and location-based entertainment. When integrated well with the

Meta Platforms, Inc.
Exhibit 1063
Page 065



Figure 3.12 **Cyberglove's Cyberforce Immersive Workstation.** (Courtesy of Haptic Workstation with HMD at VRLab in EPFL, Losanne, 2005)

rest of a VR application, motion sickness can be reduced by decreasing the conflict between visual motion and felt motion. Section 18.8 discusses how motion platforms can be used to reduce motion sickness.

Motion platforms can be active or passive. An **active motion platform** is controlled by the computer simulation. Figure 3.13 shows an example of an active motion platform that moves a base platform via hydraulic actuators. A **passive motion platform** is controlled by the user. For example, the tilting of a passive motion platform might be achieved by leaning forward, such as used with Birdly shown in Figure 3.14. Note active and passive as described here are from the point of view of the motion platform and system. When describing motion from the point of view of the user, passive implies the user is passively along for the ride, with no way to influence the experience, and active implies the user is actively influencing the experience.

3.2.5 Treadmills

Treadmills provide a sense that one is walking or running while actually staying in one place. Variable-incline treadmills, individual foot platforms, and mechanical tethers

Meta Platforms, Inc.
Exhibit 1063
Page 066

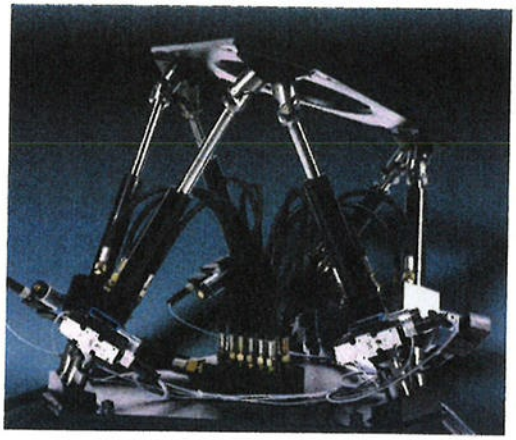


Figure 3.13 **An active motion platform that moves via hydraulic actuators. A chair can be attached to the top of the platform.** (Courtesy of Shifz, Syntharturalist Art Association)



Figure 3.14 **Birdly by Somniacs. In addition to providing visual, auditory, and motion cues, this VR experience provides a sense of taste and smell.** (Courtesy of Swissnex San Francisco and Myleen Hollero)

Meta Platforms, Inc.
Exhibit 1063
Page 067



Figure 3.15 **The Virtuix Omni.** (Image courtesy of Virtuix)

providing restraint can convey hills by manipulating the physical effort required to travel forward.

Omnidirectional treadmills enable simulation of physical travel in any direction and can be active or passive.

Active omnidirectional treadmills have computer-controlled mechanically moving parts. These treadmills move the treadmill surface in order to recenter the user on the treadmill (e.g., Darken et al., 1997, and Iwata, 1999). Unfortunately, such recentering can cause the user to lose balance.

Passive omnidirectional treadmills contain no computer-controlled mechanically moving parts. For example, the feet might slide along a low-friction surface (e.g., the Virtuix Omni as shown in Figure 3.15). A harness and surrounding encasing keeps the user from falling. Like other forms of non-real walking, the experience of walking on a passive treadmill does not perfectly match the real world (it feels more like walking on slippery ice), but can add a significant amount of presence and reduce motion sickness.

3.2.6 Other Sensory Output

VR largely focuses on the display component, but other components such as taste, smell, and wind can more fully immerse users and add to a VR experience. Figure 3.14

Meta Platforms, Inc.
Exhibit 1063
Page 068

shows Birdly by Somniacs, a system that adds smells and wind to a VR experience. Section 8.6 discusses the senses of taste and smell.

3.2.7 Input

A fully immersive VR experience is more than simply presenting content. The more a user physically interacts with a virtual world using his own body in intuitive ways, the more that user feels engaged and present in that virtual world. VR interactions consist of both hardware and software working closely together in complex ways, yet the best interaction techniques are simple and intuitive to use. Designers must take into account input device capabilities when designing experiences—one input device might work well for one type of interaction but be inappropriate for another. Other interactions might work across a wider range of input devices. Part V contains multiple chapters on interaction with Chapter 27 focusing exclusively on input devices.

3.2.8 Content

VR cannot exist without content. The more compelling the content, the more interesting and engaging the experience. Content includes not only the individual pieces of media and their perceptual cues, but also the conceptual arc of the story, the design/layout of the environment, and computer- or user-controlled characters. Part IV contains multiple chapters dedicated to content creation.

Meta Platforms, Inc.
Exhibit 1063
Page 069

Immersion, Presence, and Reality Trade-Offs

VR is about psychologically being in a place different than where one is physically located, where that place may be a replica of the real world or may be an imaginary world that does not exist and never could exist. Either way there are some commonalities and essential concepts that must be understood to have users feel like they are somewhere else. This chapter discusses immersion, presence, and trade-offs of reality.

4.1 Immersion

Immersion is the objective degree to which a VR system and application projects stimuli onto the sensory receptors of users in a way that is extensive, matching, surrounding, vivid, interactive, and plot informing (Slater and Wilbur, 1997).

Extensiveness is the range of sensory modalities presented to the user (e.g., visuals, audio, and physical force).

Matching is the congruence between sensory modalities (e.g., appropriate visual presentation corresponding to head motion and a virtual representation of one's own body).

Surroundness is the extent to which cues are panoramic (e.g., wide field of view, spatialized audio, 360° tracking).

Vividness is the quality of energy simulated (e.g., resolution, lighting, frame rate, audio bitrate).

Interactability is the capability for the user to make changes to the world, the response of virtual entities to the user's actions, and the user's ability to influence future events.

Plot is the story—the consistent portrayal of a message or experience, the dynamic unfolding sequence of events, and the behavior of the world and its entities.

Meta Platforms, Inc.
Exhibit 1063
Page 070

Immersion is the objective technology that has the potential to engage users in the experience. However, immersion is only part of the VR experience as it takes a human to perceive and interpret the presented stimuli. Immersion can lead the mind but cannot control the mind. How the user subjectively experiences the immersion is known as presence.

4.2 Presence

Presence, in short, is a sense of "being there," inside a space, even when physically located in a different location. Because presence is an internal psychological state and a form of visceral communication (Section 1.2.1), it is difficult to describe in words—it is something that can only be understood when experienced. Attempting to describe presence is like attempting to describe concepts such as consciousness or the feeling of love, and can be just as controversial. Nonetheless, the VR community craves a definition of presence, as such a definition would be useful for designing VR experiences. The definition of presence is based on a discussion that took place via the presence-l listserv during the spring of 2000 among members of a community of scholars interested in the presence concept. The lengthy explication, available on the ISPR website (http://ispr.info), begins with

> Presence is a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience. (International Society for Presence Research, 2000)

Whereas immersion is about the characteristics of technology, presence is an internal psychological and physiological state of the user; an awareness in the moment of being immersed in a virtual world while having a temporary amnesia or agnosia of the real world and the technical medium of the experience. When present, the user does not attend to and perceive the technology, but instead attends to and perceives the objects, events, and characters the technology represents. Users who feel highly present consider the experience specified by VR technology to be a place visited rather than simply something perceived.

Presence is a function of both the user and immersion. Immersion is capable of producing the sense of presence but immersion does not always induce presence—users can simply shut their eyes and imagine being somewhere else. Presence is, however, limited by immersion; the greater immersion a system/application provides then the greater potential for a user to feel present in that virtual world.

Meta Platforms, Inc.
Exhibit 1063
Page 071

A **break-in-presence** is a moment when the illusion generated by a virtual environment breaks down and the user finds himself where he truly is—in a laboratory wearing an HMD (Slater and Steed, 2000). These breaks-in-presence can destroy a VR experience and should be avoided as much as possible. Example causes of a break-in-presence include loss of tracking, a person speaking from the real world that is not part of the virtual environment, tripping on a wire, a real-world phone ringing, etc.

4.3 Illusions of Presence

Presence induced by technology-driven immersion can be considered a form of illusion (Section 6.2) since in VR stimuli are just a form of energy projected onto our receptors—pixels on a screen or audio recorded from a different time and place. Different researchers distinguish different forms of presence in different ways. Presence is divided below into four core components that are merely illusions of a non-existent reality.

The Illusion of Being in a Stable Spatial Place

Feeling as if one is in a physical environment is the most important part of presence. This is a subset of what Slater calls "place illusion" (Slater, 2009), and it occurs due to all of a user's sensory modalities being congruent such that stimuli presented to that user (ideally with no encumbrances such as no restrictions on field of view, no cables pulling on the head, and the freedom to move around) behave as if those stimuli originated from real-world objects in 3D space. Depth cues (Section 9.1.3) are especially important for providing a sense of being at a remote location, and the more depth cues that are consistent with each other the better. This illusion can be broken when the world does not feel stable due to long latency, low frame rate, miscalibration, etc.

The Illusion of Self-Embodiment

We have a lifetime of experience perceiving our own bodies when we look down. Yet many VR experiences contain no personal body—the user is a disembodied viewpoint in space. Self-embodiment is the perception that the user has a body within the virtual world. When disembodied users feel quite present and then are given a virtual body that properly matches their movements, they quickly realize that there are different levels of presence. Then if a user sees a visual object touching the skin while a physical object also touches the skin, presence is greatly strengthened and experienced more deeply (called the rubber hand illusion; Botvinick and Cohen, 1998).

Meta Platforms, Inc.
Exhibit 1063
Page 072

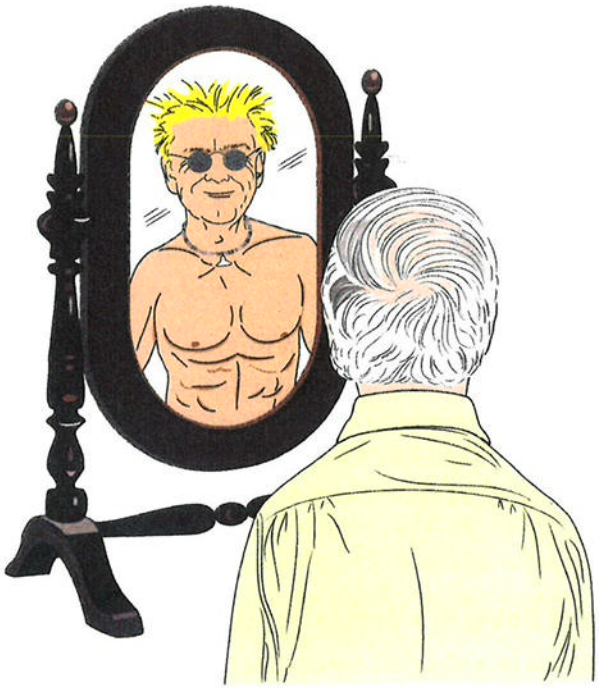


Figure 4.1 **How we perceive ourselves can be quite distorted.** (Courtesy of Gail Drinnan/Shutterstock.com)

Surprisingly, presence of a virtual body can be quite compelling even when that body does not look like one's own body. In fact, we don't necessarily perceive ourselves objectively and how we perceive ourselves, even in real life, through the lens of our subjectivity can be quite distorted (Figure 4.1; Maltz, 1960). The mind also automatically associates the visual characteristics of what is seen at the location of the body with one's own body. In VR, one can perceive oneself as a cartoon character or someone of a different gender or race, and the experience can still be quite compelling. Research has shown this to be quite effective for teaching empathy by "walking in someone else's shoes" and can reduce racial bias (Peck et al., 2013). Perhaps this is possible due to our being accustomed to changing clothing on a regular basis—just because our clothes are a different color/texture than our skin or what we were recently wearing does not mean what is underneath those clothes is not our own body. Whereas body shape and color are not so important, motion is extremely important and presence can be broken when visual body motion does not match physical motion in a reasonable manner.

Meta Platforms, Inc.
Exhibit 1063
Page 073

The Illusion of Physical Interaction

Looking around is not enough for people to believe for more than a few seconds that they are in an alternate world. Even if not realistic, adding some form of feedback such as audio, visual highlighting, or a rumble of a controller can give the user a sense that she has in some way touched the world. Ideally, the user should feel a solid physical response that matches the visual representation (as described in Section 3.2.3). As soon as one reaches out to touch something and there is no response, then a break-in-presence can occur. Unfortunately, strong physical feedback can be difficult to attain so sensory substitution (Section 26.8) is often used instead.

The Illusion of Social Communication

Social presence is the perception that one is really communicating (both verbally and through body language) with other characters (whether computer- or user-controlled) in the same environment. Social realism does not require physical realism. Users have been found to exhibit anxiety responses when causing pain to a relatively low-fidelity virtual character (Slater, Antley, et al., 2006) and when users with a fear of public speaking must talk in front of a low-fidelity virtual audience (Slater, Pertaub, et al., 2006).

Although social presence does increase as behavioral realism (the degree to which human representations and objects behave as they would in the physical world) increases (Guadagno et al., 2007), even tracking and rendering only a few points on human players can be quite compelling (Section 9.3.9). Figure 4.2 shows a screenshot of the multiplayer VR game Arena by NextGen Interactions where the player's head and both hands are directly controlled (i.e., three tracked points) and lower-body turning/walking/running animations are indirectly controlled with analog sticks on Sixense Razer Hydra Controllers.

4.4 Reality Trade-Offs

Some consider real reality the gold standard of what we are trying to achieve with VR whereas others consider reality a goal to surpass—for if we can only get to the point of matching reality, then what is the point? This section discusses the trade-offs of trying to replicate reality vs. creating more abstract experiences.

4.4.1 The Uncanny Valley

Some people say robots and computer-generated characters come across as creepy. Although our sense of familiarity with simulated characters representing real characters increases as we get closer to reality, this is only up to a point. If reality is approached,

Meta Platforms, Inc.
Exhibit 1063
Page 074

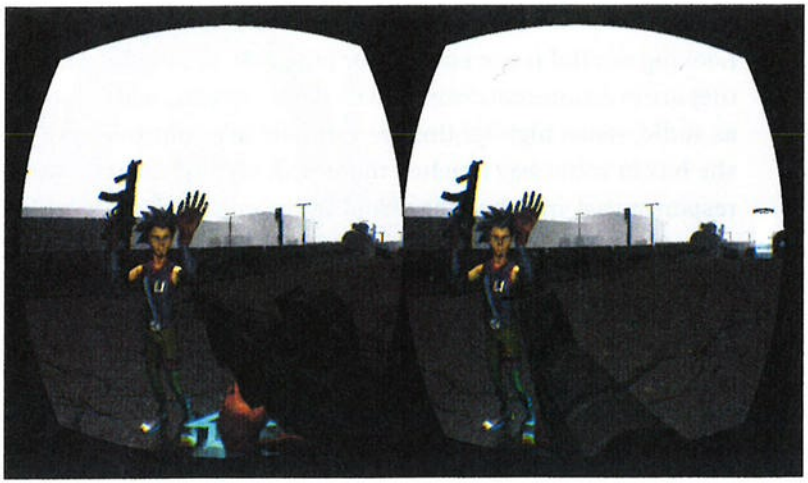


Figure 4.2 **Paul Mlyniec, President of Digital ArtForms, surrenders without having to think about the interface as he is threatened by the author. Head and hand tracking alone can be extremely socially compelling in VR due to the ability to naturally convey body language.** (Image courtesy of NextGen Interactions)

but not attained, some of our reactions shift from empathy to revulsion. This descent into creepiness is known as the Uncanny Valley as first proposed by Masahiro Mori (1970). Figure 4.3 shows a graph of observers' comfort level with virtual humans as a function of human realism. Observer comfort increases as a character becomes more human-like up until a certain point at which observers start to feel uncomfortable with the almost, but not quite, human character.

The Uncanny Valley is a controversial topic that is more a simple explanatory theory rather than backed by scientific evidence. However, there is power in simplicity and the theory helps us to think about how to design and give character to VR entities. Creating cartoon characters often yields superior results over creating near-photorealistic humans. Section 22.5 discusses such issues in further detail.

4.4.2 Fidelity Continua

The Uncanny Valley of character creepiness is not the only case of things being close to reality not necessarily being better. The goal of VR is not necessarily to replicate reality. Contrary to what one might initially think, presence does not require photorealism and there are more important presence-inducing cues such as responsiveness of the system, character motion, and depth cues. Simple worlds consisting of basic structures that provide a sense of spatial stability can be extremely compelling, and

Meta Platforms, Inc.
Exhibit 1063
Page 075

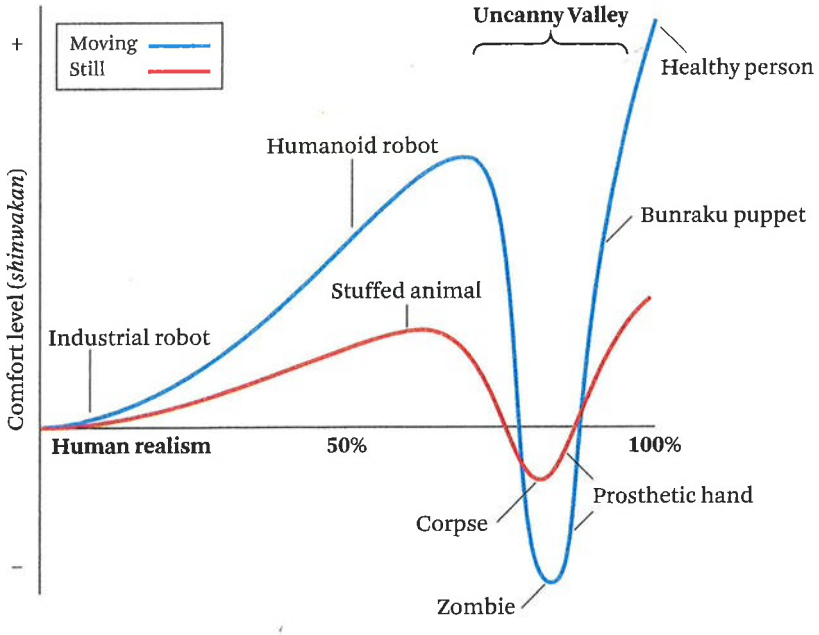


Figure 4.3 **The Uncanny Valley.** (Based on Ho and MacDorman (2010))

making worlds more photorealistic does not necessarily increase presence (Zimmons and Panter, 2003). Being in a cartoon world can feel as real as a world captured by 3D scanners.

Highly presence-inducing experiences can be ranked on different continua, and one extreme on each continuum is not necessarily better than the other extreme. What points on the continua VR creators choose depends upon the vision and goals of the project.

Listed below are some VR fidelity continua that VR creators should consider.

Representational fidelity is the degree to which the VR experience conveys a place that is, or could be, on Earth. At the high end of this spectrum is photorealistic immersive film in which the real world is captured with depth cameras and microphones, then re-created in VR. At the extreme low end of the spectrum are purely abstract or non-objective worlds (e.g., blobs of color and strange sounds). These may have no reference to the real world, simply conveying emotions, exploring pure visual events, or presenting other non-narrative qualities (Paul Mlyniec, personal communication, April 28, 2015). Cartoon worlds and abstract

Meta Platforms, Inc.
Exhibit 1063
Page 076

video games are somewhere in the middle depending on how closely the scene and characters represent the real world and its inhabitants.

Interaction fidelity is the degree to which physical actions for a virtual task correspond to physical actions for the equivalent real-world task (Section 26.1). On one end of this spectrum are physical training tasks where low interaction fidelity risks negative training effects (Section 15.1.4). At the other extreme are interaction techniques that require no physical motion beyond a button press. Magical techniques are somewhere in the middle where users are able to do things they are not able to do in the real world such as grabbing objects at a distance.

Experiential fidelity is the degree to which the user's personal experience matches the intended experience of the VR creator (Section 20.1). A VR application that closely conveys what the creator intended has high experiential fidelity. A free-roaming world where endless possibilities exist and every usage results in a different experience has low experiential fidelity.

Meta Platforms, Inc.
Exhibit 1063
Page 077

The Basics: Design Guidelines

No other technology can cause fear, running, or fleeing in the way that VR can. VR can also induce frustration and anger when things don't work, and worse—even physical sickness if not properly implemented and calibrated. Well-designed VR can produce the awe and excitement of being in a different world, improve performance and cut costs, provide new worlds to experience, improve education, and create better understanding by walking in someone else's shoes. With the foundation set in this chapter, the reader can better understand subsequent parts of this book and start creating VR experiences. As VR creators, we have the chance to change the world—let's not blow it through bad design and implementation due to not understanding the basics.

High-level guidelines for creating VR experiences are provided below.

Introduction and Background (Part I)

- In order to have a toolbox of options to intelligently choose from, learn the basics of VR, such as the different forms of augmented and virtual reality, different hardware options, ways to present information to the senses, and how to induce presence into the minds of users.

VR Is Communication (Section 1.2)

- Focus on the user experience rather than the technology (Section 1.2).
- Simplify and harmonize the communication between user and technology (Section 1.2).
- Focus on making the technical intermediary between the user and content transparent so users feel like they have direct access to the virtual world and its entities (Section 1.2.1).

Meta Platforms, Inc.
Exhibit 1063
Page 078

- Design for visceral communication (Section 1.2.1) in order to induce presence (Section 4.2) and inspire awe in users.

An Overview of Various Realities (Chapter 3)

- Choose what form of reality you wish to create (Section 3.1). Where does it fall on the virtuality continuum?
- Choose what type of input and output hardware to use (Section 3.2).
- Understand a VR application is more than just the hardware and technology. Create a strong conceptual story, an interesting design or layout of the environment, and engaging characters (Section 3.2.8).

Immersion, Presence, and Reality Trade-Offs (Chapter 4)

- Since presence is a function of both immersion and the user, we can't entirely control it. Maximize presence by focusing on what we do have control over—immersion (Section 4.1).
- Minimize breaks-in-presence (Section 4.2).
- For maximum presence, focus first on world stability and depth cues. Then consider adding physical user interactions, cues of one's own body, and social communication (Section 4.3).
- Avoid the Uncanny Valley by not trying to make characters appear too close to the way real humans look (Section 4.4.1).
- Choose what type of representational fidelity, interaction fidelity, and experiential fidelity you want to create (Section 4.4.2).

Meta Platforms, Inc.
Exhibit 1063
Page 079

PART

PERCEPTION

We see things not as they are, but as we are—that is, we see the world not as it is, but as molded by the individual peculiarities of our minds.

—G. T. W. Patrick (1890)

Of all the technologies required for a VR system to be successful, there is one component that is most essential. Fortunately, it is widely available, although most of the population is only vaguely aware of it, fewer have even a moderate understanding of how it operates, and no single individual completely understands it. Like most technologies, it originally started in a very simple form, but has since evolved into a much more sophisticated and intelligent platform. In fact, it is the single most complicated system in existence today, and will remain so for the foreseeable future. It is the human brain.

The human brain contains approximately 100 billion neurons and, on average, each of these neurons is connected by thousands of synapses—structures that allow neurons to pass an electrochemical signal to another cell—resulting in hundreds of trillions of synaptic connections that can simultaneously exchange and process prodigious amounts of information over a massively parallel neural network in milliseconds (Marois and Ivanoff, 2005). In fact, every cubic millimeter of cerebral cortex contains roughly a billion synapses (Alonso-Nanclares et al., 2008). The white matter of the nervous system contains approximately 150,000–180,000 km of myelinated nerve fibers at age 20 (i.e., the connecting wires in our white matter could circle the earth ~4 times), connecting all these neuronal elements. Despite the monumental number of components in the brain, it is estimated that each neuron is able to contact any other

Meta Platforms, Inc.
Exhibit 1063
Page 080

neuron with no more than six interneuronal connections or "six degrees of separation" (Drachman, 2005).

Many of us tend to believe that we understand human behavior and the human mind—after all we are human and should therefore understand ourselves. To some extent this is true—knowing the basics is enough to function in the real world or to experience VR. In fact, this part can be skipped if the intent is to simply re-create what others have already built or to even create new simple worlds (although a theoretical background does help to fine-tune one's creations). However, most of what we perceive and do is a result of subconscious processes. To go beyond the basics of VR to create the most innovative experiences in a way that is comfortable and intuitive to work with, a more detailed understanding of perception is helpful.

What is meant by "beyond the basics of VR"? VR is a relatively new medium and the vast space of artificial realities that could be created is largely unexplored. Thus, we get to make it all up. In fact, many believe providing direct access to digital media through VR has no limits. Whereas that may be true, that does not mean every possible arrangement of input and output will result in a quality VR experience. In fact, only a fraction of the possible input/output combinations will result in a comprehensible and satisfying experience. Understanding how we perceive reality is essential to designing truly innovative VR applications that effectively present and react to users. Some concepts discussed in this part directly apply to VR today and other concepts may apply in ways we can't yet fathom. Understanding the core concepts of human perception might just enable us to recognize new opportunities as VR improves and the vast array of new VR experiences is explored.

Of course studying perception is valuable in designing any product—e.g., desktop or hand-held applications. For VR, it is essential. Rarely has someone become sick due to a bad traditional software design (although plenty of us have become frustrated!), but sickness is common for a bad VR design. VR designers integrate more senses into a single experience, and if sensations are not consistent, then users can become physically ill (multiple chapters are devoted to this—see Part III). Thus, understanding the human half of the hybrid interaction design is more important for VR applications than it is for traditional applications. Once we have obtained this knowledge, then we can more intelligently experiment with the rules. Simply put, the better our understanding of human perception, the better we can create and iterate upon quality VR experiences.

The material in this part will help readers become more aware and appreciative of the real world, and how we humans perceive it. It addresses less common questions such as the following.

- Why do the images on my eyeballs not seem to move when I rotate my eyes? See Section 7.4.

Meta Platforms, Inc.
Exhibit 1063
Page 081

- What is the resolution of the human eye? See Section 8.1.4.
- Why does a character not seem to change sizes when that character walks away from the camera in a movie or game even though his image takes up a smaller portion of the screen? See Section 10.1.

By asking and understanding the answers to these types of questions (albeit we are far from knowing all the answers!), we can create better VR experiences in more innovative ways.

To appreciate perception, consider what you are experiencing at this very moment. You are reading the pages of this book without consciously thinking about how these abstract symbols called letters and words convey meaning in your mind. The same holds true for pictures. **Trompe-l'œil** (French for "deceive the eye") is an art technique that uses realistic 2D imagery to create the illusion of 3D. Looking at the trompe-l'œil lobby art in Figure II.1, you may perceive a lobby floor that opens up to lower levels of the building. In reality, you are seeing multiple 2D planes giving the illusion of depth (the real 3D world containing this book or display showing a 2D photograph of a 3D scene containing a planar lobby surface with 2D art). Unfortunately, trompe-l'œil art only provides an illusion of 3D from a specific static viewing location. Move a few centimeters and the illusion breaks down. VR uses head tracking so illusions work from any viewpoint, in some cases even as one physically walks around.

Subconscious perception goes beyond vision. For example, perhaps you are touching this book or a table and do not consciously consider that specific feeling of touch until this prose suggested it. Conversely, even though you may not have been aware of that touch before the previous statement, you probably would have noticed if you did not feel anything. This lack of touch in most VR applications is a major difference between the real world and virtual worlds, and is one of the significant challenges of VR (see Section 26.8). Nevertheless, innovative, well-funded companies are taking on the challenge to circumvent or otherwise solve the current lack of perfect feedback mechanisms for VR. Whereas there are undoubtedly challenges, the arena is rapidly evolving, and such problems represent opportunities rather than barriers for the proliferation of VR.

Part II consists of several chapters that focus on different aspects of human perception.

Chapter 6, Objective and Subjective Reality, discusses the differences between what actually exists in the world and how we perceive what exists. The two can often be very different as demonstrated by various perceptual illusions described in the chapter.

Meta Platforms, Inc.
Exhibit 1063
Page 082

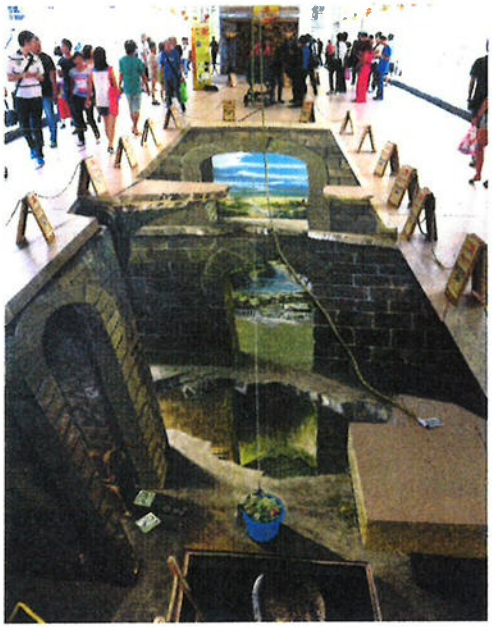


Figure II.1 **Trompé-l'œil lobby art.** (From http://kurtwenner.com/galleries/pavement/pavement_4/pages/StreetPaintingGallery4.003.htm; accessed June 1, 2015. Courtesy of Sleeping cat/Shutterstock.com)

Chapter 7, Perceptual Models and Processes, discusses different models and ways of thinking about how the human mind works. Because the brain is so complex, no single model explains everything and it is useful to think about how physiological processes and perception work from different perspectives.

Chapter 8, Perceptual Modalities, discusses our sensory receptors and how signals from those receptors are transformed into sight, hearing, touch, proprioception, balance and physical motion, smell, and taste.

Chapter 9, Perception of Space and Time, discusses how we perceive the layout of the world around us, how we perceive time, and how we perceive the motion of both objects and ourselves.

Chapter 10, Perceptual Stability, Attention, and Action, discusses how we perceive the world to be consistent even when signals reaching our sensory receptors change, how we focus on items of interest while excluding others, and how perception relates to taking action.

Chapter 11, Perception: Practitioner's Summary and Design Guidelines, summarizes Part II and provides perception guidelines for VR creators.

Meta Platforms, Inc.
Exhibit 1063
Page 083

Objective and Subjective Reality

What's reality anyway? Just a collective hunch!
—Lily Tomlin (Hamit, 1993)

Objective reality is the world as it exists independent of any conscious entity observing it. Unfortunately, it is impossible for someone to perceive objective reality exactly as it is. This chapter describes how we perceive objective reality through our own subjectivity. This is done through a philosophical discussion as well as with examples of perceptual illusion.

6.1 Reality Is Subjective

The things we sense with our eyes, ears, and bodies are not simply a reflection of the world around us. We create in our minds the realities that we think we live in. **Subjective reality** is the way an individual perceives and experiences the external world in his own mind. Much of what we perceive is the result of our own made-up fiction of what happened in the past, and now believe to be the "truth." We continually make up reality as we go along even though most of us do not realize it.

Abstract philosophy with no application to VR? That is certainly a valid opinion. Take a look at Figure 6.1. What is the objective reality? An old lady or a young lady? Look again, and try to see if you can perceive both. This is a classic textbook example of an ambiguous image. However, there is neither an old lady or a young lady. What actually exists in objective reality is a black-and-white image made of ink on paper or pixels on a digital display. This is essentially what we are doing with VR—creating artificial content and presenting that to users so that they perceive our content as being real.

Figure 6.2 demonstrates how reality is what we make of it. The image is a 100-plus-year-old face with a resolution of 14 × 18 grayscale pixels (Harmon, 1973). If you live

Meta Platforms, Inc.
Exhibit 1063
Page 084



Figure 6.1 **Old lady/young lady ambiguous image.**

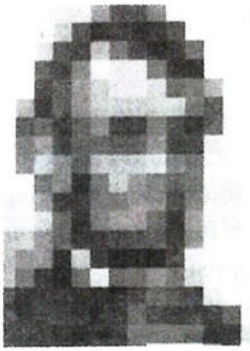


Figure 6.2 **We see what or who we think we know.** (From Harmon (1973))

in the United States then you may recognize who this is a picture of, even though there is no possible way you could have met the individual. If you do not live in the United States, then you may not recognize the face. The picture is a low-resolution image of the face on the US five dollar bill—President Abraham Lincoln. Culture and our memories (even if made up by ourselves or through someone else) contribute significantly to what we see. We don't see Abraham Lincoln because he looks like

Meta Platforms, Inc.
Exhibit 1063
Page 085

Abraham Lincoln, we see Abraham Lincoln because we "know" what Abraham Lincoln looks like even though none of us alive today have ever met him.

Einstein and Freud, although coming from very different realities, came to the same conclusion that reality is malleable, that our own point of view is an irreducible part of the equation of reality. Now with VR, not only do we get to experience the realities of others but we get to *create* objective realities that are more directly communicated (vs. indirectly through a book or television) for others to experience through their own subjectivity.

Science attempts to take the subjectivity out of the universe, but that can often only be done under controlled conditions. In our day-to-day lives outside of the lab, our perceptions depend upon context. VR enables us to take some of the subjectivity out of synthetically created worlds by controlling conditions as done in a lab. Although much of VR experiences occur within individuals' own minds, by understanding how different perceptual processes work we can better lead people's minds for them to more closely experience our created realities as we intend.

We go about our lives and subconsciously organize sensory inputs via our made-up "rules" and match those to our expected outcomes (Section 7.3). Have you ever drunk orange juice when you were expecting milk? If so, then you may have gained a new experience of orange juice; if we don't perceive what we expect to perceive then we can perceive things very differently. When we are surprised or in need of deeper understanding, then our brains further transform, analyze, and manipulate the data via conceptual knowledge. At that time, our mental model of the world can be modified to match future expectations.

Facts, algorithms, and geometry are not enough for VR. By better understanding human perception, we can build VR applications that enable us to more directly connect the objective realities we create, by manipulating digital bits, into the minds of users. We seek to better understand how humans perceive those bits in order to more directly convey our intention to users in a way that is well received and interpreted.

The good news is that VR does not need to perfectly replicate reality. Instead we just need to present the most important stimuli well and the mind will fill in the gaps.

6.2 Perceptual Illusions

For most situations, we feel like we have direct access to reality. However, in order to create that feeling of direct access, our brains integrate various forms of sensory input along with our expectations. In many cases, our brains have been hardwired to predict certain things based on repeated exposure over a lifetime of perceiving. This hard-wired subconscious knowledge influences the way we interpret and experience the

Meta Platforms, Inc.
Exhibit 1063
Page 086

world and enables our perception and thinking to take shortcuts, which provides more room for higher-level processing. Normally we don't notice these shortcuts. However, when atypical stimuli are apparent then visual illusions can result. These illusions can help us understand the shortcuts and how the brain works. The remarkable thing about some of these illusions is that even when you know things are not the way they seem, it is still difficult to perceive it in some other way. The subconscious mind is indeed powerful even when the conscious mind knows something to the contrary.

Our brains are constantly and subconsciously looking for patterns in order to make sense of the stream of information coming in from the senses. In fact, the subconscious can be thought of as a filter that only allows things out of the ordinary that are not predictable to pass into consciousness (Sections 7.9.3 and 10.3.1). In many cases, these predictable patterns are so hardwired into the brain that we perceive these patterns even when they do not exist in reality.

Perhaps the earliest philosophical discussion of perception and the illusion of reality created by patterns can be found within Plato's *The Republic* (Plato, ~350 BC). Plato used the analogy of a person facing the back of a cave that is alive with shadows, with the shadows being the only basis for inferring what real objects are. Plato's cave was actually a motivation that revolutionized the VR industry back in the 1990s with the creation of the CAVE VR system (Cruz-Neira et al., 1992) that projects light onto walls surrounding the user, creating an immersive VR experience (Section 3.2).

In addition to helping design VR systems, knowledge of perceptual illusions can help VR creators better understand human perception, identify and test assumptions, interpret experimental results, better design worlds, and solve problems as issues are discovered. The following sections demonstrate some virtual illusions that serve as building blocks for further investigation into perception.

6.2.1 2D Illusions

Even for simple 2D shapes we don't necessarily perceive the truth. The Jastrow illusion demonstrates how size can be misinterpreted—look at Figure 6.3 and try to determine which shape is larger. It turns out neither shape is larger—measure the height and width of the two shapes and you will find the shapes are identical, although the lower one appears to be larger.

The Hering illusion (Figure 6.4) is a geometrical perceptual illusion that demonstrates straight lines do not always appear to be straight. When two straight and parallel lines are presented in front of a radial background, the lines convincingly appear as if they are bowed outward.

Meta Platforms, Inc.
Exhibit 1063
Page 087

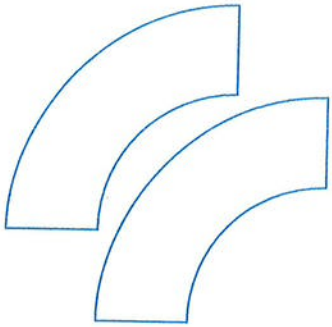


Figure 6.3 The Jastrow illusion. The two shapes are identical in size.

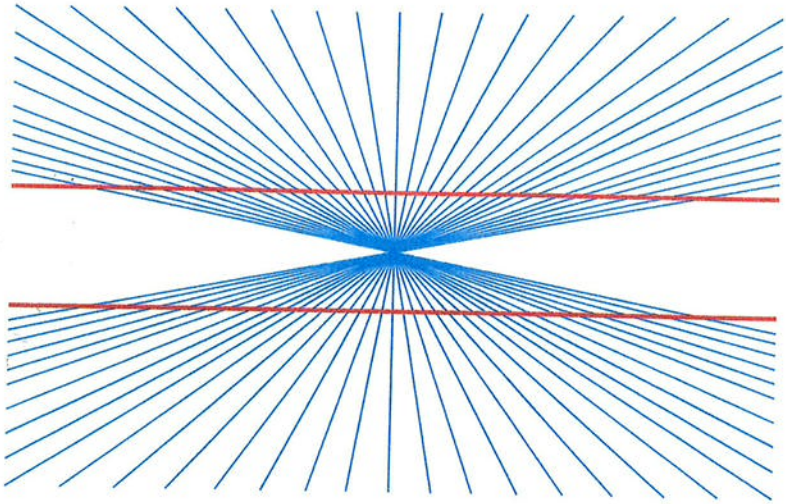


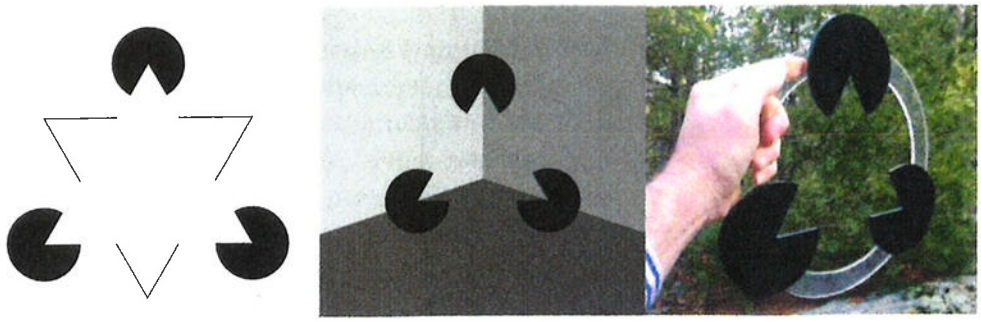
Figure 6.4 The Hering illusion. The two horizontal lines are perfectly parallel although they appear bent. (Courtesy of Fibonacci)

These simple 2D illusions show we should not necessarily trust our perceptions alone (although that is a good place to start) when calibrating VR systems, as judgments can depend upon the surrounding context.

6.2.2 Illusory Boundary Completion

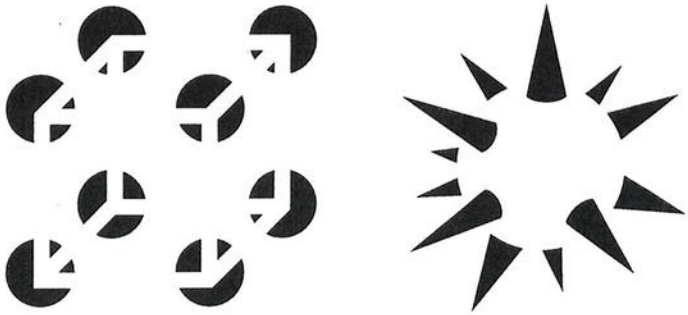
Figure 6.5 shows three images, each with three "Pac Man" shapes similar to the classic video game of the 1980s. Between the three Pac Men in the leftmost figure, you might see a large white triangle overlaid on another slightly darker upside-down triangle. The

Meta Platforms, Inc.
Exhibit 1063
Page 088



(a) (b) (c)

Figure 6.5 **The Kanizsa illusion in (a) its traditional form, (b) 3D form (From Guardini & Gamberini (2007)), (c) real-world form (Lehar (2007)). Triangles are perceived although they do not actually exist.**



(a) (b)

Figure 6.6 **Necker cube (a) and a spiky ball illusion (b). Similar to the Kanizsa illusion, shape is seen where shape does not exist.**

perception of a triangle is known as the Kanizsa illusion, a form of illusory contour where boundaries are perceived even when no boundary actually exists. If you cover up these Pac Men then the supposed edges and triangles disappear. The middle figure shows how a triangle might appear in the corner of a rendered room. The rightmost figure shows how such a non-existent shape can even appear in the context of the real world.

Figure 6.6 shows two more illusory contour images where our minds fill in the gaps resulting in perceived shapes where those shapes don't actually exist. Gestalt theory states that the whole is other than the sum of its individual parts and is described in Section 20.4.

Meta Platforms, Inc.
Exhibit 1063
Page 089

6.2.3 Blind Spot

The mind can also perceive emptiness even when something is right in front of us. The **blind spot** is an area on the retina where blood vessels leave the eye and there are no light-sensitive receptors. Blind spots are so large one can make a person's head disappear when they are sitting across the room (Gregory, 1997). We can see (or not see!) the effect of the blind spot by a simple diagram as demonstrated in Figure 6.7.

Illusory contours demonstrate that we do not need to perfectly replicate reality, but we do need to present the most essential stimuli to induce objects into users' minds.

6.2.4 Depth Illusions

The Ponzo Railroad Illusion

Figure 6.8 shows the Ponzo Railroad illusion where the top yellow rectangle appears to be larger than the bottom yellow rectangle. In actuality the two yellow lines/rectangles are identical in shape (try measuring them for yourself). This occurs because we interpret the converging sides according to linear perspective as parallel lines receding into the distance (Section 9.1.3). In this context, we interpret the upper yellow rectan-



Figure 6.7 **A simple demonstration of the blind spot. Close your right eye. With your left eye stare at the red dot and slowly move closer to the image. Notice that the two blue bars are now one!**



Figure 6.8 **The Ponzo Railroad illusion. The top yellow rectangle looks larger than the bottom yellow rectangle even though in actuality they are the same size.**

Meta Platforms, Inc.
Exhibit 1063
Page 090

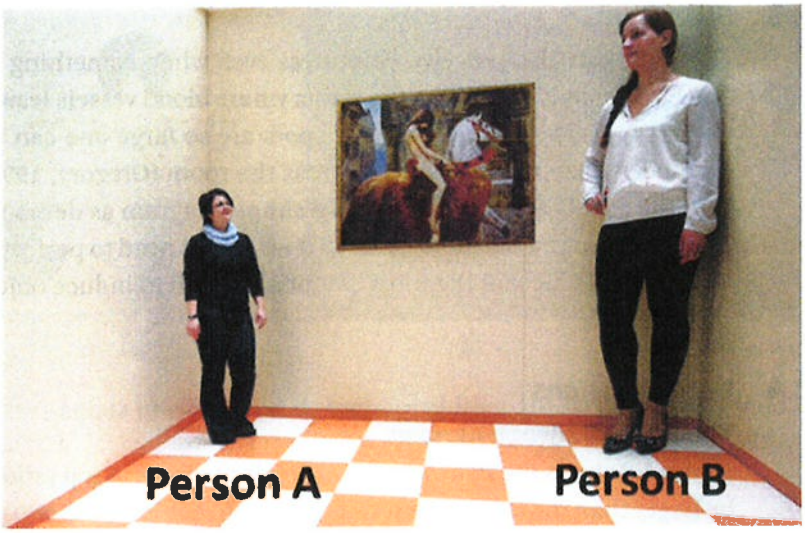


Figure 6.9 **An Ames room creates the illusion of dramatically different sized people. Even though we logically know such differences in human sizes are not possible, the scene is still convincing.** (Courtesy of PETER ENDIG/AFP/Getty Images)

gle as though it were farther away, so we see it as larger—a farther object would have to be larger in actual physical size than a nearer one for both to produce retinal images of the same size.

The Ames Room

In an Ames room such as that shown in Figure 6.9, a person standing in one corner appears to the observer to be a giant, while a person standing in the other corner appears to be impossibly small. The illusion is so convincing that a person walking back and forth from the left corner to the right corner appears to grow or shrink.

An Ames room is designed so that from a specific viewpoint, the room appears to be an ordinary room. However, this is a trick of perspective and the true shape of the room is trapezoidal: the walls are slanted and the ceiling and floor are at an incline, and the right corner is much closer to the viewer than the left corner (Figure 6.10). The resulting illusion shows our perceptions can be affected by surrounding context. The illusion also occurs because the room is viewed with one eye through a pinhole to avoid any clues from stereopsis or motion parallax. When creating VR worlds, we want to build with consistency across sensory modalities and know that misperceptions

Meta Platforms, Inc.
Exhibit 1063
Page 091

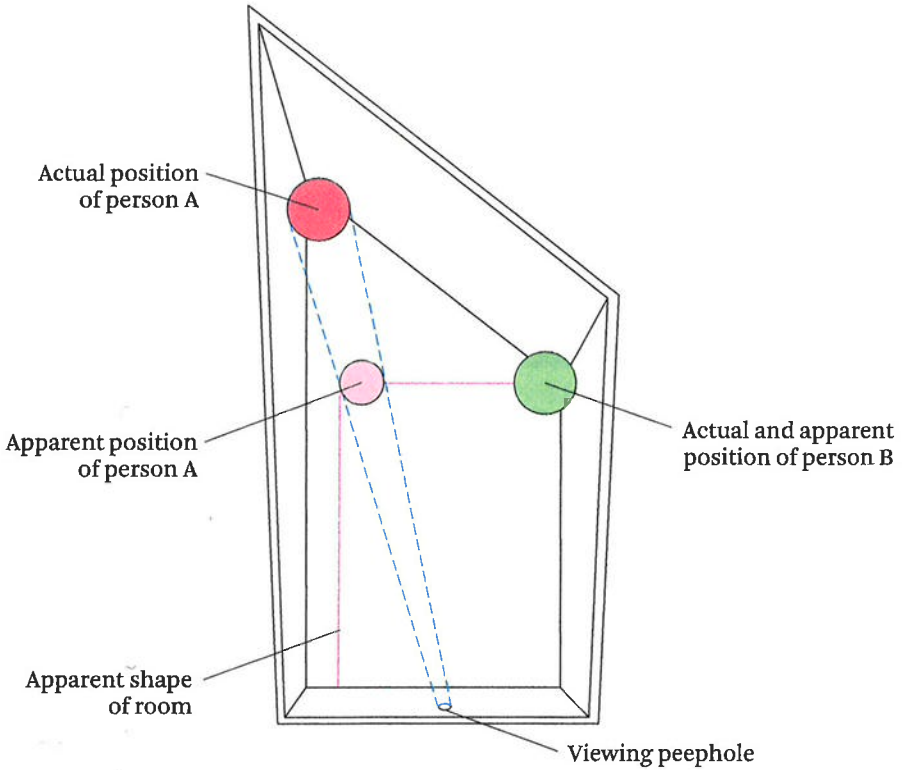


Figure 6.10 **The actual shape of an Ames room.**

can occur if we are not careful. Otherwise much stranger things than the Ames room illusion can occur!

6.2.5 The Moon Illusion

The **moon illusion** is the phenomenon that the moon appears larger when it is on the horizon than when it is high in the sky even though it remains the same distance from the observer and consists of the same visual angle. There are several explanations to the moon illusion (Figure 6.11), but the most likely is because when the moon is near the horizon there are other features on the terrain that can be directly compared to the moon. Because of the distance cues of the terrain and horizon we have a better sense that the moon is further away. The more distance cues, the stronger the illusion. Similar to the railroad perspective illusion, the moon that is thought of as further away seems larger than the zenith moon (where the distance to the sky is perceived to be closer than the horizon) even though both moons take up the same area on the retina.

Meta Platforms, Inc.
Exhibit 1063
Page 092



Figure 6.11 **The moon illusion is the perception that the moon is larger when on the horizon than when high in the sky. The illusion can be especially strong when there are strong depth cues such as occlusion shown here.**

6.2.6 Afterimages

A **negative afterimage** is an illusion where the eyes continue seeing the inverse colors of an object even after the object is no longer physically present. An afterimage can be demonstrated by staring at the small red, green, and yellow dots below the lady's eye in Figure 6.12. After staring at this dot for 30–60 seconds, look to the blank area to the right of the picture at the X. You should now see a realistically colored image of the lady. How does this occur? The photoreceptive cells in our eyes become less sensitive to a color after a period of time and when we then look at a white area (where white is a mix of all colors), the inhibited colors tend not to be seen, leaving the opposed color.

A **positive afterimage** is a shorter-term illusion that causes the same color to persist on the eye even after the original image is no longer present. This can result in the perception of multiple objects when a single object moves across a display (called strobing—see Section 9.3.6).

6.2.7 Motion Illusions

Motion illusions can certainly occur in VR, and such illusions can cause disorientation and sickness (Chapter 12). We can use motion illusions to our advantage to create a sense of self-motion called vection (Section 9.3.10).

The Ouchi Illusion

If you look at the Ouchi illusion shown in Figure 6.13, you may see the circle move left and/or right. If you yaw your head left/right while looking at the circle, the circle may

Meta Platforms, Inc.
Exhibit 1063
Page 093



Figure 6.12 **Afterimage effect. Stare at the small red, green, and yellow dots just below the eye for 30–60 seconds. Then look to the blank area to the right at the X.** (Image © 2015 The Association for Research in Vision and Opthalmology)

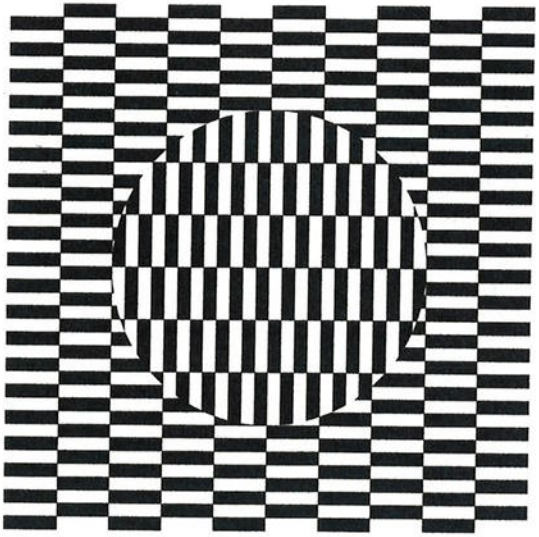


Figure 6.13 **The Ouchi illusion. The rectangles appear to move although the image is static.**

Meta Platforms, Inc.
Exhibit 1063
Page 094

appear to be unstable in front of a stable background. There are many explanations for such motion illusions, and the effect has been found to occur in a wide range of situations. One explanation is that the two differently oriented bars are perceived to be on different depth planes, causing them to seem to move relative to each other as the eyes move (similar to what would occur if the bars were actually on different depth planes).

Motion Aftereffects

Motion aftereffects are illusions that occur after one views stimuli moving in the same direction for 30 seconds or more. After this time, one may perceive the motion to slow down or stop completely due to fatigue of motion detection neurons (a form of sensory adaptation—see Section 10.2). When the motion stops or one looks away from the moving stimuli to non-moving stimuli, she may perceive motion in the opposite direction as the previously moving stimuli. We must therefore be careful when measuring perceptible motion or designing content for VR. For example, be careful of including consistently moving textures, such as moving water, in virtual worlds. Otherwise unintentional perceived motion may result after users look at such areas for long periods of time.

The Moon-Cloud Illusion

The moon-cloud illusion is an excellent example of induced motion—one may perceive clouds to be stable and the moon to be moving, when in fact the moon is stable and the clouds are moving. This illusion occurs because the mind assumes smaller objects are more likely to move than larger surrounding context (Section 9.3.5).

The Autokinetic Effect

The **autokinetic effect**, also known as autokinesis, is the apparent movement of a single stable point-light when no other visual cues are present. This effect occurs even when the head is held still. This is because the brain is not directly aware of unintentional small and slow eye movements (Section 8.1.5) and there is no surrounding visual context to judge motion (i.e., no object-relative motion cues are available—see Section 9.3.2), so the movement is ambiguous; i.e., did the movement occur due to the eye moving or due to the point-light source moving? An observer does not know the answer, and the amount and direction of perceived motion varies. The autokinetic effect decreases as the size of the target increases, due to the brain's assumption that objects taking up a large field of view are stable.

The autokinetic effect suggests that in order to create the illusion of stable worlds, large surrounding context should be used (Section 12.3.4).

Meta Platforms, Inc.
Exhibit 1063
Page 095

Perceptual Models and Processes

The human brain is quite complex and cannot be described by a single model or process. This chapter discusses some models and processes that provide different perspectives about how the mind works. Note these concepts are not necessarily absolute "truths" but are simplified models that provide useful ways to think about perception. These concepts help explain various factors that affect human perception and behavior; enable us to create virtual environments that better communicate with a diverse set of users; help us to better program interaction into our environments to better respond to users on their terms; and help us induce more engaging and fulfilling experiences into the minds of users.

7.1 Distal and Proximal Stimuli

To understand perception, we must first understand the difference between the concepts of distal (distant) stimuli and proximal (approximate = close) stimuli.

Distal stimuli are actual objects and events out in the world (objective reality).

Proximal stimuli are the patterns of stimuli from these objects and events that actually reach the senses (eyes, ears, skin, etc.).

The primary task of perception is to interpret distal stimuli in the world in order to guide effective survival-producing, goal-driven behaviors. One challenge is that a proximal stimulus is not always an ideal source of information about the distal stimulus. For example, light rays from distal objects form an image on the retina in the back of the eye. But this image continually changes as lighting varies, we move through the world, our eyes rotate, etc. Although this proximal image is what triggers the neural signals to the brain, most of the time we are often quite unaware of the overall image or pay little attention to it (Section 10.1). Instead, we are aware of and respond to the distal objects that the proximal stimulus represents.

Meta Platforms, Inc.
Exhibit 1063
Page 096

Perceptual Models and Processes

The human brain is quite complex and cannot be described by a single model or process. This chapter discusses some models and processes that provide different perspectives about how the mind works. Note these concepts are not necessarily absolute "truths" but are simplified models that provide useful ways to think about perception. These concepts help explain various factors that affect human perception and behavior; enable us to create virtual environments that better communicate with a diverse set of users; help us to better program interaction into our environments to better respond to users on their terms; and help us induce more engaging and fulfilling experiences into the minds of users.

7.1 Distal and Proximal Stimuli

To understand perception, we must first understand the difference between the concepts of distal (distant) stimuli and proximal (approximate = close) stimuli.

Distal stimuli are actual objects and events out in the world (objective reality).

Proximal stimuli are the patterns of stimuli from these objects and events that actually reach the senses (eyes, ears, skin, etc.).

The primary task of perception is to interpret distal stimuli in the world in order to guide effective survival-producing, goal-driven behaviors. One challenge is that a proximal stimulus is not always an ideal source of information about the distal stimulus. For example, light rays from distal objects form an image on the retina in the back of the eye. But this image continually changes as lighting varies, we move through the world, our eyes rotate, etc. Although this proximal image is what triggers the neural signals to the brain, most of the time we are often quite unaware of the overall image or pay little attention to it (Section 10.1). Instead, we are aware of and respond to the distal objects that the proximal stimulus represents.

Meta Platforms, Inc.
Exhibit 1063
Page 097

Factors that play a role in characterizing distal stimuli include context, prior knowledge, expectations, and other sensory input. Perceptual psychologists seek to understand how the brain extracts accurate stable perceptions of objects and events from limited and inadequate information.

Our job as VR creators is to create an artificial distal virtual world and then project that distal world onto the senses; the better we can do this in a way that the body and brain expects then the more users will feel present in the virtual world.

7.2 Sensation vs. Perception

Sensations are elementary processes that enable low-level recognition of distal stimuli through proximal stimuli. For example, in vision, sensation occurs as photons fall on the retina. In hearing, sensation occurs as waves of pulsating air are collected by the outer ear and are transmitted through the bones of the middle/inner ear to the cochlea.

Perception is a higher-level process that combines information from the senses and filters, organizes, and interprets those sensations to give meaning and to create subjective, conscious experiences. This gives us awareness of the world around us. Two people can receive the same sensory information yet perceive two different things. Moreover, even the same person can perceive the same stimulus differently (e.g., ambiguous figures as shown in Figure 6.1 or emotional responses when listening to the same song at different times).

7.2.1 Binding

Sensations of shapes, colors, motion, audio, proprioception, vestibular cues, and touch all cause independent firing of neurons in different parts of the cortex. **Binding** is the process by which stimuli are combined to create our conscious perception of a coherent object—a "red ball," for example (Goldstein, 2014). We have integrated perceptions of objects, with all the object's features bound together to create a single coherent perception of the object. Binding is also important to perceive multimodal events as coming from a single entity. However, sometimes we incorrectly assign features of one object to another object. An example is the ventriloquism effect that causes people to believe a sound comes from a source when it is actually coming from somewhere else (Section 8.7). These illusory conjunctions most typically occur when some event happens quickly (such as errors in eyewitness testimony of a crime), or the scene is only briefly presented to the observer. Maintaining spatial and temporal compliance (Section 25.1.5) across sensory modalities increases binding and is

Meta Platforms, Inc.
Exhibit 1063
Page 098

especially important for maximizing intuitive VR interactions as well as minimizing motion sickness (Section 12.3.1).

7.3 Bottom-Up and Top-Down Processing

Bottom-up processing (also called data-based processing) is processing that is based on stimuli reaching receptors (Goldstein, 2014) that provides the starting point for perception. Tracking the perceptual process from sensory stimulation to transduction to neural processing is an example of bottom-up processing. VR uses bottom-up stimuli (e.g., pixels on the display) that is sufficiently convincing (i.e., to create presence) to overcome top-down data that suggests we are just wearing an HMD.

Top-down processing (also called knowledge-based or conceptually based processing) refers to processing that is based on knowledge—that is, the influence of an observer's experiences and expectations on what he perceives. At any given time, an observer has a mental model (Section 7.8) of the world. For example, we know through experience that object properties tend to be constant (Section 10.1). Conversely, we modify what we expect based on situational properties or changeable aspects of the environment. The perception of a distal stimulus is heavily biased toward this internal model (Goldstein, 2007; Gregory, 1973). As stimuli become more complex, the role of top-down processing increases. Our knowledge of how things are in the world plays an important role in determining what we perceive. As VR creators, we take advantage of top-down processing (e.g., the user's experience from the real or virtual world) that enables us to create interesting content and more engaging stories.

7.4 Afference and Efference

Afferent nerve impulses travel from sensory receptors inward toward the central nervous system. **Efferent nerve** impulses travel from the central nervous system outward toward effectors such as muscles. **Efference copy** sends a signal equal to the efference to an area of the brain that predicts afference. This prediction allows the central nervous system to initiate responses before sensory feedback occurs. The brain compares the efference copy with the incoming afferent signals. If the signals match, then this suggests the afference is solely due to the observer's actions. This is known as **reafference**, since the afference reconfirms that the intended action has taken place. If the efference copy and afference do not match, then the observer perceives the stimulus to have occurred from a change in the external world (passive motion) instead of from the self (active motion).

Meta Platforms, Inc.
Exhibit 1063
Page 099

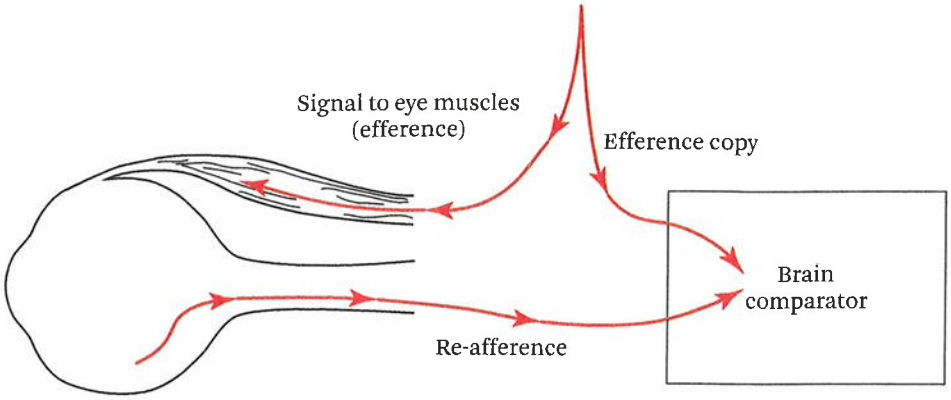


Figure 7.1 **Efference copy during rotation of the eye.** (Razzaque, 2005; Gregory, 1973)

Afference and efference copy result in the world being perceived differently depending upon whether the stimulus is applied actively or passively (Gregory, 1973). For example, when eye muscles move the eyeball (efference), efference copy matches the expected image motion on the retina (re-afference), resulting in the perception of a visually stable world (Figure 7.1). However, if an external force moves the eyeball the world seems to visually move. This can be demonstrated by closing one eye and gently pushing the skin covering the open eye near the temple. The finger forcing the movement of the eye is considered to be passive with respect to the way the brain normally controls eye movement. The perceived visual shifting of the scene is due to the brain receiving no efference copy from the brain's command to move the eye muscles. The afference (the movement of the scene on the retina) is then greater than the zero efference copy, so the external world is perceived to move.

7.5 Iterative Perceptual Processing

In a simplified form, perception can be considered an ever-evolving sequence of processing steps, as shown in Figure 7.2.

At a high level, the three steps of Goldstein's **iterative perceptual process** (Goldstein, 2007) are

1. stimulus to physiology—the external world falling onto our sensory receptors;
2. physiology to perception—our perceptual interpretation of those signals; and
3. perception to stimulus—our behavioral actions that influence the external world, resulting in an endless cycle that we call life.

Meta Platforms, Inc.
Exhibit 1063
Page 100

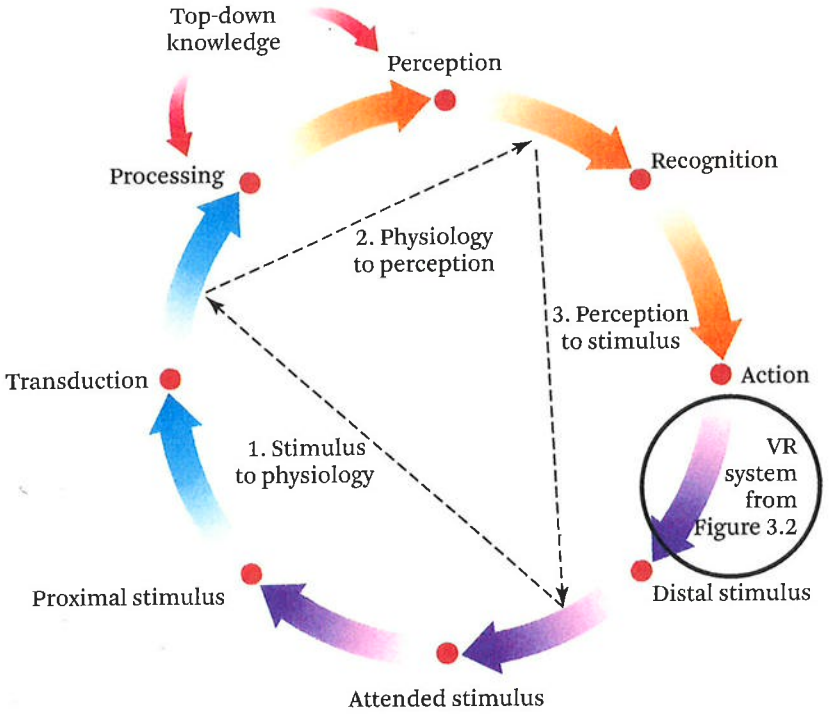


Figure 7.2 **Goldstein's (2007) perceptual process model. Perception is a continual process that is dynamic and continually changing. Purple arrows represent stimuli, blue arrows represent sensation, and orange arrows represent perception.**

These stages can be broken down further. As we journey through a world, we are confronted with a complex physical world (distal stimuli). Because there is far too much happening in our worlds, we scan the scene, looking at, listening to, or feeling from one place to another things that catch our interest (attended stimuli). Photons, vibrations, and pressure falls onto our body (proximal stimuli). Our organs that sense such physical properties are transformed into electrochemical signals (transduction). These signals then go through complex neurological interactions (processing) that are influenced by our past experiences (top-down knowledge). This results in our conscious sensory experience (perception). We then are able to categorize or give meaning to those perceptions (recognition). This determines our behavior (action), which in turn affects distal stimuli, and the process continues.

For VR, we hijack real-world distal stimuli and replace those stimuli with computer-generated distal stimuli as a function of geometric models and algorithms. If we could

Meta Platforms, Inc.
Exhibit 1063
Page 101

somehow perfectly replicate the real world with our models and algorithms, then VR would be indistinguishable from reality.

7.6 The Subconscious and Conscious

The human mind can be thought as two different components—the subconscious and the conscious. Table 7.1 summarizes characteristics of the two (Norman, 2013).

The **subconscious** is everything that occurs within our mind that we are not fully aware of but very much influences our emotions and behaviors. Because subconscious processing happens in parallel and without our awareness, the processing seems to occur naturally and automatically without effort. The subconscious is very much a pattern finder, generalizing incoming stimuli into the context of the past. Whereas the subconscious is powerful, it can also heavily bias us toward decisions based on false trends that cause us to behave inappropriately. The subconscious also is limited in the sense that it cannot arrange symbols in an intelligent fashion or make specific plans more than a couple of steps into the future.

When something comes into our awareness, it becomes part of our conscious. The **conscious** is the mind's totality of sensations, perceptions, ideas, attitudes, and feelings that we are aware of at any given time. The conscious is good at considering, pondering, analyzing, comparing, explaining, and rationalizing. However, the mind is slow as it primarily processes in a sequential manner and can become overwhelmed, resulting in analysis paralysis. The written word, mathematics, and programming are very much tools of the conscious.

Although the subconscious and conscious are each subject to mistakes, misunderstandings, and suboptimal solutions, the strengths of each can help balance disadvantages of the other. The integration of the two parts together often creates insight and innovation that would otherwise not be realized.

Table 7.1 **Subconscious and conscious thought.** (From Norman (2013))

Subconscious	Conscious
Fast	Slow
Automatic	Controlled
Multiple resources	Limited resources
Controls skilled behavior	Invoked for novel situations when learning, when in danger, when things go wrong

Meta Platforms, Inc.
Exhibit 1063
Page 102

7.7 Visceral, Behavioral, Reflective, and Emotional Processes

A useful approximate model of human cognition and reaction can be categorized into four levels of processing: visceral, behavioral, reflective, and emotional (Norman, 2013). All four can be thought of as different forms of pattern recognition and prediction, and work together to determine a person's state. High-level cognitive reflection and emotion can trigger low-level visceral and physiological responses. Conversely, low-level visceral responses can trigger high-level reflective thinking.

7.7.1 Visceral Processing

Visceral processes, often referred to as "the lizard brain," are tightly coupled with the motor system, serving as reflexive and protective mechanisms to help with immediate survival. Visceral communication (Section 1.2.1) results in immediate reaction and judgments such as a situation or person being dangerous. Visceral reactions are immediate; there is no time to determine cause, or to blame, or even to feel emotions. Visceral responses are precursors to many emotions and can also be caused by emotions. Although high-level learning does not occur through visceral processes, sensitization and desensitization can occur through sensory adaptation (Section 10.2).

Examples of visceral reactions include the startle reflex resulting from unexpected events, fight-or-flight responses, fear of heights, and the childhood anxiety produced by the lights being shut off. The visceral response to fear of heights is especially popular for demonstrating the power of VR (e.g., viewing the world from the top edge of a tall building).

Visceral processes are the initial reaction that is all about attraction or repulsion and have little to do with how usable, effective, or understandable a product or VR experience is. Many of us who consider ourselves "smart" tend to dismiss visceral responses as irrelevant and have trouble understanding the importance of an experience "feeling better." Great designers use their aesthetic intuition to drive visceral responses that result in positive states and high affection from users.

7.7.2 Behavioral Processing

Behavioral processes are learned skills and intuitive interactions triggered by situations that match stored neural patterns, and are largely subconscious. Even though we are usually aware of our actions, we are often unaware of the details.

When we consciously act, all we have to do is think of the goal; then the body and mind handle most of the details with little awareness of those details. An example is when one picks up an object, the person first wills the action and then it happens—there is no need to consciously orient and position the hand. If a VR user wishes to grab an object in the world, she simply thinks of grabbing an object, the hand moves

Meta Platforms, Inc.
Exhibit 1063
Page 103

toward the object to intersect it, then pushes a button to pick it up (assuming she has learned how to use a tracked controller).

The desire to control the environment gives motivation to learning new behaviors. When control cannot be obtained or things do not go as planned, then frustration and anger often results. This is especially true when new useful behaviors cannot be learned due to not understanding the reasons why something doesn't work. Feedback (Section 25.1.4) is critical to learning new interfaces, even if that feedback is negative, and VR applications should always provide some form of feedback.

7.7.3 Reflective Processing

Reflective processing ranges from basic conscious thinking to examining one's own thoughts and feelings. Higher-level reflection gives understanding and enables one to make logical decisions. Reflective processing is much slower than visceral and behavioral processing; reflective processing typically occurs after an event has happened. Indirect communication (Section 1.2.2) such as spoken language or internal thoughts (i.e., speaking to oneself) is the tool of reflective processing. In contrast to visceral processing, reflective processing evaluates circumstances, assigns cause, and places blame.

Reflective processing is where we create stories within our minds (that might be far from what actually happened). Reflective stories and memories are often more important than what previously actually happened because the past no longer exists—what is more important is to predict and plan into the future based off the stories that we created. Reflective stories create the highest levels of emotion—highs and lows of remembered or anticipated events. Emotion and cognition are much more tightly intertwined than most people believe.

7.7.4 Emotional Processing

Emotional processing is the affective aspect of consciousness that powerfully processes data, resulting in physiological and psychological visceral and behavioral responses. This is done in a way that requires less effort and occurs faster than more reflective thinking—providing intuitive insight to assess a situation or whether a goal is desirable or not.

If a person did not have emotions, then making decisions would be difficult as emotion assigns subjective value and judgments, whereas rational thinking assigns more objective understanding. Emotions are also more easily recalled than objective facts that are not associated with emotions. Logic is often used to backward-rationalize a decision that has already been made emotionally. Once a person has made an emotional decision, then some significant event or insight must occur for them to change

Meta Platforms, Inc.
Exhibit 1063
Page 104

that decision. People use emotion to pursue rational thought, such as being motivated to pursue the study of mathematics, art, or VR. Emotions can also be detrimental because they evolved in order to maximize immediate survival and reproduction rather than to achieve longer-term goals that empower us to move beyond just survival.

The emotional connection at the end of a VR experience is often what is remembered, so if the end experience is good the memories may be positive—but only if the user can make it that far. Conversely, one bad initial exposure can be so bad that they may never want to experience that application again, or in the worst case never want to experience any VR again.

7.8 Mental Models

A **mental model** is a simplified explanation in the mind of how the world or some specific aspect of the world works. The primary purpose of a mental model is prediction, and the best model is the simplest one that enables one to predict how a situation will turn out well enough to take appropriate action.

An analogy of a mental model is a physical map of the real world. The map is not the territory—meaning we form subjective representations (or maps) of objective reality (the territory), but the representation is not actually the reality, just as a map of a place is not actually the place itself but a more or less accurate representation (Korzybski, 1933). Most people confuse their maps of reality with reality itself. A person's interpretation/model created in the mind (the map) is not the truth (the territory), although it might be a simplified version of the truth. The value of a model is in the result that can be produced (e.g., an entertaining experience or learning new skills).

We all create mental models of ourselves, others, the environment, and objects we interact with. Different people often have different mental models of the same thing as these models are created through top-down processes that are derived from previous experience, training, and instruction that may not be consistent. In fact, an individual might have multiple models of the same thing, with each model dealing with a different aspect of its operation (Norman, 2013). These models might even be in conflict with each other.

A good mental model need not be complete or even accurate as long as it is useful. Ideally, complexity should be able to be intuitively understood with the simplest mental model possible that can achieve the desired results, but not simpler. An oversimplified model can be dangerous when implicit assumptions are not valid, in which case the model might be erroneous and therefore result in difficulties. Useful models serve as guides to help with understanding of a world, predict how objects in

Meta Platforms, Inc.
Exhibit 1063
Page 105

that world will behave, and interact with the world in order to achieve goals. Without such a mental model, we would fumble around mindlessly without the intelligence of fully appreciating why something works and what to expect from our actions. We could manage by following directions precisely, but when problems occur or an unexpected situation arises, then a good mental model is essential. VR creators should use signifying cues, feedback, and constraints (Section 25.1) to help the user form quality mental models and to make assumptions explicit.

Learned helplessness is the decision that something cannot be done, at least by the person making the decision, resulting in giving up due to a perceived absence of control. It has been difficult for VR to gain wide adoption, and this can be compounded by poor interfaces. Bad interaction design for even a single task can quickly lead to learned helplessness—a couple of failures not only can erode confidence for the specific task but can generalize to the belief that interacting in VR is difficult. If users don't get feedback of at least a sign of minimal success, then they may get bored, not realizing they could do more, and mentally or physically eject themselves from the experience.

7.9 Neuro-Linguistic Programming

Neuro-linguistic programming (NLP) is a psychological approach to communication, personal development, and psychotherapy that is built on the concept of mental models (Section 7.8). NLP itself is a model that explains how humans process stimuli that enter into the mind through the senses, and helps to explain how we perceive, communicate, learn, and behave. Its creators, Richard Bandler and John Grinder, claim NLP connects neurological processes ("neuro") to language ("linguistic") and our behaviors are the results of psychological "programs" built by experiences. Furthermore, they claim that skills of experts can be acquired by modeling the neuro-linguistic patterns of those experts. NLP is quite controversial and some of the claims/details are disputed. However, even if some of the specific details of NLP do not describe neurological processes accurately, some NLP high-level concepts are useful in thinking about how users interpret stimuli presented to them by a VR system.

Everyone perceives and acts in their own unique way, even when they are in the same situation. This is because humans manipulate pure sensory information depending upon their "program." Each individual's program is determined by the sum of one's entire life experiences, so each program is unique. With this being said, there are patterns to programs and we can make different people's experiences of an event more consistent if we understand the mind's programs. In the case of VR, we can con-

Meta Platforms, Inc.
Exhibit 1063
Page 106

trol what stimuli are presented to users and influence, not control, what that person experiences. Understanding how sensory cues are integrated into the mind and precisely controlling stimuli enables us to more directly and consistently communicate and influence users' experiences. This understanding also helps us to change what is presented to users based on users' behaviors and personal preferences.

7.9.1 The NLP Communication Model

The NLP communication model describes how external stimuli affect our internal state and behavior. As shown in Figure 7.3, the NLP communication model (James and Woodsmall, 1988) can be divided into three components: external stimuli, filters, and internal state. External stimuli enter through our senses and then are filtered in various ways before integrating with our internal state. The person then subconsciously and consciously behaves in some way that potentially modifies external stimuli, creating a continuously repeating communication cycle.

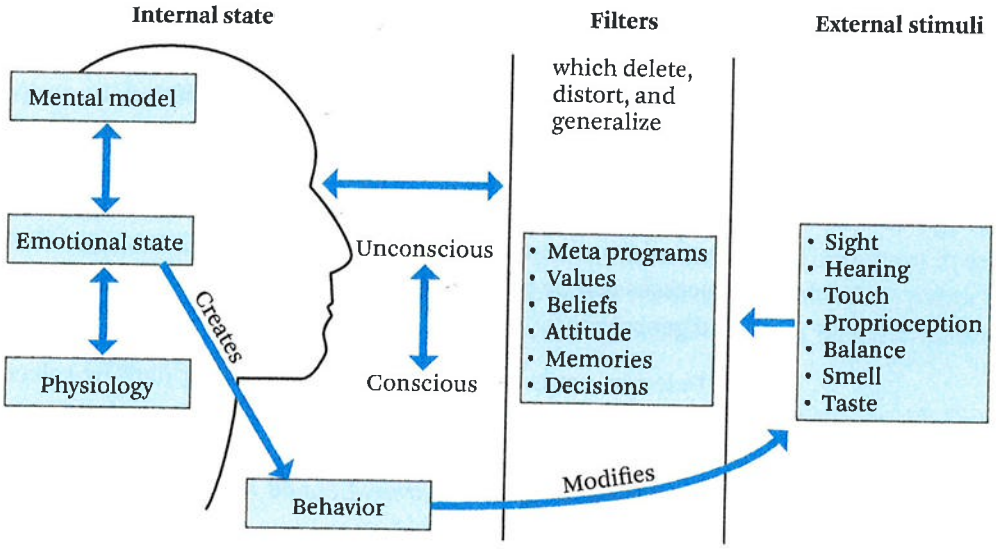


Figure 7.3 **The NLP communication model states that external events are filtered as they enter into the mind, and then result in an individual's specific behavior.** (Adapted from James and Woodsmall (1988))

Meta Platforms, Inc.
Exhibit 1063
Page 107

7.9.2 External Stimuli

NLP asserts that humans perceive events and external stimuli through the sensory input channels. Furthermore, individuals have a dominant or **preferred sensory modality**. Not only do individuals perceive their preferred modality better but their primary method of thinking is in terms of that preferred modality. So a visual person thinks more with pictures than an auditory person, who thinks more with sounds. Even when external stimuli are in one form, they are often perceived partially in the preferred modality. For example, visually dominant individuals will react more to visual words that convey color or form. These concepts of preferred modalities are especially important for education as students learn in different ways. By providing sensory cues across all modalities, individuals are more likely to comprehend the concept being presented. For example, kinesthetic-oriented students learn best by doing rather than by watching or listening. One trap VR creators often fall into is that they focus on their preferred modality, when many users of their system may have a different preferred modality. VR creators should present sensory cues across all modalities to cover a wider range of users.

7.9.3 Filters

A vast majority of sensory information that we absorb is assimilated by our minds subconsciously. While our subconscious minds are able to absorb and store a large portion of what is happening around us, our conscious minds can only handle about 7 ± 2 "chunks" (Csikszentmihalyi, 2008). In addition, our minds interpret incoming data in different ways depending on the context and our past experiences in order to comprehend, discern, and utilize the complexity of the world. If our minds objectively accepted all information, then our lives would immediately become overwhelmed and we would not be able to sustain our individual selves or society as a whole. These processes can be thought of as perceptual filters that do three things: delete, distort, and generalize.

Deletion omits certain aspects of incoming data by selectively paying attention to only parts of the world. For specific moments of time, we focus on what is most important and the rest is deleted from our conscious awareness in order to prevent becoming overwhelmed with information.

Distortion modifies incoming sensory information, enabling us to take into account the context of the environment and previous experiences. This is often beneficial but can be deceiving when something that happens in the world does not match one's mental model of the world. Being intimidated by technology, frightened of situations that we logically know are not real, or misinterpreting

Meta Platforms, Inc.
Exhibit 1063
Page 108

what occurred in a simulation are examples of how we distort incoming information. Examples of distortion include polarization (thinking of things in absolute extremes—e.g., the enemy in a game (or war) is wholly evil), mind reading (e.g., that avatar doesn't like me), and catastrophizing (e.g., I'm going to die if the goblin hits me).

Generalization is the process of drawing global conclusions based off of one or more experiences. A majority of the time generalization is beneficial. For example, a user that works through a tutorial that teaches how to open a door for the first time quickly generalizes their new ability so that all types of doors can be opened from then on. Designing VR experiences that are consistent will, over time, cause users to form filters that generalize perception of events and interactions. Generalization can also cause problems such as occurs with learned helplessness (Section 7.8). For example, a user might generalize that memorizing VR controls is difficult after a couple of failed attempts.

The filters that delete, distort, and generalize vary from deeply unconscious processes to more conscious processes. Different people perceive things in different ways even when perceiving the same thing. Listed below are descriptions of these filters in approximate order of increasing awareness.

Meta programs are the most unconscious filters and are content-free applying to all situations. Some consider meta programs to be one's "hardware" that individuals are born with and are difficult to change. Meta programs are similar to personality types. An example of a meta program is "optimism" where a person is generally optimistic regardless of the situation.

Values are filters that automatically judge if things are good or bad, or right or wrong. Ironically, values are context specific; therefore, what's important in one area of a person's life may not be important in other areas of his life. A player of a video game may value game characters very differently than how that person values real living people.

Beliefs are convictions about what is true or real in and about the world. Deep beliefs are assumptions that are not consciously realized. Weaker beliefs can be modified when brought to conscious thought and questioned by others. Beliefs can very much empower or hinder performance in VR (e.g., the belief "I'm not smart enough" can dramatically reduce performance).

Attitudes are values and belief systems about specific subjects. Attitudes are sometimes conscious and sometimes subconscious. An example of an attitude is "I don't like this part of the training."

Meta Platforms, Inc.
Exhibit 1063
Page 109

Memories are past experiences that are reenacted consciously in the present moment. Individual and collective experiences influence current perceptions and behaviors. Designers can take advantage of memories by providing cues in the environment that remind users of some past event.

Decisions are conclusions or resolutions reached after reflection and consideration. Past decisions affect present decisions and people tend to stick with a decision once it is made. Past decisions are what create values, beliefs, and attitudes, and therefore influence how we make current decisions and respond to current situations. A person who decided they liked a first VR experience will certainly be more open to new VR experiences, whereas someone who had an initial bad VR experience will be more difficult to convince VR is the future, even with higher-quality subsequent experiences. Make sure users have a positive experience from the beginning of each session so that they are more likely to decide they like what they experience overall.

Although VR creators don't have control over users' filters, learning about the target user's general values, beliefs, attitudes, and memories can help in creating a VR experience. Section 31.10 discusses how to create personas to better target a VR application to users' wants and needs.

7.9.4 Internal State

As incoming information passes through our filters, thoughts are constructed in the form of a mental model called an internal representation. This internal representation takes the form of sensory perceptions (e.g., a visual scene with sounds, tastes, and smells) that may or may not exist as external stimuli. This model triggers and is triggered by emotional states, which in turn changes one's physiology and motivates behavior. Although internal representation and emotional state cannot be directly measured, physiological state and behavior can.

Meta Platforms, Inc.
Exhibit 1063
Page 110

Perceptual Modalities

Mind-machine collaboration demands a two-way channel. The broadband path into the mind is via the eyes. It is not, however, the only path. The ears are especially good for situational awareness, monitoring alerts, sensing environmental changes, and speech. The haptic (feeling) and the olfactory systems seem to access deeper levels of consciousness. Our language is rich in metaphors suggesting this depth. We have "feelings" about complex cognitive situations on which we need to "get a handle" because we "smell" a rat.

—Frederick P. Brooks, Jr. (2010)

We interact with the world through sight, hearing, touch, proprioception, balance/motion, smell, and taste. This chapter discusses all of these with a focus on aspects of each that most relate to VR.

8.1 Sight

8.1.1 The Visual System

Light falls onto photoreceptors in the retina of the eye. These photoreceptors transduce photons into electrochemical signals that travel through different pathways in the brain.

The Photoreceptors: Cones and Rods

The retina is a multilayered network of neurons covering the inside back of the eye that processes photon input. The first retinal layer includes two types of receptors—cones and rods. Cones are primarily responsible for vision during high levels of illumination, color vision, and detailed vision. The fovea is a small area in the center of the retina that contains only cones packed densely together. The fovea is located on the line of sight, so that when a person looks at something, its image falls on the fovea. Rods are primarily responsible for vision at low levels of illumination and are located across

Meta Platforms, Inc.
Exhibit 1063
Page 111

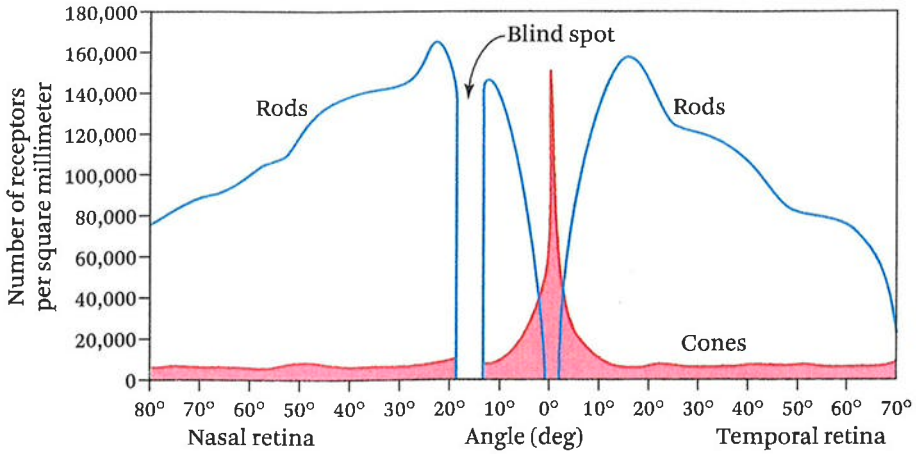


Figure 8.1 **Distribution of rods and cones in the eye.**

the retina everywhere except the fovea. Rods are extremely sensitive in the dark but cannot resolve fine details. Figure 8.1 shows the distribution of rods and cones across the retina.

Electrochemical signals from multiple photoreceptors converge to single neurons in the retina. The number of converging signals per neuron increases toward the periphery, resulting in higher sensitivity to light but decreased visual acuity. In the fovea, some cones have a "private line" to structures deeper within the brain (Goldstein, 2007).

Beyond Cones and Rods: Parvo and Magno Cells

Beyond the first layer of the retina, neurons relating to vision come in many shapes and sizes. Scientists categorize these neurons into parvo cells (those with smaller bodies) and magno cells (those with larger bodies). There are more parvo cells than magno cells with magno cells increasing in distribution toward peripheral vision whereas parvo cells decrease in distribution toward peripheral vision (Coren et al. 1999).

Parvo cells have a slower conduction rate (20 m/s) compared to magno cells, have a sustained response (continual neural activity as long as the stimulus remains), have a small receptive field (the area on the retina that influences the firing rate of the neuron), and are color sensitive. These qualities result in parvo cells being optimized for local shape, spatial analysis, color vision, and texture. Although parvo cells are located throughout all areas of the visual field, a greater portion of parvo neurons are dedicated to the fovea.

Meta Platforms, Inc.
Exhibit 1063
Page 112

Magno cells are color blind, have a large receptive field, have a rapid conduction rate (40 m/s), and have a transient response (the neurons briefly fire when a change occurs and then stop responding). Magno cells are optimized for motion detection, time keeping operations/temporal analysis, and depth perception. Magno neurons enable observers to quickly perceive large visuals before perceiving small details.

Multiple Visual Pathways

Signals from the retina travel along the optic nerve before diverging along different paths.

The primitive visual pathway. The **primitive visual pathway** (also known as the tectopulvinar system) starts to branch off at the end of the optic nerve (about 10% of retinal signals, mostly magno cells) toward the **superior colliculus**, a part of the brain just above the brain stem that is much older in evolutionary terms and a more primitive visual center than the cortex. The superior colliculus is highly sensitive to motion and is a primary cause of VR sickness (Part III) as it alters the response to the vestibular system, plays an important role in reflexive eye/neck movement and fixation, changes the curvature of the lens to bring objects into focus (eye accommodation), mediates postural adjustments to visual stimuli, induces nausea, and coordinates the diaphragm and abdominal muscles that evoke vomiting (Goldstein, 2014; Coren et al., 1999; Siegel and Sapru, 2014; Lackner, 2014).

The superior colliculus also receives input from the auditory and somatic (e.g., touch and proprioception) sensory systems as well as the visual cortex. Signals from the visual cortex are called **back projections**—a form of feedback from higher-level areas of the brain based on previous information that has already been processed (i.e., top-down processing).

The primary visual pathway. The **primary visual pathway** (also known as the geniculostriate system) goes through the **lateral geniculate nucleus** (LGN) in the thalamus. The LGN acts as a relay center to send visual signals to different parts of the visual cortex. The LGN is represented by central vision more than peripheral vision. In addition to receiving information from the eye, the LGN receives a large amount of information (as much as 80–90%) from higher visual centers in the cortex (back projections) (Coren et al., 1999). The LGN also receives information from the reticular activating system, the part of the brain that helps to mediate increased alertness and attention (Section 10.3.1). Thus, LGN processes are not solely a function of the eyes—visual processing is influenced by both bottom-up (from the retina) and top-down processing (from the reticular activating system and cortex) as described in Section 7.3. In fact,

Meta Platforms, Inc.
Exhibit 1063
Page 113

the LGN receives more information from the cortex than it sends out to the cortex. What we see is largely based on our experience and how we think.

Signals from the parvo and magno cells in the LGN feed into the visual cortex (and conversely back projections from the visual cortex feed into the LGN). After the visual cortex, signals diverge into two different pathways toward the temporal and parietal lobes. The path to the temporal lobe is known as the ventral or "what" pathway. The path to the parietal lobe is known as the dorsal or "where/how/action" pathway. Note the pathways are not totally independent and signals flow in both directions along each pathway.

The what pathway. The **ventral pathway** leads to the temporal lobe, which is responsible for recognizing and determining an objects' identity. Those with damaged temporal lobes can see but have difficulty in counting dots in an array, recognizing new faces, and placing pictures in a sequence that relates a meaningful story or pattern (Coren et al., 1999). Thus, the ventral pathway is commonly called the "what" pathway.

The where/how/action pathway. The **dorsal pathway** leads to the parietal lobe, which is responsible for determining an object's location (as well as other responsibilities). Picking up an object requires knowing not only where the object is located but also where the hand is located and how it is moving. The parietal reach region at the end of this pathway is the area of the brain that controls both reaching and grasping (Goldstein, 2014). Thus, the dorsal pathway is commonly called the "where," "how," or "action" pathway. Section 10.4 discusses action, which is heavily influenced by the dorsal pathway.

Central vs. Peripheral Vision

Central and peripheral vision have different properties, due to not only the retina but also the different visual pathways described above. Central vision

- has high visual acuity,
- is optimized for bright daytime conditions, and
- is color-sensitive.

Peripheral vision

- is color insensitive,
- is more sensitive to light than central vision in dark conditions,
- is less sensitive to longer wavelengths (i.e., red),
- has fast response and is more sensitive to fast motion and flicker, and
- is less sensitive to slow motions.

Meta Platforms, Inc.
Exhibit 1063
Page 114

Sensitivity to motion in central and peripheral vision is described further in Section 9.3.4.

Field of View and Field of Regard

Although we use central vision to see detail, peripheral vision is extremely important to function in real or virtual worlds. In fact, peripheral vision is so important that the US government defines those who cannot see more than 20° in the better eye as legally blind.

The **field of view** is the extent of the world that can be seen at a single point in time. As can be seen in Figure 8.2, a human eye has about a 160° horizontal field of view and both eyes can see the same area over an angle of about 120° when looking straight ahead (Badcock et al., 2014). The total horizontal field of view when looking straight ahead is about 200°—thus we can see "behind" us by 10° on each side of the head! If

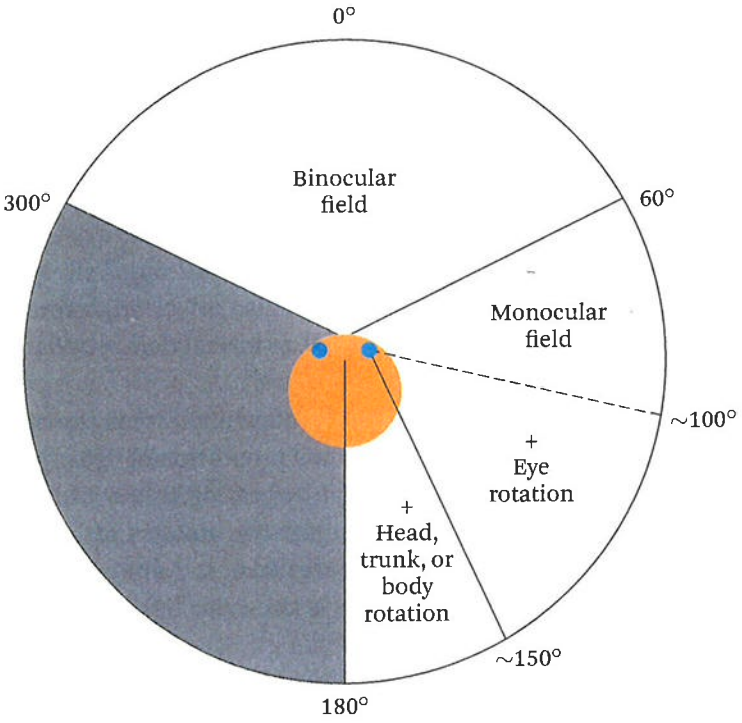


Figure 8.2 **Horizontal field of view of the right eye with straight-ahead fixation (looking toward the top of diagram), maximum lateral eye rotation, and maximum lateral head rotation.** (Based on Badcock et al. (2014))

Meta Platforms, Inc.
Exhibit 1063
Page 115

we rotate our eyes to one side or another, we can see an additional 50° on each side—thus for an HMD to cover our entire visual potential range, we would need an HMD with 300° horizontal field of view! Of course we can also turn our bodies and head to see 360° in all directions. The measure for what can be seen by physically rotating the body and head is known as **field of regard.** Fully immersive VR has the capability to provide a 360° horizontal and vertical field of regard.

Our eyes cannot see nearly as much in the vertical direction due to our foreheads, torso, and the eyes not being vertically additive. We only see about 60° above due to the forehead getting in the way whereas we can see 75° below, for a total of 135° vertical field of view (Spector, 1990).

8.1.2 Brightness and Lightness

Brightness is the apparent intensity of light that illuminates a region of the visual field. Under ideal conditions, a single photon can stimulate a rod and we can perceive as few as six photons hitting six rods (Coren et al., 1999). Even at higher levels of illumination, fluctuations of only a few photons can affect brightness perception.

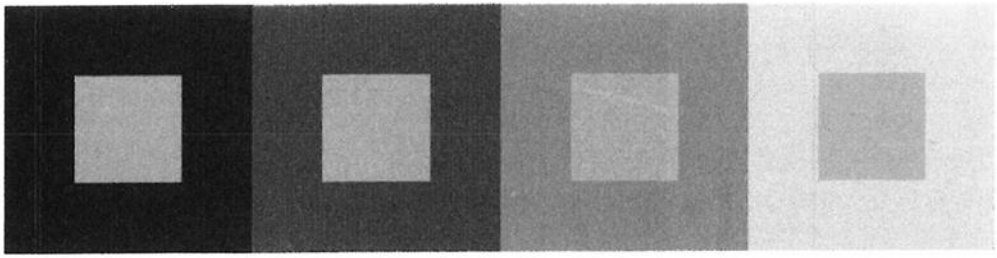
Brightness is not simply explained by the amount of light reaching the eye, but is also dependent on several other factors. Different wavelengths appear as different brightnesses (e.g., yellow light appears to be brighter than blue light). Peripheral vision is more sensitive than foveal vision (looking directly at a dim star can cause it to disappear) for all but longer wavelengths that appear as red in foveal vision. Dark-adapted vision provides six orders of magnitude more sensitivity than light-adapted vision (Section 10.2.1). Longer bursts of light (up to about 100 ms) are more easily detected than shorter bursts, and increasing stimulus sizes are more easily detected up to about 24°. Surrounding stimuli also affect brightness perception. Figure 8.3 shows how a light square can be made to appear darker by darkening the surrounding background.

Lightness (sometimes referred to as whiteness) is the apparent reflectance of a surface, with objects reflecting a small proportion of light appearing dark and objects reflecting a larger proportion of light appearing light/white. The lightness of an object is a function of much more than just the amount of light reaching the eyes. The perception that objects appear to maintain the same lightness even as the amount of light reaching the eyes changes is known as lightness constancy and is discussed in Section 10.1.1.

8.1.3 Color

Colors do not exist in the world outside of ourselves, but are created by our perceptual system. What exists in objective reality are different wavelengths of electromagnetic

Meta Platforms, Inc.
Exhibit 1063
Page 116



(a) (b) (c) (d)

Figure 8.3 **The luminance of surrounding stimuli affects our perception of brightness. All four squares have the same luminance, but (d) is perceived to be darkest.** (Based on Coren et al. (1999))

radiation. Although what we call colors is systematically related to light wavelengths, there is nothing intrinsically "blue" about short wavelengths or "red" about longer wavelengths.

We perceive colors in wavelengths from about 360 nm (violet) to 830 nm (red) (Pokorny and Smith, 1986), and bands of wavelengths within this range are associated with different colors. Colors of objects are largely determined by wavelengths of light that are reflected from the objects into our eyes, which contain three different cone visual pigments with different absorption spectra. Figure 8.4 shows plots of reflected light as a function of wavelength for some common objects. Black paper and white paper both reflect all wavelengths to some extent. When some objects reflect more of some wavelengths than other wavelengths, we call these chromatic colors or hues.

We can add more variation to hues by changing the intensity and by adding white to change the saturation (e.g., changing red to pink). **Color vision** is the ability to discriminate between stimuli of equal luminance on the basis of wavelength alone. Given equal intensity and saturation, people can discriminate between about 200 colors, but by changing the wavelength, intensity, and saturation, people can differentiate between about one million colors (Goldstein, 2014). Color discrimination ability decreases with luminance, especially at lower wavelengths, but remains fairly constant at higher luminance. Chromatic discrimination is poor at eccentricities beyond 8°.

Colors are more than just a direct effect of the physical variation of wavelengths—colors can subconsciously evoke our emotions and affect our decisions (Coren et al., 1999). Colors can delight and impress. The color of detergent changes how consumers rate the strength of the detergent. The color of pills may affect whether patients take prescribed medications—black, gray, tan, or brown pills are rejected whereas blue,

Meta Platforms, Inc.
Exhibit 1063
Page 117

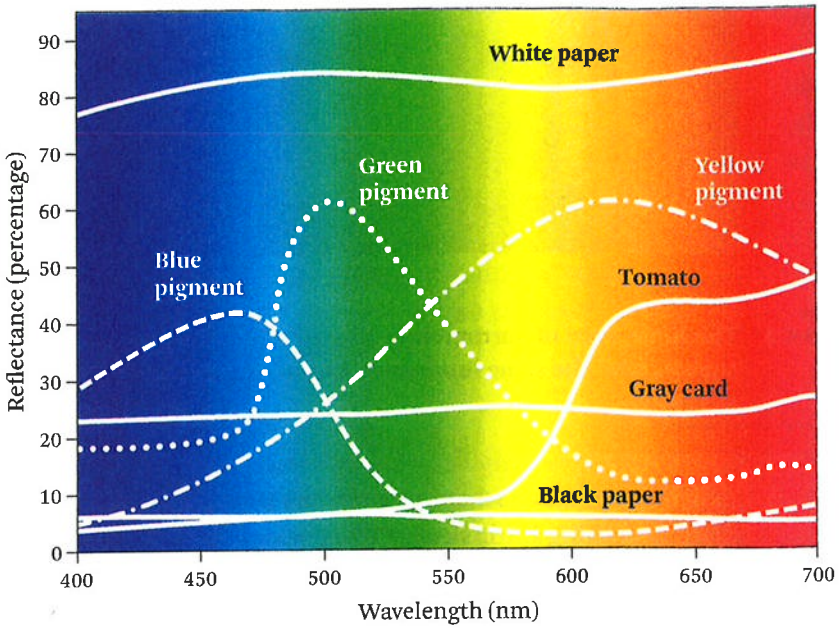


Figure 8.4 **Reflectance curves for different colored surfaces.** (Based on Clulow (1972))

red, and yellow pills are preferred. Blue colors are described as cool whereas yellows tend to be described as warm. In fact, people turn a heat control to a higher setting in a blue room than in a yellow room. VR creators should be very cognizant of what colors they choose as arbitrary colors may result in unintended experiences.

8.1.4 Visual Acuity

Visual acuity is the ability to resolve details and is often measured in visual angle. A "rule of thumb" is that the thumb or a quarter viewed at arm's length subtends an angle of about 2° on the retina. A person with normal eyesight can see a quarter at 81 meters (nearly the length of a football field), which corresponds to 1 arc min (1/60th of a degree) (Coren et al., 1999). Under ideal conditions, we can see a line as thin as 0.5 arc sec (1/7200th of a degree) (Badcock et al., 2014)!

Factors

Several factors affect visual acuity. As can be seen in Figure 8.5, acuity dramatically falls off with eye eccentricity. Note how visual acuity matches quite well with the distribution of cones shown in Figure 8.1. We also have better visual acuity with

Meta Platforms, Inc.
Exhibit 1063
Page 118

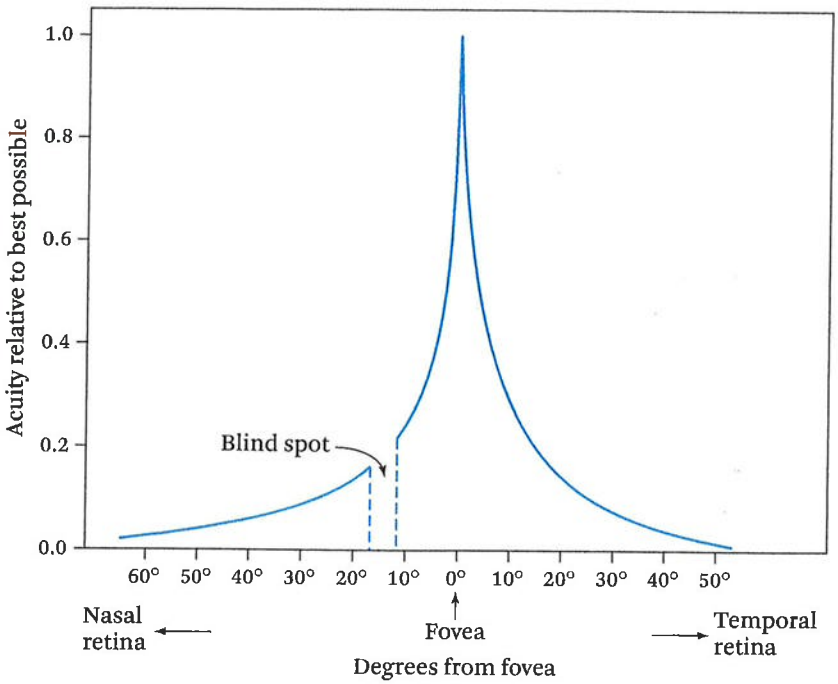


Figure 8.5 **Visual acuity is greatest at the fovea.** (Based on Coren et al., (1999))

illumination, high contrast, and long line segments. Visual acuity also depends upon the type of visual acuity that is being measured.

Types of Visual Acuity

There are different types of visual acuity: detection acuity, separation acuity, grating acuity, vernier acuity, recognition acuity, and stereoscopic acuity. Figure 8.6 shows some acuity targets used for measuring some of these visual acuities.

Detection acuity (also known as visible acuity) is the smallest stimulus that one can detect in an otherwise empty field and represents the absolute threshold of vision. Under ideal conditions a person can see a line as thin as 0.5 arc sec (1/7200° or 0.00014°) (Badcock et al., 2014). Increasing the target size up to a point is equivalent to increasing its relative intensity. This is because the mechanism of detection is contrast, and for small objects, minimum visible acuity does not actually depend on the width of the line that cannot be discerned. As the line gets thinner, it appears to get fainter but not thinner.

Meta Platforms, Inc.
Exhibit 1063
Page 119

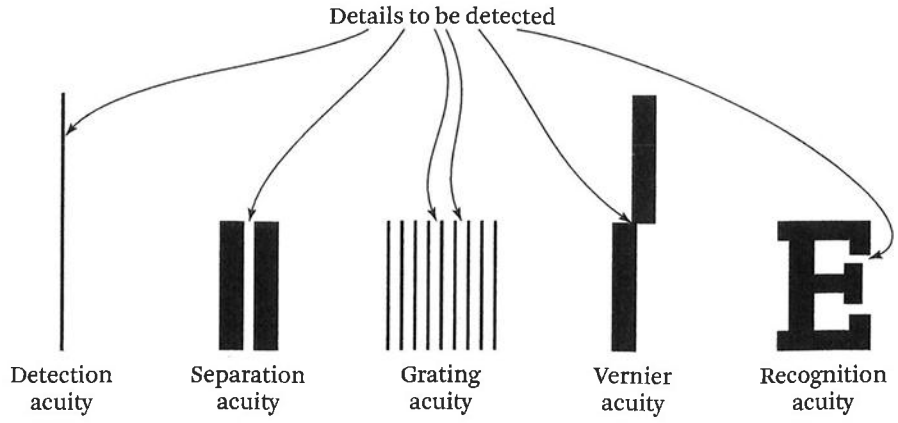


Figure 8.6 **Typical acuity targets for different methods of measuring visual acuity.** (Based on Coren et al., 1999))

Separation acuity (also known as resolvable acuity or resolution acuity) is the smallest angular separation between neighboring stimuli that can be resolved, i.e., the two stimuli are perceived as two. More than 5,000 years ago, Egyptians measured separation acuity by the ability to resolve double stars (Goldstein, 2010). Today, minimum separation acuity is measured by the ability to resolve two black stripes separated by white space. Using this method, observers are able to see the separation of two lines with a cycle of 1 arc min.

Grating acuity is the ability to distinguish the elements of a fine grating composed of alternating dark and light stripes or squares. Grating acuity is similar to separation acuity, but thresholds are lower. Under ideal conditions, grating acuity is 30 arc sec (Reichelt et al., 2010).

Vernier acuity is the ability to perceive the misalignment of two line segments. Under optimal conditions vernier acuity is very good at 1–2 arc sec (Badcock et al., 2014).

Recognition acuity is the ability to recognize simple shapes or symbols such as letters. The most common method of measuring recognition acuity is to use the Snellen eye chart. The observer views the chart from a distance and is asked to identify letters on the chart, and the smallest recognizable letters determine acuity. The smallest letter on the chart is about 5 arc min at 6 m (20 ft); this also happens to be the size of the average newsprint at a normal viewing distance. For the Snellen eye chart, acuity is measured relative to the performance of a normal observer. An acuity of 6/6 (20/20) means the observer is able to recognize letters at a distance of 6 m (20 ft) that a nor-

Meta Platforms, Inc.
Exhibit 1063
Page 120

mal observer is able to recognize. An acuity of 6/9 (20/30) means the observer is able recognize letters at 6 m (20 ft) that a normal observer can recognize at 9 m (30 ft), i.e., the observer cannot see as well.

Stereoscopic acuity is the ability to detect small differences in depth due to the disparity between left/right images (Section 9.1.3). For complex stimuli, stereoscopic acuity is similar to monocular visual acuity. Under well-lit conditions, stereo acuity of 10 arc sec is assumed to be a reasonable value (Reichelt et al., 2010). For simpler targets such as vertical rods, stereoscopic acuity can be as good as 2 arc sec. Factors that affect stereoscopic acuity (Reichelt et al., 2010) include spatial frequency, location on the retina (e.g., at an eccentricity of 2°, stereoscopic acuity decreases to 30 arc sec, and observation time (fast-changing scenes result in reduced acuity).

Stereopsis can result in better visual acuity than monocular vision. In one recent experiment, disparity-based depth perception was determined to occur beyond 40 meters when no other cues were present (Badcock et al., 2014). Interestingly, the contribution of stereopsis to depth was considerably greater than that of motion parallax that occurred by moving the head.

8.1.5 Eye Movements

Six extraocular muscles control rotations of each eye around three axes. Eye movements can be classifed in several different ways. They are categorized here as gaze-shifting, fixational, and gaze-stabilizing eye movements.

Gaze-Shifting Eye Movements

Gaze-shifting eye movements enable people to track moving objects or look at different objects.

Pursuit is the voluntary tracking with the eye of a visual target. The purpose of pursuit is to stabilize a target on the fovea, in order to provide maximum resolution, and to prevent motion blur, as the brain is too slow to completely process all details of foveal images moving faster than a few degrees per second.

Saccades are fast voluntary or involuntary movements of the eye that allow different parts of the scene to fall on the fovea and are important for visual scanning (Section 10.3.2). Saccades are the fastest moving external part of the body with speeds up to 1,000°/s and are ballistic where once started the destination cannot be changed (Bridgeman et al., 1994). Durations are about 20–100 ms, with most being about 50 ms in duration (Hallett, 1986), and amplitudes are up to 70°. Saccades typically occur at about three times per second.

Meta Platforms, Inc.
Exhibit 1063
Page 121

Saccadic suppression greatly reduces vision during and just before saccades, effectively blinding observers. Although observers do not consciously notice this loss of vision, events can occur during these saccades and observers will not notice. A scene can rotate by 8–20% of eye rotation during these saccades without observers noticing (Wallach, 1987). If eye tracking were available in VR, then the system could perhaps (for the purposes of redirected walking; Section 29.3.1) move the scene during saccades without users perceiving motion.

Vergence is the simultaneous rotation of both eyes in opposite directions in order to obtain or maintain binocular vision for objects at different depths. **Convergence** (the root con means "toward") rotates the eyes toward each other. **Divergence** (the root di means "apart") rotates the eyes away from each other in order to look further into the distance.

Fixational Eye Movements

Fixational eye movements enable people to maintain vision when holding the head still and looking in a single direction. These small movements keep rods and cones from becoming bleached. Humans do not consciously notice these small and involuntary eye movements, but without such eye movements the visual scene would fade into nothingness.

Small and quick movements of the eyes can be classified as microtremors (less than 1 arc min at 30–100 Hz) and microsaccades (about 5 arc min at variable rates) (Hallett, 1986).

Ocular drift is slow movement of the eye, and the eye may drift as much as a degree without the observer noticing (May and Badcock, 2002). Involuntary drifts during attempts at steady fixation have a median extent of 2.5 arc min and have a speed of about 4 arc min per second (Hallett, 1986). In the dark, the drift rate is faster. Ocular drift plays an important part in the autokinetic illusion described in Section 6.2.7. As discussed there, this may play an important role in judgments of position constancy (Section 10.1.3).

Gaze-Stabilizing Eye Movements

Gaze-stabilizing eye movements enable people to see objects clearly even as their heads move. Gaze-stabilizing eye movements play a role in position constancy adaptation (Section 10.2.2), and the eye movement theory of motion sickness (Section 12.3.5) states problems with gaze-stabilizing eye movements can cause motion sickness.

Retinal image slip is movement of the retina relative to a visual stimulus being viewed (Stoffregen et al., 2002). The greatest potential source of retinal image slip is due to rotation of the head (Robinson, 1981). Two mechanisms work together

Meta Platforms, Inc.
Exhibit 1063
Page 122

to stabilize gaze direction as the head moves—the vestibulo-ocular reflex and the optokinetic reflex.

Vestibulo-ocular reflex. The **vestibulo-ocular reflex** (VOR) rotates the eyes as a function of vestibular input and occurs even in the dark with no visual stimuli. Eye rotations due to the VOR can reach smooth speeds up to 500°/s (Hallett, 1986). This fast reflex (4–14 ms from the onset of head motion) serves to keep optical flow on the retina in the low-frequency range to which the retina is sensitive—i.e., it enables stable vision during rotation of the head (Razzaque, 2005). There are also proprioceptive eye-motion reflexes similar to VOR that use the neck, trunk, and even leg motion to help stabilize the eyes.

Optokinetic reflex. The **optokinetic reflex** (OKR) stabilizes retinal gaze direction as a function of visual input from the entire retina. If uniform motion of the visual scene occurs on the retina, then the eyes reflexively rotate to compensate. Eye rotations due to OKR can reach smooth speeds up to 80°/s (Hallett, 1986).

Eye rotation gain. **Eye rotation gain** is the ratio of the velocity of eye rotation divided by the velocity of head rotation (Hallett, 1986; Draper, 1998). A gain of 1.0 means that as the head rotates to the right the eyes rotate by an equal amount to the left so that the eyes are looking in the same direction as they were at the start of the head turn. Gain due to the VOR alone (i.e., in the dark) is approximately 0.7. If the observer imagines a stable target in the world, gain increases to 0.95. If the observer imagines a target that turns with the head, then gain is suppressed to 0.2 or lower. Thus, VOR is not perfect, and OKR corrects for residual error. VOR is most effective at 1–7 Hz (e.g., VOR helps maintain fixation on objects while walking or running) and progressively less effective at lower frequencies, particularly below 0.1 Hz. OKR is most effective at frequencies below 0.1 Hz and has decreasing effectiveness at 0.1–1 Hz. Thus, the VOR and OKR complement each other for typical head motions—both VOR and OKR working together result in a gain close to 1 over a wide range of head motions.

Gain also depends on the distance to the visual target being viewed. For an object at an infinite distance, the eyes are looking straight ahead in parallel. In this case, gain is ideally equal to 1.0 so that the target image remains on the fovea. For a closer target, eye rotation must be greater than head rotation for the image to remain on the fovea. These differences in gains are because rotational axis for the head is different than the rotational axis of the eyes. The differences in gain can quickly be demonstrated by comparing eye movements while looking at a finger held in front of the eyes versus an object further in the distance.

Meta Platforms, Inc.
Exhibit 1063
Page 123

Active vs. passive head and eye motion. Active head rotations by the observer result in more robust and more consistent VOR gains and phase lag than passive rotations by an external force (e.g., being automatically rotated in a chair) (Draper, 1998). Similarily, sensitivity to motion of a visual scene while the head moves depends upon whether the head is actively moved or passively moved (Howard, 1986a; Draper, 1998). Afference and efference help explain motion perception by taking into account active eye movements. See Section 7.4 for an explanation of afference and efference, and an explanation of how the world appears stable even as the eyes move.

Nystagmus. Nystagmus is a rhythmic and involuntary rotation of the eyes (Howard, 1986a) caused by both VOR and OKR (Razzaque, 2005) that can help stabilize gaze. Researchers typically discuss nystagmus caused by rotating the observer continuously at a constant angular velocity. The slow phase of nystagmus is the rotation of the eyes to keep the observer looking straight ahead in the world as the chair rotates. As the eyes reach their maximum amount of rotation in the eye sockets, a saccade snaps the eyes back to looking straight ahead relative to the head. This rotation is called the fast phase of nystagmus. This pattern repeats, resulting in a rhythmic rotation of the eyes. Nystagmus can also be demonstrated by looking at someone's eyes after they have become dizzy from spinning that stimulates their vestibular system (Coren et al., 1999).

Pendular nystagmus occurs when one rotates her head back and forth at a fixed frequency. This results in an always-changing slow phase with no fast phase. Pendular nystagmus occurs when walking and running. Those looking for latency in HMD often quickly rotate their head back and forth to try to estimate latency by seeing how the scene moves as they rotate their head. Little is known about how nystagmus works when perceiving such latency-induced scene motion in an HMD (see Chapter 15). A user's eye gaze might remain stationary in space (because of the VOR) when looking for scene motion, resulting in retinal image slip, or might follow the scene (because of OKR), resulting in no retinal image slip. This likely varies by the amount of head motion, the type of task, the subject, etc. Eye tracking would allow an investigator to study this in detail. Subjects in one latency experiment (Ellis et al., 2004) claimed to concentrate on a single feature in the moving visual field. However, this was anecdotal evidence and was not verified.

8.1.6 Visual Displays

Given that each eye can see about 210° by rotating the eye and assuming vernier or stereoscopic acuity of 2 arc sec, then a display would require horizontal resolution of 378,000 pixels for each eye to match what we can see in reality! This number is a simple

Meta Platforms, Inc.
Exhibit 1063
Page 124

extreme analysis as these are for perfectly ideal situations and we can design around limitations for different circumstances. For example, since we do not perceive such high resolution outside of central vision, eye tracking could be used to only render at high resolution where the user is looking. This would require new algorithms, non-uniform display hardware, and fast eye tracking. Clearly, there is plenty of work to be done to get to the point of truly simulating visual reality. Latency will be even more of a concern in this case, as the system will be racing the eye to display at high resolution by the time the user is looking at a specific location.

8.2 Hearing

Auditory perception is quite complex and is affected by head pose, physiology, expectation, and its relationship to other sensory modality cues. We can deduce qualities of the environment from sound (e.g., large rooms sound different than small rooms), and we can determine where an object is located by its sound alone. Below are discussed general concepts of sound, and Section 21.3 discusses audio as it applies specifically to VR.

8.2.1 Properties of Sound

Sound can be broken down into two distinct parts: the physical aspects and the perceptual aspects.

Physical Aspects

A person need not be present for a tree falling in the woods to make a sound. **Physical sound** is a pressure wave in a material medium (i.e., air) created when an object vibrates rapidly back and forth. **Sound frequency** is the number of cycles per second—hertz (Hz)—or vibrations that a change in pressure repeats. **Sound amplitude** is the difference in pressure between the high and low peaks of the sound wave. The decibel (dB) is a logarithmic transformation of sound amplitude where doubling the sound amplitude results in an increase of 3 dB.

Perceptual Aspects

Physical sound enters the ear, the eardrum is stimulated, and then receptor cells transduce those sound vibrations into electrical signals. The brain then processes these electrical signals into qualities of sound such as loudness, pitch, and timbre.

Loudness is the perceptual quality most closely related to the amplitude of a sound, although frequency can affect loudness. A 10 dB increase (a bit over three doublings of amplitude) results in approximately twice the subjective loudness.

Meta Platforms, Inc.
Exhibit 1063
Page 125

Pitch is most closely related to the physical property of fundamental frequency. Low frequencies are associated with low pitches and high frequencies are associated with high pitches. Most real-world sounds are not strictly periodic, as they do not have repeated temporal patterns but have fluctuations from one cycle to the next. The perceived pitch of such imperfect sound patterns is the average period of the cyclical variations. Other sounds such as noise are non-periodic and do not have a perceived pitch.

Timbre is closely related to the harmonic structure (both strength of the harmonics and number of harmonics) of a sound and is the quality that distinguishes between two tones that have the same loudness, pitch, and duration, but sound different. For example, a guitar has more high-frequency harmonics than either the bassoon or alto saxophone (Goldstein, 2014). Timbre also depends on the attack (the buildup of sound at the beginning of the tone) and the tone's decay (the decrease in sound at the end of the tone). Playing a recording of a piano backwards sounds more like an organ because the original decay has become the attack and the attack has become the decay.

Auditory Thresholds

Humans can hear frequencies from about 20 to 22,000 Hz (Vorlander and Shinn-Cunningham, 2014) with the most sensitivity occurring at about 2,000–4,000 Hz, which is the range of frequencies that is most important for understanding speech (Goldstein, 2014). We hear sound in a wide range of about a factor of a trillion (120 dB) before the onset of pain begins. Most sounds encountered in everyday experience span a dynamic intensity range of 80–90 dB. Figure 8.7 shows what amplitudes and frequencies we can hear and how perceived volume is significantly dependent upon frequency—lower- and higher-frequency sounds require greater amplitude than midrange frequencies for us to perceive equal loudness. This is largely due to the outer ear's auditory canal reinforcing/amplifying midrange frequencies. Interestingly, we only perceive musical melodies for pitches below 5,000 Hz (Attneave and Olson, 1971).

The auditory channel is much more sensitive to temporal variations than both vision and proprioception. For example, amplitude fluctuations can be detected at 50 Hz (Yost, 2006) and temporal fluctuations can be detected at up to 1,000 Hz. Not only can listeners detect rapid fluctuations but they can react quickly. Reaction times to auditory stimuli are faster than visual reaction times by 30–40 ms (Welch and Warren, 1986).

8.2.2 Binaural Cues

Binaural cues (also known as stereophonic cues) are two different audio cues, one for each ear, that help to determine the position of sounds. Each ear hears a slightly different sound—different in time and different in level. Interaural time differences provide

Meta Platforms, Inc.
Exhibit 1063
Page 126

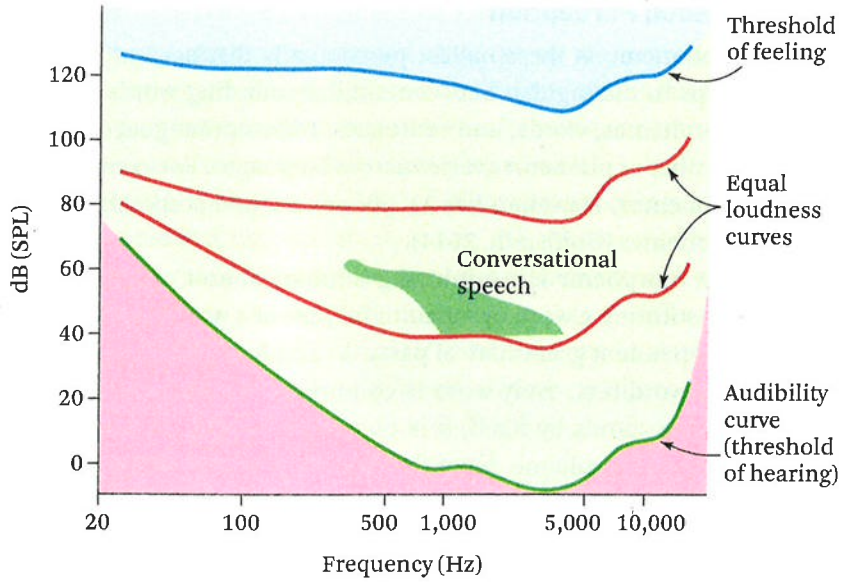


Figure 8.7 **The audibility curve and auditory response area. The area above the audibility curve represents volume and frequencies that we can hear. The area above the threshold of feeling can result in pain.** (Based on Goldstein, (2014))

an effective cue for localizing low-frequency sounds. Interaural time differences as small as ~10 ms can be discerned (Klumpp, 1956). Interaural level differences (called acoustic shadows) occur due to acoustic energy being reflected and diffracted by the head, providing an effective cue for localizing sounds above 2 kHz (Bowman et al., 2004).

Monaural cues use differences in the distribution (or spectrum) of frequencies entering the ear (due to the shape of the ear) to help determine the position of sounds. This monaural cue is helpful for determining the elevation (up/down) direction of a sound where binaural cues are not helpful. Head motion also helps determine where a sound is coming from due to changes in interaural time differences, interaural level differences, and monaural spectral cues. See Section 21.3.1 for a discussion of head-related transfer functions.

Spatial acuity of the auditory system is not nearly as good as vision. We can detect differences of sound at about 1° in front of or behind us, but our sensitivity decreases to 10° when the sound is to the far left/right side of us and 15° when the sound is above/below us. Vision can play a role in locating sounds—see Section 8.7.

Meta Platforms, Inc.
Exhibit 1063
Page 127

8.2.3 Speech Perception

A **phoneme** is the smallest perceptually distinct unit of sound in a language that helps to distinguish between similar-sounding words. Phonemes combine to form morphemes, words, and sentences. Different languages use different sounds, so the number of phonemes varies across languages. For example, American English has 47 phonemes, Hawaiian has 11 phonemes, and some African languages have up to 60 phonemes (Goldstein, 2014).

A **morpheme** is a minimal grammatical unit of a language, with each morpheme constituting a word or meaningful part of a word that cannot be divided into smaller independent grammatical parts. A morpheme can be, but is not necessarily, a subset of a word (i.e., every word is comprised of one or more morphemes). When a morpheme stands by itself, it is considered a root because it has meaning on its own. When a morpheme depends on another morpheme to express an idea, it is an affix because it has a grammatical function (e.g., adding an *s* to the end of a word to make it plural). The more ways morphemes can be combined with other morphemes, the more productive that morpheme is. Phoneme and morpheme awareness is the ability to identify speech sounds, the vocal gestures from which words are constructed, when they are found in their natural context—spoken words.

Speech segmentation is the perception of individual words in a conversation even when the acoustic signal is continuous. Our perception of words is not solely based upon energy stimulating our receptors, but is heavily influenced by our experience with those sounds and relationships between those sounds. To someone listening to an unfamiliar foreign language, words seem to speed by in a single stream of sound.

It is easier to perceive phonemes in a meaningful context. Warren (1970) had test subjects listen to the sentence "The state governors met with their respective legislatures convening in the capital city." When the first *s* in "legislatures" was replaced with a cough sound, none of the subjects were able to state where in the sentence the cough occurred or that the *s* in "legislatures" was missing. This is called the **phonemic restoration effect.**

Samuel (1981) used the phonemic restoration effect to show that speech perception is determined both by the acoustic signal (bottom-up processing) and by context that produces expectations in the listener (top-down processing). Samuel also found longer words increase the likelihood of the phonemic restoration effect. A similar effect occurs for meaningfulness of spoken words in a sentence and knowledge of the rules of grammar; it is easier to perceive spoken words when heard in the context of familiar grammatical sentences (Miller and Isard, 1963).

Meta Platforms, Inc.
Exhibit 1063
Page 128

8.3 Touch

When we touch something or we are touched, receptors in the skin provide information about what is happening to our skin and about the object contacting the skin. These receptors enable us to perceive information about small details, vibration, textures, shapes, and potentially damaging stimuli.

Although those who are deaf or blind can get along surprisingly well, those with a rare condition that results in losing the sensation of touch often suffer constant bruising, burns, and broken bones due to the absence of warnings provided by touch and pain. Losing the sense of touch also makes it difficult to interact with the environment. Without the feedback of touch, actions as simple as picking objects up or typing on a keyboard can be difficult. Unfortunately, touch is extremely challenging to implement in VR but by understanding how we perceive touch, we can at least take advantage of some simple cues to users.

Adults have 1.3–1.7 square m (14–18 sq ft) of skin. However, the brain does not consider all skin to be created equally—humans are very dependent on speech and manipulation of objects through the use of the hands, thus we have large amounts of our brain devoted to the hands. The right side of Figure 8.8 shows the **sensory homunculus** ("little man" in Latin), which represents the location and proportion of sensory cortex devoted to different body parts. The proportion of sensory cortex devoted to each body part also correlates with the density of tactile receptors on that

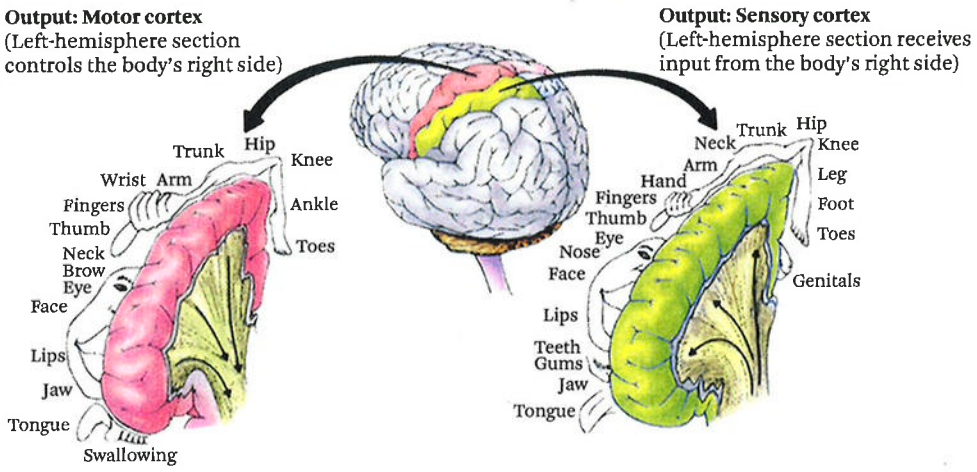


Figure 8.8 **This motor and sensory homunculus represents "the body within the brain." The size of a body part in the diagram represents the amount of cerebral cortex devoted to that body part.**

Meta Platforms, Inc.
Exhibit 1063
Page 129

body part. The left side shows the motor homunculus that helps to plan and execute movements. Some areas of the body are represented by a disproportionately large area of the sensory and motor cortexes, notably the lips, tongue, and hands.

8.3.1 Vibration

The skin is capable of detecting not only spatial details but vibrations as well. This is due to mechanoreceptors called Pacinian corpuscles. Nerve fibers located within corpuscles respond slowly to slow or constant pushing, but respond well to high vibrational frequencies.

8.3.2 Texture

Depending on vision for perceiving texture is not always sufficient, because seeing that texture is dependent on lighting. We can also perceive texture through touch. Spatial cues are provided by surface elements such as bumps and grooves, resulting in feelings of shape, size, and distribution of surface elements.

Temporal cues occur when the skin moves across a textured surface, and occur as a form of vibration. Fine textures are often only felt with a finger moving across the surface. Temporal cues are also important for feeling surfaces indirectly through the use of tools (e.g., dragging a stick across a rough surface); this is typically felt not as vibration of the tool but texture of the surface even though the finger is not touching the texture.

Most VR creators do not consider physical textures. Where they are considered is with passive haptics (Section 3.2.3), such as when building physical devices (e.g., hand-held controllers or a steering system) and real-world surfaces that users interact with while immersed.

8.3.3 Passive Touch vs. Active Touch

Passive touch occurs when stimuli are applied to the skin. Passive touch can be quite compelling in VR when combined with visuals. For example, rubbing a real feather on the skin of users when they see a virtual feather rubbing their physical body is used quite effectively to embody users into their avatars (Section 4.3).

Active touch occurs when a person actively explores an object, usually with the fingers and hands. Note passive and active touch is not to be confused with passive and active haptics as discussed in Section 3.2.3. Humans use three distinct systems together when using active touch.

- The sensory system, used in detecting cutaneous sensations such as touch, temperature, textures, and positions/movements of the fingers.

Meta Platforms, Inc.
Exhibit 1063
Page 130

- The motor system, used in moving the fingers and hands.
- The cognitive system, used in thinking about the information provided by the sensory and motor systems.

These systems work together to create an experience that is quite different from passive touch. For passive touch, the sensation is typically experienced on the skin, whereas for active touch, we perceive the object being touched.

8.3.4 Pain

Pain functions to warn us of dangerous situations. Three types of pain are

- neuropathic pain caused by lesions, repetitive tasks (e.g., carpal tunnel syndrome), or damage to the nervous system (e.g., spinal cord injury or stroke);
- nociceptive pain caused by activation of receptors in the skin called nociceptors, which are specialized to respond to tissue damage or potential damage from heat, chemicals, pressure, and cold; and
- inflammatory pain caused by previous damage to tissue, inflammation of joints, or tumor cells.

The perception of pain is strongly affected by factors other than just stimulation of skin, such as expectation, attention, distracting stimuli, and hypnotic suggestion. One specific example is phantom limb pain, where individuals who have had a limb amputated continue to experience the limb, and in some cases continue to experience pain in a limb that does not exist (Ramachandran and Hirstein, 1998). An example of reducing pain using VR is Snow World, a distracting VR game set in a cold-conveying environment that doctors use when they have burn victims' bandages removed (Hoffman, 2004).

8.4 Proprioception

Proprioception is the sensation of limb and whole body position and movement derived from the receptors of muscles, tendons, and joint capsules. Proprioception enables us to touch our nose with a hand even when the eyes are closed. Proprioception includes both conscious and subconscious components, not only enabling us to sense the position and motion of our limbs, but also providing us the sensation of force generation enabling us to regulate force output.

Whereas touch seems straightforward because we are often aware of the sensations that result, proprioception is more mysterious because we are largely unaware of it and take it for granted during daily living. However, as VR creators, becoming familiar

Meta Platforms, Inc.
Exhibit 1063
Page 131

with the sense of proprioception is important for understanding how users physically move to interact with a virtual environment (at least until directly connected neural interfaces become common). This includes moving the head, eyes, limbs, and/or whole body. Without the senses of touch and proprioception, we would crush brittle objects when picking them up. Proprioception can be very useful for designing VR interactions as discussed in Section 26.2.

8.5 Balance and Physical Motion

The **vestibular system** consists of labyrinths in the inner ears that act as mechanical motion detectors (Figure 8.9), which provide input for balance and sensing physical motion. The vestibular organs are composed of the otolith organs and the semicircular canals.

Each set (right and left) of two **otolith organs** acts as a three-axis accelerometer, measuring linear acceleration. Cessation of linear motion is sensed almost immedi-

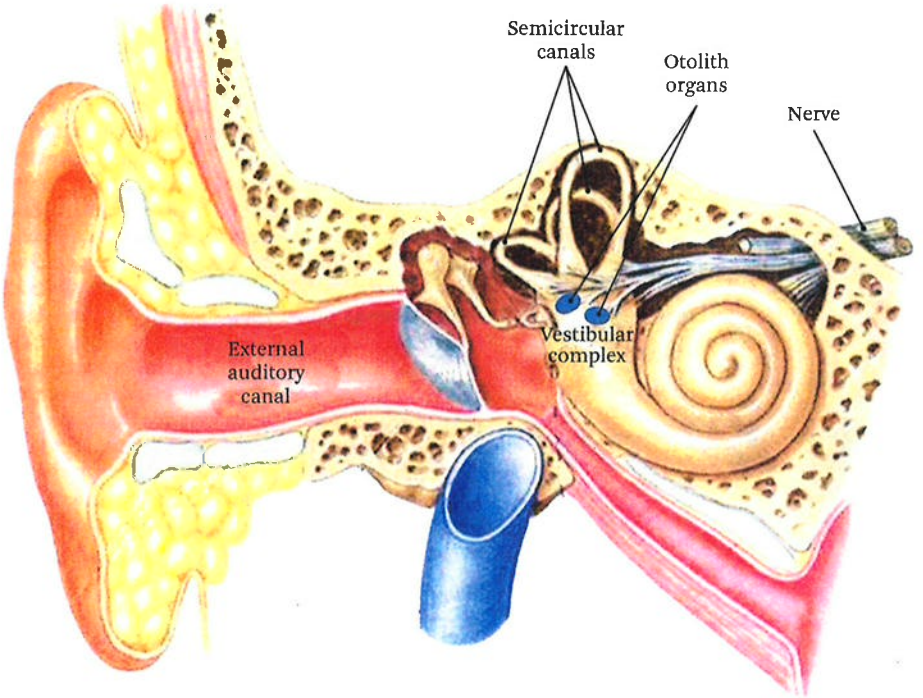


Figure 8.9 **A cutaway illustration of the outer, middle, and inner ear, revealing the vestibular system.** (Jerald (2009), adapted from Martini (1998))

Meta Platforms, Inc.
Exhibit 1063
Page 132

ately by the otolith organs (Howard, 1986b). The nervous system's interpretation of signals from the otolith organs relies almost entirely on the direction and not the magnitude of acceleration (Razzaque, 2005).

Each set of the three nearly orthogonal **semicircular canals** (SCCs) acts as a three-axis gyroscope. The SCCs act primarily as gyroscopes measuring angular velocity, but only for a limited amount of time if velocity is constant. After 3–30 seconds the SCCs cannot disambiguate between no velocity and some constant velocity (Razzaque, 2005). In real and virtual worlds, head angular velocity is almost never constant other than zero velocity. The SCCs are most sensitive between 0.1 Hz and 5.0 Hz. Below 0.1 Hz, SCC output is roughly equal to angular acceleration, 0.1–5.0 Hz roughly equal to angular velocity, and above 5.0 Hz roughly equal to angular displacement (Howard, 1986b). These ranges fit well within typical head motions—head motions while walking (at least one foot always touching the ground) are in the 1–2 Hz range and 3–6 Hz while running (moments when neither foot is touching the ground) (Draper, 1998; Razzaque, 2005).

Although we are not typically aware of the vestibular system in normal situations, we can become very aware of this sense in atypical situations when things go wrong. Understanding the vestibular system is extremely important for creating VR content as motion sickness can result when vestibular stimuli does not match stimuli from the other senses. The vestibular system and its relationship to motion sickness are discussed in more detail in Chapter 12.

8.6 Smell and Taste

Smell and taste both work through chemoreceptors, where the chemoreceptors detect chemical stimuli in the environment.

Smell (also known as olfactory perception) is the ability to perceive odors when odorant airborne molecules bind to specific sites on the olfactory receptors high in the nose. Very low concentrations of material can be detected by the olfactory system. The number of smells the olfactory system can distinguish between is highly debated with different researchers claiming a wide range from hundreds to trillions. Different people smell different odors due to genetic differences.

Taste (also known as gustatory perception) is the chemosensory sensation of substances on the tongue. The human tongue by itself can distinguish among five universal tastes: sweet, sour, bitter, salty, and umami.

Smell, taste, temperature, and texture all combine together to provide the sensation of flavor. Combining these senses provides a much wider range of flavor than if taste alone determined flavor. In fact, a pseudo-gustatory VR system has been shown to fool

Meta Platforms, Inc.
Exhibit 1063
Page 133

users into thinking a plain cookie tasted differently by providing different visual and smell cues (Miyaura et al., 2011).

8.7 Multimodal Perceptions

Integration of our different senses occurs automatically and rarely does perception occur as a function of a single modality. Examining each of the senses as if it were independent of the others leads to only partial understanding of everyday perceptual experience (Welch and Warren, 1986). Sherrington (1920) states, "All parts of the nervous system are connected together and no part of it is probably ever capable of action without affecting and being affected by various other parts."

Perception of a single modality can influence other modalities. A surprising example is if audio cues are not precisely synchronized with visual cues, then visual perception may be affected. For example, auditory stimulation can influence the visual flicker-fusion frequency (the frequency at which subjects start to notice the flashing of a visual stimulus; see Section 9.2.4) (Welch and Warren, 1986).

Sekuler et al. (1997) found that when two moving shapes cross on a display, most subjects perceived these shapes as moving past each other and continuing their straight-line motion. When a "click" sounded just when the shapes appeared adjacent to each other, a majority of people perceived the shapes as colliding and bouncing off in opposite directions. This is congruent with what typically happens in the real world—when a sound occurs just as two moving objects become close to each other, a collision has likely occurred that caused that sound.

Perceiving speech is often a multisensory experience involving both audition and vision. **Lip sync** refers to the synchronization between the visual movement of the speaker's lips and the spoken voice. Thresholds for perceiving speech synchronization vary depending on the complexity, congruency, and predictability of the audiovisual event as well as the context and applied experimental methodology (Eg and Behne, 2013). In general, we are more tolerant of visuals leading sound and thresholds increase with sound complexity (e.g., we notice asynchronies less for sentences than for syllables). One study found that when audio precedes video by 5 video fields (about 80 ms), viewers evaluated speakers more negatively (e.g., less interesting, more unpleasant, less influential, more agitated, and less successful) even when they did not consciously identify the asynchrony (Reeves and Voelker, 1993).

The McGurk effect (McGurk and MacDonald, 1976) illustrates how visual information can exert a strong influence on what we hear. For example, when a listener hears the sound /ba-ba/ but is viewing a person making lip movements for the sound /ga-ga/, the listener hears the sound /da-da/.

Meta Platforms, Inc.
Exhibit 1063
Page 134

Vision tends to dominate other sensory modalities (Posner, Nissen, and Klein, 1976). For example, vision dominates spatialized audio. **Visual capture** or the ventriloquism effect occurs when sounds coming from one location (e.g., the speakers in a movie theater) are mislocalized to seem to come from a place of visual motion (e.g., an actor's mouth on the screen). This occurs due to the tendency to try to identify visual events or objects that could be causing the sound.

In many cases, when vision and proprioception disagree, we tend to perceive hand position to be where vision tells us it is (Gibson, 1933). Under certain conditions, VR users can be made to believe that their hand is touching a different seen shape than a felt shape (Kohli, 2013). Under other conditions, VR users are more sensitive to their virtual hand visually penetrating a virtual object than they are to the proprioceptive sense of their hand not being colocated with their visual hand in space (Burns et al., 2006). Sensory substitution can be used to partially make up for the lack of physical touch in VR systems as discussed in Section 26.8.

Mismatch of visual and vestibular cues is a major problem of VR as it can cause motion sickness. Chapter 12 discusses such cue conflicts in detail.

8.7.1 Visual and Vestibular Cues Complement Each Other

The vestibular system provides primarily first-order approximations of angular velocity and linear acceleration, and positional drift occurs over time (Howard, 1986b). Thus, absolute position or orientation cannot be determined from vestibular cues alone. Visual and vestibular cues combine to enable people to disambiguate between moving stimuli and self-motion.

The visual system is good at capturing lower-frequency motions whereas the vestibular system is better at detecting high-frequency motions. The vestibular system is a mechanical system and has a faster response (as fast as 3–5 ms!) than the slower electrochemical visual response of the eyes. Thus, the vestibular system is initially more responsive to a sudden onset of physical motion. After some time of sustained constant velocity, vestibular cues subside and visual cues take over. Mid-range frequencies use both vestibular and visual cues.

Missing or misleading vestibular cues can lead to life-threatening motion illusions in pilots (Razzaque, 2005). For example, the vestibular otolith organs cannot distinguish between linear acceleration and tilt. This ambiguity can be resolved from the semicircular canals when within the canals' sensitive range. However, when not within the sensitive range, the ambiguity is resolved from visual cues. Aircraft accidents, and loss of many lives, have occurred due to the ambiguity when visual cues are missing (e.g., low visibility conditions). This ambiguity can be used to our advantage in VR when motion platforms are available by tilting the platform to give a sense of forward acceleration that can be extremely compelling.

Meta Platforms, Inc.
Exhibit 1063
Page 135

Perception of Space and Time

How do we perceive the layout, timing, and motion of objects in the world around us? This question is not new, as it has fascinated artists, philosophers, and psychologists for centuries. The success of painting, photography, cinema, video games, and now VR is largely contingent upon the convincing portrayal of spatial 3D relationships and their timings.

9.1 Space Perception

Trompe-l'œil art (Figure II.1) and the Ames room (Figure 6.9) prove the eye can be fooled when it comes to space perception. However, these illusions are very constrained and only work when viewed from a specific viewpoint—when the observer moves and explores resulting in the perception of more information, then the illusion goes away and the truth becomes more apparent. It is relatively easy to fool observers from a single perspective, compared to fooling VR users into believing they are present in another world.

If the senses can be deceived to misinterpret space, then how can we be sure the senses do not deceive us all the time? Perhaps we cannot be 100% sure of the space around us. However, evolution tells us that our perceptions must be truthful to some degree—otherwise if our perceptions were overly deceived, we would have become extinct long ago (Cutting and Vishton, 1995). Perhaps deception of space is the goal of VR, and if we can someday completely overcome the challenges of VR, then we can defeat the argument from evolution, and hopefully not extinguish ourselves from existence in the process.

Multiple sensory modalities often work together to give us spatial location. Vision is the superior spatial modality as it is more precise, is faster for perceiving location (although audition is faster when the task is detection), is accurate for all three

Meta Platforms, Inc.
Exhibit 1063
Page 136

directions (whereas audition, for example, is reasonably good only in the lateral direction), and works well at long distances (Welch and Warren, 1986).

Perceiving the space around us is not just about passive perception but is important for actively moving through the world (Section 10.4.3) and interaction (Section 26.2).

9.1.1 Exocentric and Egocentric Judgments

Exocentric judgments, also known as object-relative judgments, are the sense of where objects are relative to other objects, the world, or some other external reference. Exocentric judgments include the sense of gravity, geographic direction (e.g., north), and distance between two objects. Gogel's principle of adjacency states "The effectiveness of relative cues decreases as the perceived distance between objects producing the relative cues increases" (Mack, 1986). That is, objects are more difficult to compare the further they are apart.

The perception of the location of objects also includes the direction and distance relative to our bodies. **Egocentric judgments**, also known as subject-relative judgments, are the sense of where (direction and distance) aspects in an environment are relative to oneself. Left/right and forward/backward judgments are egocentric. Egocentric judgments can be performed by the primary sensory modalities of vision, audition, proprioception, and touch.

Several variables affect our egocentric perception. The more stimuli available, the more stable our judgments; our judgments are much less stable in the dark as demonstrated by the autokinetic effect (Section 6.2.7). A stimulus that is not centered in the visual field (e.g., an offset square) can result in the straight-ahead being perceived as biased toward that off-centered stimulus. The **dominant eye**, the eye that is used for sighting tasks, also influences direction judgments. Non-visual factors, such as eye movement (Section 8.1.5), also affect our judgment of the straight-ahead.

Reference frames of exocentric and egocentric space as they relate to VR are discussed in Section 26.3. Such reference frames are especially important for designing VR interfaces.

9.1.2 Segmenting the Space around Us

The layout of the space around each of us can be divided into three circular egocentric regions: personal space, action space, and vista space (Cutting and Vishton, 1995).

Our **personal space** is generally considered to be the natural working volume within arm's reach and slightly beyond. We typically are only comfortable with others in this space in intimate situations or due to public necessity (e.g., a subway or concert). This space is within 2 meters of the person's viewpoint, which includes users' feet while standing and the space users can physically manipulate without requiring locomo-

Meta Platforms, Inc.
Exhibit 1063
Page 137

tion. Working within personal space offers advantages of proprioceptive cues, more direct mapping between hand motion and object motion, stronger stereopsis and head motion parallax cues, and finer angular precision of motion (Miné et al., 1997).

Beyond personal space, **action space** is the space of public action. In action space, we can move relatively quickly, speak to others, and toss objects. This space starts at about 2 meters and extends to about 20 meters from the user.

Beyond 20 meters is **vista space**, where we have little immediate control and perceptual cues are fairly consistent (e.g., binocular vision is almost non-existent). Because effective depth cues are pictorial at this range, large 2D trompe-l'œil paintings at this range are most effective in deceiving the eye that the space is real, as demonstrated in Figure 9.1 showing the ceiling of the Jesuit Church in Vienna painted by Andrea Pozzo.



Figure 9.1 **This trompe-l'œil art on the dome of the Jesuit Church in Vienna demonstrates that at a distance it can be difficult to distinguish 2D painting from 3D architecture.** (Courtesy of Paolo Gianti/Shutterstock.com)

Meta Platforms, Inc.
Exhibit 1063
Page 138

9.1.3 Depth Perception

This page you are reading is likely less than a meter away and the words and lettering are less than a centimeter in height. If you look further up you might see walls, windows, and various objects further than a meter. You can intuitively estimate these distances and interact with objects without thinking much about distance even though the projection of those objects onto the retinas of your eyes are 2D images. Based on these projected scenes of the world onto the surface of the retinas alone, we somehow are able to intuitively perceive how far away objects are.

How is it that we can take individual photons that fall onto the retina's surface and re-create from them an understanding of a 3D world? If spatial perception was solely a function of a 2D image on each retina (Section 8.1.1), then there would be no way to know if a particular visual stimulus was a foot away or a mile away. Clearly, there must be more going on than simply detecting points of light on the retina.

Higher-level processing occurs that results in the perception of a 3D world. Human perception relies on a large number of ways we perceive space, shapes, and distance. In this section, several cues are described that help us perceive egocentric depth or distance. The more cues available and the more consistent they are with each other, the stronger the perception of depth. If implemented well in a VR system, such depth cues can significantly add to presence.

Depth cues can be organized into four categories: pictorial cues, motion cues, binocular cues, and oculomotor cues. The context or psychological state of the user can also affect a person's judgment of depth/distance. Table 9.1 shows an organization of the cues and factors described in this section, and Table 9.2 shows the approximate order of importance for several of the specific cues.

Pictorial Depth Cues

Pictorial depth cues are projections of stimuli onto the retina, resulting in 2D images. These pictorial cues are present even when viewed with only a single eye and when that eye or anything in the view is not moving. Described below are pictorial cues listed in approximate declining order of importance.

Occlusion. **Occlusion** is the strongest depth cue as it is always the case for close opaque objects to hide other more distant objects and is effective over the entire range of perceivable distances. Although occlusion is an important depth cue, it only provides relative depth information, and thus precision is dependent upon the number of occluded objects in the scene.

Occlusion is an especially important concept when designing VR information displays. Many heads-up displays found in 2D video games do not work well with VR

Meta Platforms, Inc.
Exhibit 1063
Page 139

Table 9.1 **Organization of depth cues and factors presented in this chapter.**

- Pictorial Depth Cues
  - Occlusion
  - Linear perspective
  - Relative/familiar size
  - Shadows/shading
  - Texture gradient
  - Height relative to horizon
  - Aerial perspective
- Motion Depth Cues
  - Motion parallax
  - Kinetic depth effect
- Binocular Depth Cues
- Oculomotor Depth Cues
  - Vergence
  - Accommodation
- Contextual Distance Factors
  - Intended action
  - Fear

Table 9.2 **Approximate order of importance (with 1 being most important) of visual cues for perceiving egocentric distance in personal space, action space, and vista space.**

	Space		
Source of Information	Personal Space	Action Space	Vista Space
Occlusion	1	1	1
Binocular	2	8	9
Motion	3	7	6
Relative/familiar size	4	2	2
Shadows/shading	5	5	7
Texture gradient	6	6	5
Linear perspective	7	4	4
Oculomotor	8	10	10
Height relative to horizon	9	3	3
Aerial perspective	10	9	8

Meta Platforms, Inc.
Exhibit 1063
Page 140

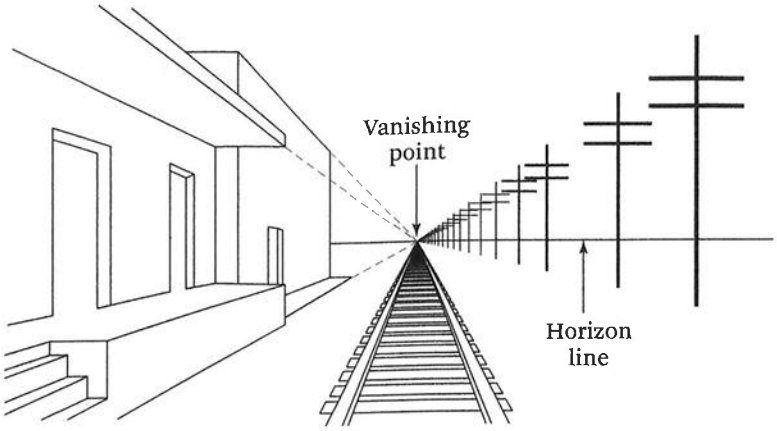


Figure 9.2 **Linear perspective causes parallel lines receding into the distance to appear to converge at the vanishing point.**

displays due to conflicts with occlusion, motion, physiological, and binocular cues. Because of these conflicts, 2D heads-up displays should be transformed into 3D geometry that has a distance from the viewpoint so that occlusion is properly handled. Otherwise users will get confused when the text or geometry appears at a distance but is not occluded by closer objects and can lead to eye strain and headaches (Section 13.2).

Linear perspective. **Linear perspective** causes parallel lines that recede into the distance to appear to converge together at a single point called the vanishing point (Figure 9.2). The Ponzo railroad illusion in Figure 6.8 demonstrates how strong of a cue linear perspective is. The sense of depth provided by linear perspective is so strong that non-realistic grid patterns may be more effective to conveying spatial presence than more common textures with little detail such as carpet or flat walls.

Relative/familiar size. When two objects are of equal size, the further object's projected image on the retina will take up less visual angle than the closer object. **Relative/familiar size** enables us to estimate distance when we know the size of an object, whether due to prior knowledge or comparison to other objects in the scene (Figure 9.3). Seeing a representation of one's own body not only adds presence (Section 4.3) but can also serve as a relative-size depth cue since users know the size of their own body.

Meta Platforms, Inc.
Exhibit 1063
Page 141



Figure 9.3 **Relative/familiar size. The visual angle of an object of known size provides a sense of distance to that object (far left). Two identical objects having different angular sizes implies the smaller object is at a further distance.**

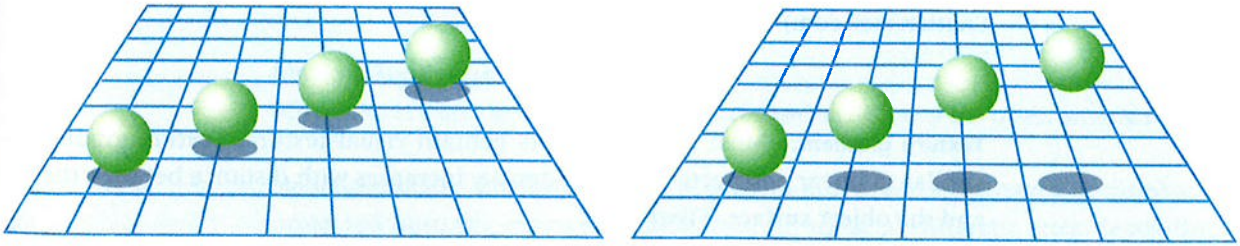


Figure 9.4 **Example of how objects can appear at different depth and heights depending on where their shadows lie.**

Shadows/shading. **Shadows and shading** can provide information regarding the location of objects. Shadows can immediately tell us if an object is lying on the ground or floating above the ground (Figure 9.4). Shadows can also give us clues about the shape (and different depths of a single object).

Meta Platforms, Inc.
Exhibit 1063
Page 142

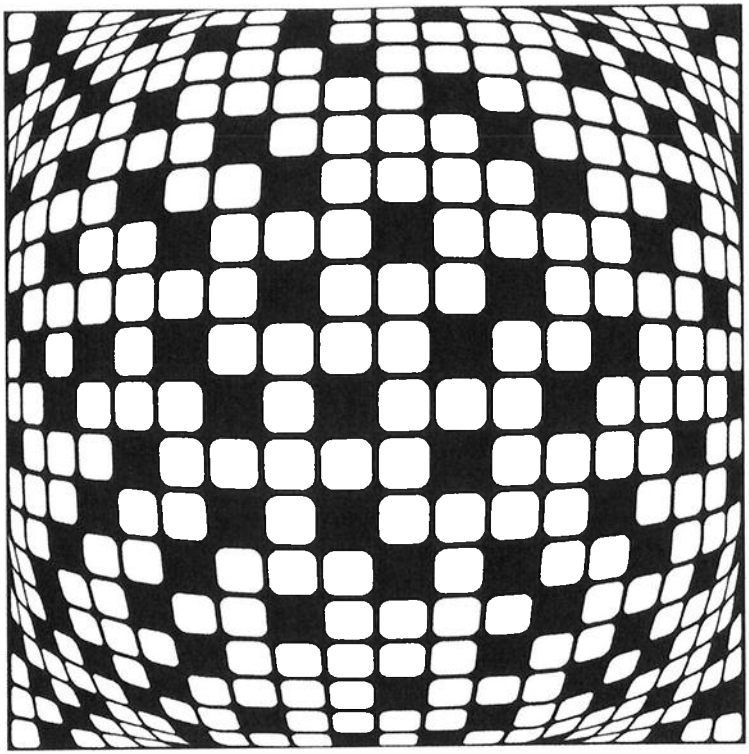


Figure 9.5 **The textures of objects can provide a sense of shape and depth. The center of the image seems closer than the edges due to texture gradient.** (Courtesy of Galkin Grigory/ Shutterstock.com)

Texture gradient. Most natural objects contain visual texture. **Texture gradient** is similar to linear perspective; texture density increases with distance between the eye and the object surface (Figure 9.5).

Height relative to horizon. **Height relative to the horizon** makes objects appear to be further away than when they are closer to the horizon (Figure 9.6).

Aerial perspective. Aerial perspective, also called atmospheric perspective, is a depth cue where objects with more contrast (e.g., brighter, sharper, and more colorful) appear closer than duller objects (Figure 9.7). This results from scattering of light by particles in the atmosphere.

Meta Platforms, Inc.
Exhibit 1063
Page 143

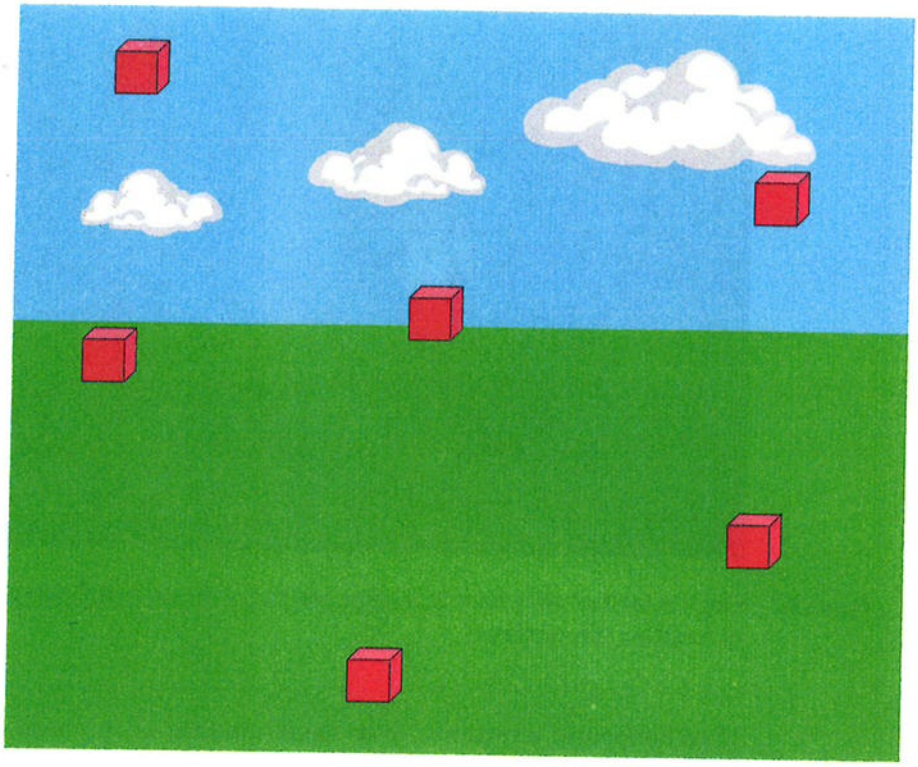


Figure 9.6 Objects closer to the horizon result in the perception that they are further away.

Motion Depth Cues

Motion cues are relative movements on the retina between the observer and the world.

Motion parallax. Motion parallax is a depth cue where the images of distal stimuli projected onto the retina (proximal stimuli) move at different rates depending on their distance. For example, when you look outside the side window of a moving car, nearby objects appear to speed by in a blur, whereas distant objects appear to move more slowly. If while moving, you focus on a single point called the fixation point, objects further and closer than that fixation point appear to move in opposite directions (closer objects move against the direction of physical motion and further objects appear to move with the direction of physical motion). Motion parallax enables us to look around objects and is a strong cue for relative perception of depth. Motion parallax works equally in all directions, whereas binocular disparity only occurs in

Meta Platforms, Inc.
Exhibit 1063
Page 144

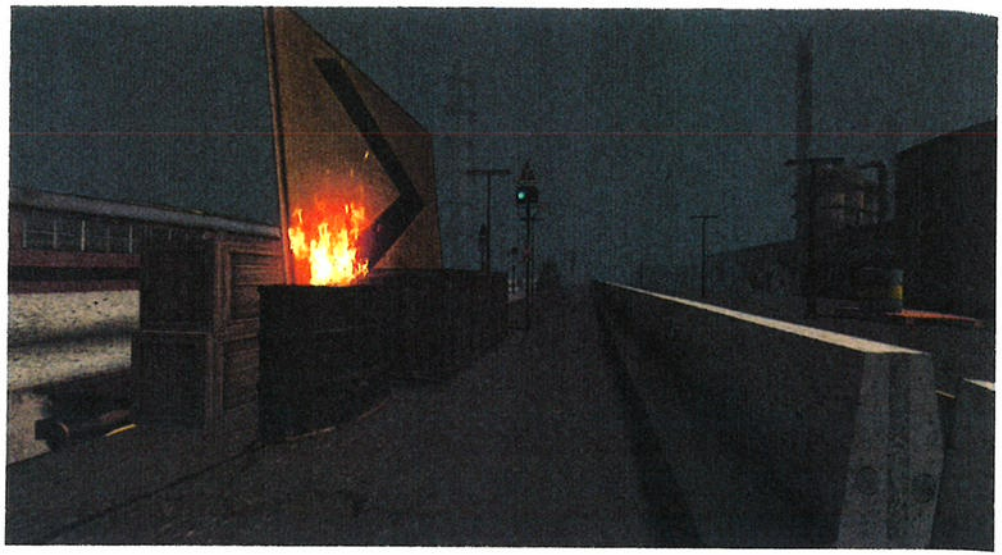


Figure 9.7 **Objects further away have less color and detail due to aerial perspective.** (Image courtesy of NextGen Interactions)

head-horizontal directions. Even small motions of the head due to breathing can result in subtle motion parallax that adds to presence (Andrew Robinson and Sigurdur Gunnarsson, personal communication, May 11, 2015).

Motion parallax is a problem for world-fixed displays (Section 3.2.1) when there are multiple viewers but only one person's viewpoint is being tracked. The shape and motion of the scene is correct for the person being tracked but appears as distortion and objects swaying for the non-tracked viewers.

Kinetic depth effect. The **kinetic depth effect** is a special form of motion parallax that describes how 3D structural form can be perceived when an object moves. For example, observers can perceive the 3D structure of a wire simply by the motion and deformation of the shadow cast by that wire as it is rotated.

Binocular Depth Cues

Those of us with two healthy eyes receive slightly different views of the world on each of our retinas, with the differences being a function of distance. **Binocular disparity** is the difference in image location of a stimulus seen by the left and right eyes resulting from the eyes' horizontal separation and the stimulus distance. **Stereopsis** (also known

Meta Platforms, Inc.
Exhibit 1063
Page 145

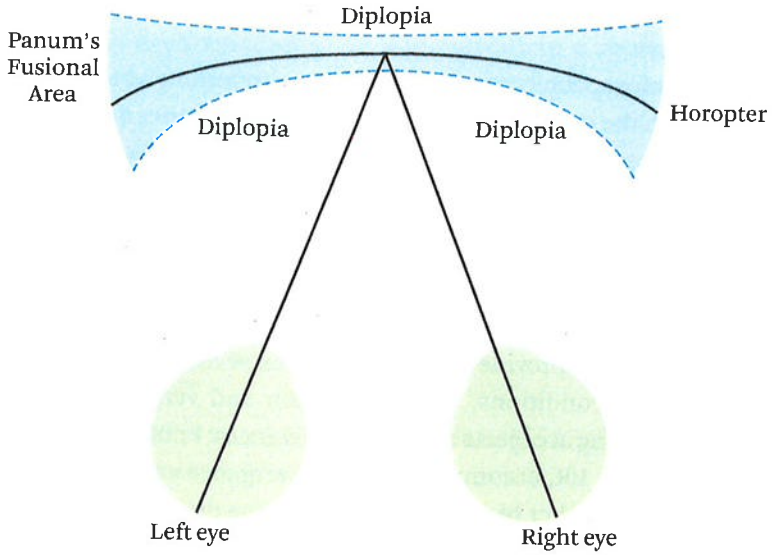


Figure 9.8 **The horopter and Panum's fusional area.**

as binocular fusion) occurs when the brain combines separate and slightly different images from each eye to form a single percept that contains a vivid sense of depth.

Stimuli at a distance resulting in no disparity form a surface in space called the **horopter** (Figure 9.8). The surrounding area in front of and behind the horopter where we can perceive stereoscopic depth is called **Panum's fusional area.** The area outside Panum's fusional area contains disparity that is too great to fuse, causing double images resulting in diplopia.

HMDs can be **monocular** (one image for a single eye), **biocular** (two identical images for each eye), or binocular (two different images for each eye, providing a sense of stereoscopic depth). Binocular images can cause conflicts with accommodation, vergence, and disparity (Section 13.1). Biocular HMDs cause a conflict with motion parallax, but there are few complaints of the conflict.

The value of binocular cues is often debated for stereoscopic displays and VR. Part of this is likely due to some proportion of the population being stereoblind—it is estimated that 3–5% of the population is unable to extract depth signaled purely by disparity differences in modern 3D displays (Badcock et al., 2014). The incidence of such stereoblindness increases with age, and there are likely different levels of

Meta Platforms, Inc.
Exhibit 1063
Page 146

stereoblindness beyond the 3–5%, resulting in disagreement on the value of binocular displays.

For most people, stereopsis becomes important when interacting in personal space through the use of the hands, especially when other depth cues are not strong. We do perceive some binocular depth cues at further distances (up to ~40 meters under ideal conditions—see Section 8.1.4), but stereopsis is much less important at such distance.

Oculomotor Depth Cues

The musculature of the eyes, specifically those dealing with accommodation and vergence, can provide subtle depth cues at distances up to 2 meters. Under normal real-world conditions, accommodation and vergence automatically work together when looking at objects at different distances, known as the accommodation-vergence reflex. With VR, accommodation and vergence can be in conflict with each other (as well as with other binocular cues) when the display focus is set at a distance different than where the eyes feel like they are looking (Section 13.1).

Vergence. As discussed in Section 8.1.5, **vergence** is the simultaneous rotation of the eyes in opposite directions, triggered by retinal disparity, with the primary purpose being to obtain or maintain both sharp and comfortable binocular vision (Reichelt et al., 2010). We can feel inward (**convergence**) and outward (**divergence**) eye movements that provide depth cues. Evidence suggests the effectiveness of convergence alone as a source of information for depth is confined to a range of up to about 2 meters (Cutting and Vishton, 1995). Convergence is a more effective cue than accommodation, described below.

Accommodation. Our eyes contain elastic lenses that change curvature, enabling us to focus the image of objects sharply on the retina. **Accommodation** is the mechanism by which the human eye alters its optical power to hold objects at different distances into focus on the retina. We can feel the tightening of eye muscles that change the shape of the lens to focus on nearby objects, and this muscular contraction can serve as a cue for distance for up to about 2 meters (Cutting and Vishton, 1995). Full accommodation response requires a minimum fixation time of one second or longer and is triggered primarily by retinal blur (Reichelt et al., 2010). Once we have accommodated to a specific distance, objects much closer or further from that distance are out of focus. This blurriness can also provide subtle depth cues, but it is ambiguous if the blurred object is closer or further away from the object that is in focus. Objects having low contrast represent a weak stimulus for accommodation.

Meta Platforms, Inc.
Exhibit 1063
Page 147

The ability to accommodate decreases with age, and most people cannot maintain close-up accommodation for any more than a short period of time. VR designers should not place essential stimuli that must be consistently viewed close to the eye.

Contextual Distance Factors

Although research on visual depth cues has been around for centuries, more recent research suggests distance perception is influenced not only by visual cues but also by environmental context and personal variables. For example, the accuracy of distance perception has been found to differ depending on whether objects being judged are indoors or outdoors, or even in different types of indoor environments (Renner et al., 2013).

Future effect cues are internal psychological beliefs of how distance may personally affect the viewer in the future. Such cues include one's intended actions and fear.

Intended action. Perception and action are closely intertwined (Section 10.4). **Action-intended distance cues** are psychological factors of future actions that influence distance perception. People in the same situation perceive the world differently depending upon the actions they intend to perform and benefits/costs of those actions (Proffitt, 2008). We see the world as "reachers" only if we intend to reach, as "throwers" only if we intend to throw, and as "walkers" only if we intend to walk. Objects within reach appear closer than those out of reach, and when reachability is extended by holding a tool that is intended to be used, apparent distances are diminished. Hills appear to be steeper than they actually are. A manipulation that influences the effort to walk will influence people's perception of distance if they intend to walk. Hills of 5° are typically judged to be about 20°, and 10° hills appear to be 30° (Proffitt et al., 1995). This overestimation increases for those encumbered by wearing a heavy backpack, fatigue, low physical health, old age, and/or declining health (Bhalla and Proffitt, 1999). Hills appear steeper and egocentric distances farther, as the anticipated metabolic energy costs associated with walking increase.

Fear. **Fear-based distance cues** are fear-causing situations that increase the sense of height. We overestimate vertical distance, and this overestimation is much greater when the height is viewed from above than from below. A sidewalk appears steeper for those fearfully standing on a skateboard compared to those standing fearlessly on a box (Stefanucci et al., 2008). This is likely due to an evolved adaptation—enhancing the perception of vertical height motivates people to avoid falling. Fear influences spatial perception.

Meta Platforms, Inc.
Exhibit 1063
Page 148

9.1.4 Measuring Depth Perception

There are several methods to test distance perception and those methods can be divided into three categories: (1) verbal estimation, (2) perceptual matching, and (3) visually directed actions (Renner et al., 2013). The different methods often result in different estimates. Even the phrasing of instruction within the same method can influence how distance estimates are made. For example, "how far away does the object visually appear to be" or "how far away the object really is" may result in different estimates. With carefully designed methods, depth perception can be quite accurate for distances of up to 20 meters.

9.1.5 VR Challenges of Space Perception

Natural perception of space within VR still remains a challenge for display technology as well as content creation. For example, whereas egocentric distance perception can be quite good for the real world, VR egocentric distance estimates are reported to be compressed (Loomis and Knapp, 2003; Renner et al., 2013). Veridical spatial perception is not essential for all VR applications, but it is extremely important where perception of distance and object size is essential, such as architecture, automotive, and military applications.

9.2 Time Perception

Time can be thought of as a person's own perception or experience of successive change. Unlike objectively measured time, **time perception** is a construction of the brain that can be manipulated and distorted under certain circumstances. The concept of time is essential to most cultures, with languages having distinct tenses of past, present, and future, plus innumerable modifiers to specify the "when" more precisely.

9.2.1 A Breakdown of Time

The **subjective present** is always with us as the precious few seconds of our ongoing conscious experience. Everything else does not exist, but is either past or future. The subjective present consists of (1) the "now"—i.e., the current moment, and (2) the experience of passing time (Coren et al., 1999). The experience of passing time can further be broken down into (a) duration estimation, (b) order/sequence (consisting of simultaneity and successiveness), and (c) anticipation/planning for the immediate future (e.g., knowing what note to play next with a musical instrument).

Psychological time can be considered a sequence of moments instead of a continuous experience. A **perceptual moment** is the smallest psychological unit of time that an observer can sense. Stimuli presented within the same perceptual moment are per-

Meta Platforms, Inc.
Exhibit 1063
Page 149

ceived as occurring simultaneously. Based on several research findings, Stroud (1986) estimated that perceptual moments are in the 100 ms range in duration. The length of these perceptual moments ranges from about 25 to 150 ms depending upon the sensory modality, stimulus, task, etc. Evidence suggests perceptual moments become unstable when one judges with more than one sensory modality, such as when judging the order of presentation of a sound and a light (Ulrich, 1987).

An **event** is a segment of time at a particular location that is perceived to have a beginning and an end or a series of perceptual moments that unfolds over time. Our ability to understand relationships between actions and objects depends on this unfolding. Events are similar to sentences and phrases in language, where the sequence and timing of nouns and verbs gives meaning. There is a big difference between "bear eats man" and "man eats bear." The two phrases contain the same elements, but it is the sequence and timing of these stimuli at the eye or ear that gives meaning.

9.2.2 The Brain and Time

Delayed Perception

Despite what most of us would like to believe, we all live in the past with our consciousness always lagging behind reality. It takes between 10 ms and 50 ms for visual signals to reach the LGN (Section 8.1.1), another 20–50 ms for those signals to reach the visual cortex, and an additional 50–100 ms for reaching parts of the brain that are responsible for action and planning (Coren et al., 1999). This can cause problems when response time is essential. For example, it takes a driver over 200 ms to perceive a stimulus that requires the decision to apply the brakes. This number can be even worse if the driver is dark adapted (Section 10.2.1). After perceiving the stimulus, it takes a relatively short amount of time at 10–15 ms to initiate muscle responses. For a driver traveling at 100 kph (62 mph), the car will have traveled 6 m (20 ft) before the foot starts moving toward the brakes. For cases where we can anticipate events, this delay can be reduced via top-down processing that primes cells to respond faster.

Although we live in the past couple hundred milliseconds, this does not mean it is okay to have a couple hundred milliseconds of VR system delay. Such additional external-to-the-body latency is a major cause of motion sickness as discussed in Chapter 15.

Persistence

In addition to delayed perception, the length of an experience depends on neural activity, not necessarily the duration of a stimulus. **Persistence** is the phenomenon by which an afterimage seemingly persists visually. A single 1 ms flash of light is often

Meta Platforms, Inc.
Exhibit 1063
Page 150

experienced as a 100 ms perceptual moment when the observer is light adapted and as long as 400 ms when dark adapted (Coren et al., 1999).

When two stimuli are close in time and space, the two stimuli can be perceived as a single moving stimulus (Section 9.3.6). For example, pixels displayed in different locations at different times can appear as a single moving stimulus. **Masking** can also occur where one of two stimuli (usually visuals or sounds) are not perceived. The second stimulus is typically perceived more accurately (backward masking) as if the second stimulus goes back in time and erases the first stimulus.

Perceptual Stability

Perceptual stability is the illusion of continuity and completeness, even in cases when at any given moment only parts of stimuli from the world reach our sensory receptors (Coren et al., 1999). The brain often constructs events after multiple stimuli have been presented, helping to provide continuity for the perception of a single event across breaks or interruptions that occur because of extraneous stimuli in the environment.

The **phonemic restoration effect** (Section 8.2.3) causes us to perceive missing parts of speech, where what we hear in place of the missing parts is dependent on context. The visual system behaves in a similar way. When an object moves behind other occluding objects, then we perceive the entire object even when there is no single moment in time when the occluded object is completely seen. This can be seen by looking through a narrow slit created by a partially opened door (e.g., even for an opening of only 1 mm). As the head is moved or an object is moved behind the slit, we perceive the entire scene or object to be present due to the brain stitching together the stimuli over time. Perceptual stability is one reason why we don't normally perceive the blind spot (Section 6.2.3).

9.2.3 Perceiving the Passage of Time

Unlike the sensory modalities, there is no obvious time organ that enables us to perceive the passage of time. Instead it is through change that we perceive time. Two theories that describe how we are able to perceive time are biological clocks and cognitive clocks (Coren et al., 1999).

Biological Clocks

The world and its inhabitants have their own rhythms of change, including day-night cycles, moon cycles, seasonal cycles, and physiological changes. **Biological clocks** are body mechanisms that act in periodic manners, with each period serving as a tick of the clock, which enable us to sense the passing of time. Altering the speed of our physiological rhythm alters the perception of the speed of time.

Meta Platforms, Inc.
Exhibit 1063
Page 151

A **circadian rhythm** (circadian comes from the root circa, meaning "approximate" and dies, meaning "days") is a biologically recurring natural 24-hour oscillation. Circadian rhythms are largely controlled by the supra chiasmatic nucleus (SCN), a small organ near the optic chiasm of only a few thousand cells. Circadian rhythms result in sleep-wakefulness timed behaviors that run through a day-night cycle. Although circadian rhythms are endogenous, they are entrained by external cues called zeitgebers (German for "time-givers"), the most important being light. Without light, most of us actually have a rhythm of 25 hours.

Biological processes such as pulse, blood pressure, and body temperature vary as a function of the day-night cycle. Other effects on our mind and body include changes to mood, alertness, memory, perceptual motor tasks, cognition, spatial orientation, postural stability, and overall health. The experiences of VR users may not be as good at times when they are normally sleeping.

Slow circadian rhythms do not help us with our perception of short time periods. We have several biological clocks for judgments of different amounts of time and different behavioral aspects. **Short-term biological timers** that have been suggested for internal biological timing mechanisms include heartbeats, electrical activity in the brain, breathing, metabolic activities, and walking steps.

Our perception of time also seems to change when our biological clocks change. Durations seem to be longer when our biological clocks tick faster due to more biological ticks per unit time than normally occur. Conversely, time seems to whiz by when our internal clock is too slow. Evidence demonstrates body temperature, fatigue, and drugs cause change in time duration estimates.

Cognitive Clocks

Cognitive clocks are based on mental processes that occur during an interval. Time is not directly perceived but rather is constructed or inferred. The tasks a person engages in will influence the perception of the passage of time.

Cognitive clock factors affecting time perception are (1) the amount of change, (2) the complexity of stimulus events that are processed, (3) the amount of attention given to the passage of time, and (4) age. All these affect the perception of the passage of time, and all are consistent with the idea that the rate at which the cognitive clock ticks is affected by how internal events are processed.

Change. The ticking rate of cognitive clocks is dependent on the number of changes that occur. The more changes that occur during an interval, the faster the clock ticks, and thus the longer the estimate of the amount of time passed.

Meta Platforms, Inc.
Exhibit 1063
Page 152

The **filled duration illusion** is the perception that a duration filled with stimuli is perceived as longer than an identical time period empty of stimuli. For example, an interval filled with tones is perceived as being longer than the same interval with fewer tones. This is also the case for light flashes, words, and drawings. Conversely, those in a soundproof darkened chamber tend to underestimate the amount of time they have spent in the chamber.

Processing effort. The ticking rate of cognitive clocks is also dependent on cognitive activity. Stimuli that are more difficult to process result in judgments of longer durations. Similarly, increasing the amount of memory storage required to process information results in judgments of longer durations.

Temporal attention. Time perception is complicated by the way observers attend to a task. The more attention one pays to the passage of time, the longer the time interval seems to be. Paying attention to time may even reverse the filled duration illusion mentioned above because it is difficult to process many events while attending at the same time to the passage of time.

When observers are told in advance that they will have to judge the time a task takes, they tend to judge the duration as being longer than they do when they are unexpectedly asked to judge the duration after the task is completed. This is presumably due to drawing attention to time.

Conversely, anything that draws attention away from the passage of time results in a faster perception of time. Duration estimates are typically shorter for difficult tasks than for easy tasks because it is more difficult to attend to both time and a difficult task.

Age. As people age, the passage of larger units of time (i.e., days, months, or even years) seems to speed up. One explanation for this is that the total amount of time someone has experienced serves as a reference and a perceived duration of a time interval as compared to that reference. One year for a 4-year old is 20% of his life so time seems to drag by. One year for a 50-year old is 2% of his life, so time seems to move more quickly.

9.2.4 Flicker

Flicker is the flashing or repeating of alternating visual intensities. Retinal receptors respond to flickering light at up to several hundred cycles per second, although sensitivity in the visual cortex is much less (Coren et al., 1999). Perception of flicker covers a wide range depending on many factors including adaptation, light intensity, retinal

Meta Platforms, Inc.
Exhibit 1063
Page 153

location, stimulus size, blanking time between stimuli, time of day (due to circadian rhythms), gender, age, bodily functions, and wavelength. We are most sensitive to a light-adapted eye with a high light intensity over a wide range of the visual periphery when the body is most awake. The **flicker-fusion frequency threshold** is the flicker frequency where flicker becomes visually perceptible (Section 9.2.4).

9.3 Motion Perception

Perception is not static but varies continuously over time. Perception of motion at first seems quite trivial. For example, visual motion simply seems as a sense of stimuli moving across the retina. However, motion perception is a complex process that involves a number of sensory systems and physiological components (Coren et al., 1999).

We can perceive movement even when visual stimuli are not moving on the retina, such as when tracking a moving object with our eyes (even when there are no other contextual stimuli to compare to). At other times, stimuli move on the retina yet we do not perceive movement; for example, when the eyes move to examine different objects. Despite the image of the world moving on our retinas, we perceive a stable world (Section 10.1.3).

A world perceived without motion can be both disturbing and dangerous. A person who has akinetopsia (blindness to motion) sees people and objects suddenly appearing or disappearing due to not being able to see them approach. Pouring a cup of coffee becomes nearly impossible. The lack of motion perception is not just a social inconvenience but can be quite dangerous; consider crossing a street where a car seems far away and then suddenly appears very close.

9.3.1 Physiology

The primitive visual pathway through the brain (Section 8.1.1) is largely specialized for motion perception along with the control of reflexive responses such as eye and head movements. The primary visual pathway also contributes to motion detection through magno cells (and to a lesser extent parvo cells) as well as through both the ventral pathway (the "what" pathway) and the dorsal pathway (the "where/action/how" pathway).

The primate brain has visual receptor systems dedicated to motion detection (Lisberger and Movshon, 1999), which humans use to perceive movement (Nakayama and Tyler, 1981). Whereas visual velocity (i.e., speed and direction of a visual stimulus on

Meta Platforms, Inc.
Exhibit 1063
Page 154

the retina) is sensed directly in primates, visual acceleration is not, but is instead inferred through processing of velocity signals (Lisberger and Movshon, 1999). Most visual perception scientists agree that for perception of most motion, visual acceleration is not as important as visual velocity (Regan et al., 1986). However, visual linear acceleration can evoke motion sickness more so than visual linear velocity when there is a mismatch with linear acceleration detected by the vestibular system (Section 18.5).

9.3.2 Object-Relative vs. Subject-Relative Motion

Object-relative motion perception (similar to exocentric judgments; Section 9.1.1) occurs when the spatial relationship between stimuli changes. Because the relationship is relative, which object is moving is ambiguous. Visual object-relative judgments depend solely upon stimuli on the retina and do not take into account extra-retinal information such as eye or head motion.

Subject-relative motion perception (similar to egocentric judgments; Section 9.1.1) occurs when a stimulus moves relative to the observer. This subject-relative frame of reference for visual perception is provided by extra-retinal information such as vestibular input. Even when the head is not moving the vestibular system still receives input. A purely subject-relative cue without any object-relative cues occurs only when a single visual stimulus is visible or all visual stimuli have the same motion.

People use both object-relative and subject-relative cues to judge motion of the external world. For all but the shortest intervals, humans are much more sensitive to object-relative motion than subject-relative motion (Mack, 1986) due to perceiving displacement relative to a visual context rather than velocity. For example, one study found that observers can detect a luminous dot moving in the dark (subject-relative) at about 0.2°/s whereas observers can detect a moving target in the presence of a stationary visual context (object-relative) at about 0.03°/s.

Visuals in augmented reality are considered to be mostly object-relative to real-world cues, as the visuals can be directly compared with the real world. Incorrectly moving visuals are more easily noticed in optical-see-through HMDs (i.e., augmented reality) than non-see-through HMDs (Azuma, 1997), since users can directly compare rendered visuals relative to the real world and hence see spurious object-relative motion. Visual error in VR due to problems such as latency and miscalibration is usually subject-relative, as such error causes the entire scene to move as the user moves the head (other than for motion parallax, where visuals move differently as a function of viewpoint translation and depth).

Meta Platforms, Inc.
Exhibit 1063
Page 155

9.3.3 Optic Flow

Optic flow is the pattern of visual motion on the retina—the pattern of apparent motion of objects, surfaces, and edges on the retina caused by the relative motion between a person and the scene. As one looks to the left, optic flow moves to the right. As one moves foward, optic flow expands in a radial pattern. **Gradient flow** is the different speed of flow for different parts of the retinal image—fast near the observer and slower further away. The focus of expansion is the point in space around which all other stimuli seem to expand as one moves forward toward it.

9.3.4 Factors of Motion Perception

Many factors influence our perception of motion. For example, we are more sensitive to visual motion as contrast increases. Some other factors are discussed below.

Motion Perception in Peripheral vs. Central Vision

The literature conflicts as to whether sensitivity to motion increases or decreases with eye eccentricity (the distance from the stimulus on the retina to the center of the fovea). These conflicting claims are likely due to differences of experimental conditions as well as interpretations.

Anstis (1986) states that it is sometimes mistakenly claimed that peripheral vision is more sensitive to motion than central vision. In fact, the ability to detect slow-moving stimuli actually decreases steadily with eye eccentricity. However, since sensitivity to static detail decreases even faster, peripheral vision is relatively better at detecting motion than form. A moving object seen in the periphery is perceived as something moving, but it is more difficult to see what that something is.

Coren et al. (1999) states that the detection of movement depends on both the speed of the moving stimulus and eye eccentricity. A person's ability to detect slow-moving stimuli (less than 1.5°/s) decreases with eye eccentricity, which is consistent with Anstis. For faster-moving stimuli, however, the ability to detect moving stimuli increases with eye eccentricity. These differences are due to the dominance in the periphery of the dorsal or "action" pathway (Section 8.1.1), which consists of more transient cells (cells that respond best to fast-changing stimuli). Scene motion in the peripheral visual field is also important in sensing self-motion (Sections 10.4.3 and 9.3.10).

Judging motion consists of more than just looking for moving stimuli. Qualitative evidence suggests slight visual motion in the periphery may more easily be judged by "feeling" (e.g., vection as discussed in Section 9.3.10) whereas slight motion in central

Meta Platforms, Inc.
Exhibit 1063
Page 156

vision may more easily be judged by direct visual and logical thinking. For example, two subjects judging a VR scene motion perception task (Jerald et al., 2008) stated:

> Most of the times I detected the scene with motion by noticing a slight dizziness sensation. On the very slight motion scenes there was an eerie dwell or 'suspension' of the scene; it would be still but had a floating quality.
>
> At the beginning of each trial [sic: session—when scene velocities were greatest because of the adaptive staircases], I used visual judgments of motion; further along [later in sessions] I relied on the feeling of motion in the pit of my stomach when it became difficult to discern the motion visually.

These different ways of judging visual motion are likely due to the different visual pathways (Section 8.1.1).

Head Motion Suppresses Motion Perception

Although head motion suppresses perception of visual motion, perception of visual motion while the head is moving can be surprisingly accurate. Loose and Probst (2001) found that increasing the angular velocity of the head significantly suppresses the ability to detect visual motion when the visual motion moves relative to the head. Adelstein et al. (2006) and Li et al. (2006) also showed that head motion suppresses perception of visual motion. They used an HMD without head tracking where the slight motion to detect was relative to the head. Jerald (2009) measured how head motion suppresses perception of visual motion where the slight motion to detect was relative to the real world, and then mathematically related that suppression to the perception of latency.

Depth Perception Affects Motion Perception

The **pivot hypothesis** (Gogel, 1990) states that a point stimulus at a distance will appear to move as the head moves if its perceived distance differs from its actual distance. A related effect is demonstrated by focusing on a finger held in front of the eyes and noticing that the background further in the distance seems to move with the head. Likewise if one focuses on the background then the finger seems to move against the direction of the head.

Objects in HMDs tend to appear closer to users than their intended distance (Section 9.1.5). According to the pivot hypothesis, if a user looks at an object as if it is close, but the system moves the object's image in the HMD as if it is further away when turning the head, then the object will appear to move in the same direction as the head (similar to the finger example above).

Meta Platforms, Inc.
Exhibit 1063
Page 157

9.3.5 Induced Motion

Induced motion occurs when motion of one object induces the perception of motion in another object.

When a dot moves near an existing dot in an otherwise uniform visual field, it is clear one of the dots is moving but often it is difficult to identify which of the two dots is moving (due to the object-relative cues telling us how objects move relative to each other, but not relative to the world). However, when a rectangular frame and dot are observed, we tend to perceive the dot to be moving even if the dot is stable and the frame is what is moving. This is because the mind assumes smaller objects are more likely to move than larger surround stimuli and the larger surround stimuli serve as a stable context (see the moon-cloud illusion and the autokinetic effect in Section 6.2.7 and the rest frame hypothesis in Section 12.3.4).

Induced motion effects are most dramatic when the context is moving slowly, the context is square vs. circular, the surround is larger, and the target and context are the same distance from the observer.

9.3.6 Apparent Motion

Perception of motion does not require continuously moving stimuli. Apparent motion (also known as stroboscopic motion) is the perception of visual movement that results from appropriately displaced stimuli in time and space even though nothing is actually moving (Anstis, 1986; Coren et al., 1999). It's as if the brain is unwilling to conclude two similar-looking stimuli happened to disappear and appear next to each other close in time. Therefore, the original stimulus must have moved to the new location.

The ability to see discontinuous changes as if they were continuous motion has benefited technology since the 1800s and continues to this day, ranging from flashing neon signs (e.g., older Las Vegas signs giving the illusion of movement) to rolling text displays, television, movies, computer graphics, and VR. For VR creators, understanding apparent motion and the brain's rules of visual motion perception for filling in physically absent motion information is important for conveying not only a sense of motion from an array of pixels, but a sense of stability since the pixels must change appropriately as the head moves in order to appear stable in space.

The mind most often perceives apparent motion by following the shortest distance between two neighboring stimuli. However, in some cases the perceived path of motion can also seem to deflect around an object located between the two flashing stimuli. It is as if the brain tries to figure out how the object could have gotten from point A to point B. Two stimuli of different shapes, orientations, brightnesses, and colors can also be perceived to be a single stimulus that transforms while moving.

Meta Platforms, Inc.
Exhibit 1063
Page 158

Strobing and Judder

Strobing and judder can be a problem for HMDs due to movement of the head and movement of the eyes relative to the screen and stimulus.

Strobing is the perception of multiple stimuli appearing simultaneously, even though they are separated in time, due to the stimuli persisting on the retina. If two adjacent stimuli are flashed too quickly, then strobing can result. Stimuli separated by larger distances require longer time intervals for motion to be perceived.

Judder is the appearance of jerky or unsmooth visual motion. If adjacent stimuli are flashed too slowly, then judder can result (for traditional video, judder can also result from converting between different timing formats).

Factors

Three variables, as shown in Figure 9.9, are involved with the temporal aspects of apparent motion.

Interstimulus interval: The blanking time between stimuli. Increasing the blanking time increases strobing and decreases judder.

Frame duration: The amount of time each stimulus is displayed. Increasing the frame duration reduces strobing and increases judder.

Stimulus onset asynchrony: The total time that elapses between the onset of one frame and the onset of the next frame (interstimulus interval + frame duration). This is the inverse of the display refresh rate.

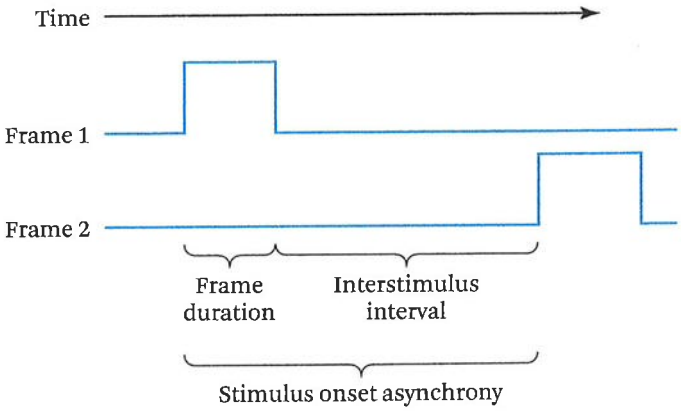


Figure 9.9 **Three variables affecting apparent motion.** (Based on Coren et al., (1999))

Meta Platforms, Inc.
Exhibit 1063
Page 159

The timings required to not notice judder or strobing depend not only on timings but on the spatial properties of the stimuli such as contrast and distance between stimuli. Filmmakers control many of these challenges through limited camera motion, object movement, lighting, and motion blur. The stimulus onset asynchrony required to perceive motion varies widely depending on various factors and ranges from 200 ms for a flashing Vegas casino sign to ~10 ms for smooth motion in a high-quality HMD. VR is much more challenging, partly because the user controls the viewpoint that often is quite fast. The best way to minimize judder for VR is to minimize the stimulus onset asynchrony and frame duration, although strobing can appear if the frame duration is too short (Abrash, 2013).

9.3.7 Motion Coherence

Seemingly randomly placed dots in individual frames can be perceived as having form and motion if those dots move in a coherent manner across frames. **Motion coherence** is the correlation between movements of dots in successive frames. Complete random motion of the dots between frames has 0% motion coherence, and if all dots move in the same way there is 100% motion coherence. Humans are highly sensitive to motion coherence, and we can see form and motion with as little as 3% coherence (Coren et al., 1999). The most optimal condition of perceiving motion coherence is when the dots move at 2°/s. Sensitivity to motion coherence increases as the visual field size increases up to 20° in diameter.

9.3.8 Motion Smear

Motion smear is the trail of visible persistence that is left by an object in motion (Coren et al., 1999). The persistence of the streak is long enough in duration to allow the path to be seen. Although motion can be seen when moving a small lit cigarette or hot campfire ember, motion smear does not require either dark conditions or bright stimuli to be seen. For example, motion smear can be seen by waving the hand back and forth in front of the eyes, resulting in the hand appearing to be in multiple locations simultaneously. Research has shown there is substantial motion smear suppression when a moving stimulus follows a predictable path, so motion smear cannot be due to persistence on the retina alone.

9.3.9 Biological Motion

Humans have especially fine-tuned perception skills when it comes to recognizing human motion. **Biological motion** perception refers to the ability to detect such motion.

Meta Platforms, Inc.
Exhibit 1063
Page 160

Johansson (1976) found that some observers were able to identify human motion represented by 10 moving point lights in durations as short as 200 ms and all observers tested were able to perfectly recognize 9 different motion patterns for durations of 400 ms or more.

In another study, people who were well acquainted with each other were filmed with only lit portions of several of their joints. Several months later, the participants were able to identify themselves and others on many trials just from viewing the moving lit joints (Coren et al., 1999). When asked how they made their judgments, they mentioned a variety of motion components such as speed, bounciness, rhythm, arm swing, and length of steps. In other studies, observers could determine above chance whether moving dot patterns were from a male or female within 5 seconds of viewing even when only the ankles were represented!

These studies imply avatar motion in VR is extremely important, even when the avatar is not seen well or when presented over a short period of time. Any motion simulation of humans must be extremely accurate in order to be believed.

9.3.10 Vection

Vection is an illusion of self-motion when one is not actually physically moving in the perceived manner. Vection is similar to induced motion (Section 9.3.5), but instead of inducing an illusion of a small visual stimulus motion, vection induces an illusion of self-motion. In the real world, vection can occur when one is seated in a car and an adjacent stopped car pulls away. One experiences the car one is seated in, which is actually stationary, to move in the opposite direction of the moving car.

Vection often occurs in VR due to the entire visual world moving around the user even though the user is not physically moving. This is experienced as a compelling illusion of self-motion in the direction opposite the moving stimuli. If the visual scene is moved to the left, the user may feel that she is moving or leaning to the right and will compensate by leaning into the direction of the moving scene (the left). In fact, vection in carefully designed VR experiences can be indistinguishable from actual self-motion (e.g., the Haunted Swing described in Section 2.1).

Although not nearly as strong as vision, other sensory modalities can contribute to vection (Hettinger et al., 2014). Auditory vection can produce vection by itself in the absence of other sensory stimuli or can enhance vection provided by other sensory stimuli. Biomechanical vection can occur when one is standing or seated and repeatedly steps on a treadmill or an even, static surface. Haptic/tactile cues via touching a rotating drum surrounding the user or airflow via a fan can also add to the sense of vection.

Meta Platforms, Inc.
Exhibit 1063
Page 161

Factors

Vection is more likely to occur for large stimuli moving in peripheral vision. Increasing visual motions up to about 90°/s result in increased vection (Howard, 1986a). Visual stimuli with high spatial frequencies also increase vection.

Normally, users correctly perceive themselves to be stable and the visuals to be moving before the onset of vection. This delay of vection can last several seconds. However, at stimulus accelerations of less than about $5°/s^2$, vection is not preceded by a period of perceived stimulus motion as it is at higher rates of acceleration (Howard, 1986b).

Vection is not only attributed to low-level bottom-up processing. Research suggests higher-level top-down vection processing of globally consistent natural scenes with a convincing rest frame (Section 12.3.4) dominates low-level bottom-up vection processing (Riecke et al., 2005). Stimuli perceived to be further away cause more vection than closer stimuli, presumably because our experience with real life tells us background objects are typically more stable. Moving sounds that are normally perceived as landmarks in the real world (e.g., church bells) can cause greater vection than from artificial sounds or sounds typically generated from moving objects.

If vection depends on the assumption of a stable environment, then one would expect the sensation of vection would be enhanced if the stimulus is accepted as stationary. Conversely, if we can reverse that assumption by thinking of the world as an object that can be manipulated (e.g., grabbing the world and moving it toward us as done in the 3D multi-touch pattern; Section 28.3.3), then vection and motion sickness can be reduced (Section 18.3). Informal feedback from users of such manipulable worlds indeed suggests this to be the case. However, further research is required to determine to what degree.

Meta Platforms, Inc.
Exhibit 1063
Page 162

Perceptual Stability, Attention, and Action

10.1 Perceptual Constancies

Our perception of the world and the objects within it is relatively constant. A **perceptual constancy** is the impression that an object tends to remain constant in consciousness even though conditions (e.g., changes in lighting, viewing position, head turning) may change. Perceptual constancies occur partly due to the observer's understanding and expectation that objects tend to remain constant in the world.

Perceptual constancies have two major phases: registration and apprehension (Coren et al., 1999).

Registration is the process by which changes in the proximal stimuli are encoded for processing within the nervous system. The individual need not be consciously aware of registration. Registration is normally oriented on a focal stimulus, the object being paid attention to. The surrounding stimuli are the context stimuli.

Apprehension is the actual subjective experience that is consciously available and can be described. During apprehension, perception can be divided into two properties: object properties and situational properties. **Object properties** tend to remain constant over time. **Situational properties** are more changeable, such as one's position relative to an object or the lighting configuration.

Table 10.1 summarizes some common perceptual constancies that are further discussed below (Coren et al., 1999).

10.1.1 Size Constancy

Why do objects not seem to change size when we walk toward them? After all, the projected size of an object on the retina changes as we walk toward or away from it. For example, if we walk toward an animal in the distance then we do not gasp in horror as if a monster is growing. The sensory reality is that the animal projected on our retina

Meta Platforms, Inc.
Exhibit 1063
Page 163

Table 10.1 **Perceptual constancies.** (Adapted from Coren et al. (1999))

	Registered Stimulus (may be unconscious)		Apprehended Stimulus (conscious)	
Constancy	Focal Stimulus	Context	Constant	Changes
Size constancy	retinal image size	distance cues	object size	object distance
Shape constancy	retinal image shape	orientation cues	object shape	object orientation
Position constancy	retinal image location	sensed head or eye pose	object position in space	head or eye pose
Lightness constancy	retinal image intensity	illumination cues	surface whiteness intensity	apparent illumination
Color constancy	retinal image color	illumination cues	surface colors	apparent illumination color
Loudness constancy	ear sound intensity	distance cues	loudness of sounds	distance from sound

is growing as we walk toward it, but we do not perceive the animal is changing size. Our past experiences and mental model of the world tells us things do not change size when we walk toward them—we have learned object properties remain constant independent of our movement through the world. This learned experience that objects tend to stay the same size is called **size constancy.**

Size constancy is largely dependent on the apparent distance of an object being correct, and in fact changes in misperceived distance can change perceived size. The more distance cues (Section 9.1.3) available, the better size constancy. When only a small number of depth cues are available, distant objects tend to be perceived as too small.

Size constancy does not seem to be consistent across all cultures. Pygmies, who live deep in the rain forest of tropical Africa, are not often exposed to wide-open spaces and do not have as much opportunity to learn size constancy. One Pygmy, outside of his typical environment, saw a herd of buffalo at a distance and was convinced he was seeing a swarm of insects (Turnbull, 1961). When driven toward the buffalo, he became frightened and was sure some form of witchcraft was at work as the insects "grew" into buffalo.

Meta Platforms, Inc.
Exhibit 1063
Page 164

Size constancy does not occur only in the real world. Even with first-person video games containing no stereoscopic cues, we still perceive other characters and objects to remain the same size even when their screen size changes. With VR, size constancy can be even stronger due to motion parallax and stereoscopic cues. However, if those different cues are not consistent with each other, then the perception of size constancy can be broken (e.g., objects become smaller on the screen, but our stereo vision tells us the object is at the same distance). The same applies for shape constancy and position constancy described below.

10.1.2 Shape Constancy

Shape constancy is the tendency for objects to retain their shape, even as we look at them from different angles resulting in a changing shape of their image on the retina. For example, we perceive the rim of a coffee mug as a circle even though it is only a circle on the retina when viewed from directly above (otherwise it is an ellipse when viewed at an angle or a line when viewed straight on). The reason for this is experience tells us cups are round so we label in our minds that the rims of cups are circular. Imagine the chaos that would result if we all described objects as the shapes that appear on our retinas with each changing moment.

Not only does shape constancy provide us with consistent perception of objects, but it helps us to determine the orientation and relative depth of those objects.

10.1.3 Position Constancy

Position constancy is the perception that an object appears to be stationary in the world even as the eyes and the head move (Mack and Herman, 1972). A perceptual matching process between motion on the retina and extra-retinal cues causes the motion to be discounted.

The real world remains stable as one rotates the head. Turning the head 10° to the right causes the visual field to rotate to the left by 10° relative to the head. The **displacement ratio** is the ratio of the angle of environmental displacement to the angle of head rotation. The stability of the real world is independent of head motion and thus has a displacement ratio of zero. If the environment rotates with the head then the displacement ratio is positive. Likewise, if the environment rotates against head direction the displacement ratio is negative.

Minifying glasses, i.e., glasses with concave lenses, increase the displacement ratio and cause objects to appear to move with head rotation. Magnifying glasses decrease the displacement ratio below zero and cause objects to appear to move in the opposite direction as head rotation. Rendering a different field of view than the displayed

Meta Platforms, Inc.
Exhibit 1063
Page 165

field of view causes similar effects as minifying/magnifying glasses. For HMDs, this is perceived as scene motion that can result in motion sickness (Chapter 12).

The **range of immobility** is the range of displacement ratio where position constancy is perceived. Wallach (1987) determined the range of immobility to be 0.04–0.06 displacement ratios wide. If an observer rotates her head 100° to the right, then the environment can move by up to 2°–3° with or against the direction of her head turn without her noticing that movement.

In an untracked HMD, the scene moves with the user's head and has a displacement ratio of one. In a tracked HMD with latency, the scene first moves *with* the direction of the head turn (a positive displacement ratio), then after the head decelerates the scene moves back toward its correct position, *against* the direction of the head turn (a negative displacement ratio), after the system has caught up to the user (Jerald, 2009).

Lack of position constancy can result from a number of factors and can be a primary cause of motion sickness as discussed in Part III.

10.1.4 Lightness and Color Constancy

Lightness constancy is the impression that the lightness (Section 8.1.2) of an object is more the perception of the reflectance of an object and intensity of the surrounding stimuli rather than the amount of light reaching the eye (Coren et al., 1999). For example, white paper and black coal still appear to be white and black even as lighting conditions change. Lightness constancy based on relative intensity to surrounding objects holds over a million-to-one range of illumination. Lightness constancy is also affected by shadows, the shape of the object and distribution of illumination on it, and the relative spatial relationship among objects.

Color constancy is the perception for the colors of familiar objects to remain relatively constant even under changing illumination. When a shadow partially falls on a book, we do not perceive the shadowed part of the book to change colors. Like lightness constancy, color constancy is largely due to the relationship with surrounding stimuli. The greater the number of objects and colors in the scene that can serve as comparison stimuli, the greater the perception of color constancy. Prior experience (i.e., top-down processing) also is used to maintain color constancy. A banana almost always appears yellow even in different lighting conditions in part because we know bananas are yellow.

10.1.5 Loudness Constancy

Loudness constancy is the impression that a sound source maintains its loudness even when the sound level at the ear diminishes as the listener moves away from the sound source. Like any constancy, loudness constancy only works up to some threshold.

Meta Platforms, Inc.
Exhibit 1063
Page 166

10.2 Adaptation

Humans must be able to adapt to new situations in order to survive. We must not only adapt to an ever-changing environment, but also adapt to internal changes such as neural processing times and spacing between the eyes, both of which change over a lifetime. When studying perception in VR, investigators must be aware of adaptation, as it may confound measurements and change perceptions.

Adaptation can be divided into two categories: sensory adaptation and perceptual adaptation (Wallach, 1987).

10.2.1 Sensory Adaptation

Sensory adaptation alters a person's sensitivity to detect a stimulus. Sensitivity increases or decreases over time; one starts or stops detecting a stimulus after a period of constant stimulus intensity or after removal of that stimulus. Sensory adaptations are typically localized and only last for short periods of time. Dark adaptation is an example of sensory adaptation.

Dark adaptation increases one's sensitivity to light in dark conditions. **Light adaptation** decreases one's sensitivity to light in light conditions. The sensitivity of the eye changes by as much as six orders of magnitude, depending on the lighting of the environment.

The cones of the eye (Section 8.1.1) reach maximum dark adaptation in approximately 10 min after initiation of dark adaptation. The rods reach maximum dark adaptation in approximately 30 min after initiation of dark adaptation. Complete light adaptation occurs within 5 min after initiation of light adaptation. Dark adaptation causes a person's perception of stimuli to be delayed as discussed in Section 15.3. Occasionally presenting bright large stimuli can keep users light-adapted. Rods are relatively insensitive to red light. Thus, if red lighting is used, rods will dark-adapt whereas cones will maintain high acuity.

10.2.2 Perceptual Adaptation

Perceptual adaptation alters a person's perceptual processes. Welch (1986) defines **perceptual adaptation** to be "a semipermanent change of perception or perceptual-motor coordination that serves to reduce or eliminate a registered discrepancy between or within sensory modalities or the errors in behavior induced by this discrepancy." Six major factors facilitate perceptual adaption (Welch and Mohler, 2014):

- stable sensory rearrangement
- active interaction
- error-corrective feedback

Meta Platforms, Inc.
Exhibit 1063
Page 167

- immediate feedback
- distributed exposures over multiple sessions
- incremental exposure to the rearrangement

Dual adaptation is the perceptual adaptation to two or more mutually conflicting sensory environments that occurs after frequently alternating between those conflicting environments. Going back and forth between a VR experience and the real world may be less of an issue for seasoned users.

One of the first examples of extreme perceptual adaptation goes back to the late 1800s. George Stratton wore inverting goggles that turned his visual world upside-down and right-side-left over a period of eight days (Stratton, 1897). At first, everything seemed backwards and it was difficult to function, but he slowly adapted over several days' time. Stratton eventually achieved near full adaptation so that the world seemed right-side-up and he could fully function while wearing the inverting goggles.

Position-Constancy Adaptation

The compensation process that keeps the environment perceptually stable during head rotation can be altered by perceptual adaptation. Wallach (1987) calls this visual-vestibular sensory rearrangement "adaptation to constancy of visual direction," and in this book it is called **position-constancy adaptation** to be consistent with position constancy described in Section 10.1.3.

VOR (Section 8.1.5) adaptation is an example of position-constancy adaptation. Perhaps the most common form of position-constancy adaptation is eyeglasses that at first cause the stationary environment to seemingly move during head movements but is eventually perceptually stabilized after some use. In a more extreme case, Wallach and Kravitz (1965) had subjects wear minifying glasses that caused the world to seemingly move with the direction of head turns. They found that over time, subjects' perceived motion of the environment during head turns subsided, resulting in apparent position constancy. After adaptation and removal of the optical device, subjects reported a negative aftereffect (Section 10.2.3) such that the real world seemed to move against the direction of their head turns. Draper (1998) found similar adaptation for subjects in HMDs when the rendered field of view was intentionally modified to be different from the true field of view of the HMD.

People are able to achieve dual adaptation of position constancy so that they perceive a stable world for different displacement ratios (Section 10.1.3) if there is a cue (e.g., glasses on the head or scuba-diver face masks; Welch, 1986).

Meta Platforms, Inc.
Exhibit 1063
Page 168

Temporal Adaptation

As discussed in Section 9.2.2, our conscious lags behind actual reality by hundreds of milliseconds. **Temporal adaptation** changes how much time in the past our consciousness lags.

Until relatively recently, no studies have been able to show adaptation to latency (although the Pulfrich pendulum effect from Section 15.3.1 suggests dark adaptation also causes some form of temporal adaptation). Cunningham et al. (2001a) found behavioral evidence that humans can adapt to a new intersensory visual-temporal relationship caused by delayed visual feedback. A virtual airplane was displayed moving downward with constant velocity on a standard monitor. Subjects attempted to navigate through an obstacle field by moving a mouse that controlled only left/right movement. Subjects first performed the task in a pre-test with a visual latency of 35 ms. Subjects were then trained to perform the same task with 200 ms of additional latency introduced into the system. Finally, subjects performed the task in a post-test with the original minimum latency of 35 ms. The subjects performed much worse in the post-test than in the pre-test.

Toward the end of training, with 235 ms of visual latency, several subjects spontaneously reported that visual and haptic feedback seemed simultaneous. All subjects showed very strong negative aftereffects. In fact, when the latency was removed, some subjects reported the visual stimulus seemed to move before the hand that controlled the visual stimulus, i.e., a reverse of causality occurred, where effect seemed to occur before cause!

The authors reasoned that sensorimotor adaptation to latency requires exposure to the consequences of the discrepancy. Subjects in previous studies were able to reduce discrepancies by slowing down their movements when latency was present, whereas in this study subjects were not allowed to slow the constant downward velocity of the airplane.

These results suggest VR users might be able to adapt to latency, thereby changing latency thresholds over time. Susceptibility to VR sickness in general is certainly reduced for many individuals, and VR latency-induced sickness is likely specifically reduced. However, it is not clear if VR users truly perceptually adapt to latency.

10.2.3 Negative Aftereffects

Negative aftereffects are changes in perception of the original stimulus after the adapting stimulus has been removed. Negative aftereffects occur with both sensory adaptation and perceptual adaptation. These negative aftereffects provide the most common measure of adaptation (Cunningham et al., 2001a). Aftereffects as they relate to VR usage are discussed in Section 13.4.

Meta Platforms, Inc.
Exhibit 1063
Page 169

10.3 Attention

Our senses continuously provide an enormous amount of information that we can't possibly perceive all at once. **Attention** is the process of taking notice or concentrating on some entity while ignoring other perceivable information. Attention is far more than simply looking around at things—attending brings an object to the forefront of our consciousness by enhancing the processing and perception of that object. As we pay more attention, we start to notice more detail and perhaps realize the item is different from that which we first thought. Our perception and thinking becomes sharp and clear, and the experience is easy to recall. Attention informs us what is happening, enables us to perceive details, and reduces response time. Attention also helps bind different features, such as shape, color, motion, and location, into perceptible objects (Section 7.2.1).

10.3.1 Limited Resources and Filtering

Our brains have limited physiological capacity and processing capabilities, and attention can be thought of as the allocation of these limited resources. To prevent overloading, the mind-body has evolved to apply more resources to a single area of interest at a time. For example, the fovea is the only location on the retina that provides high-resolution vision. We also filter out events not relevant to our current interests, attending to only one of several available sources of information about the world.

Perceptual capacity is one's total capability for perceiving. **Perceptual load** is the amount of a person's perceptual capacity that is currently being used. Easy or well-practiced tasks have low perceptual loads whereas difficult tasks, such as learning a new skill, have higher perceptual loads.

As discussed in Section 7.9.3, deletion filtering omits certain aspects of incoming data by selectively paying attention to only parts of the world. Things we do not attend to are less distinct and more difficult to remember (or not even perceived as described below as inattentional blindness). Filtering out information allows little of that information to make a lasting impression.

The personal growth industry claims the reticular activating system serves as an automatic goal-seeking mechanism by bringing those things that we think about to the forefront of our attention and filtering out things that we don't care about. Unfortunately, none of these "backed by scientific evidence" claims include references, and a literature search results in little or no scientific evidence supporting such claims. Whereas the **reticular activating system** does enhance wakefulness and general alertness, this part of the brain is only one small piece of the attention puzzle. However,

Meta Platforms, Inc.
Exhibit 1063
Page 170

regardless of what parts of the brain control our attention, the concept of perceiving and attending to what we think about is important.

The Cocktail Party Effect

The **cocktail party effect** is the ability to focus one's auditory attention on a particular conversation while filtering out many other conversations in the same room. **Shadowing** is the act of repeating verbal input one is receiving, often among other conversations. Shadowing is easier if the messages come from two different spatial locations, are different pitches, and/or are presented at different speeds.

Inattentional Blindness

Inattentional blindness is the failure to perceive an object or event due to not paying attention and can occur even when looking directly at the event. Researchers created a video of two teams playing a game similar to basketball and observers were told to count the number of passes, causing them to focus their attention on one of the teams (Simons and Chabris, 1999). After 45 seconds, either a woman carrying an umbrella or a person dressed in a gorilla suit walked through the game over a period of 5 seconds. Amazingly, nearly half of all observers failed to report seeing the woman or the gorilla.

Inattentional blindness also refers to more specific types of perceptual blindness. Discussed below are change blindness, choice blindness, the video overlap phenomenon, and change deafness.

Change blindness. **Change blindness** is the failure to notice a change of an item on a display from one moment to the next. Change blindness can most easily be demonstrated by making a change in a scene when the display is blanked or there is a flash of light between images. Changes in the background are noticed less often than changes in the foreground.

Change blindness can occur even when observers are explicitly told to look for a change. **Continuity errors** are changes between shots in film, and many viewers fail to notice significant changes. Viewers also often miss changes when those changes are introduced during saccadic eye movement.

Most people believe they are good at detecting changes and are not aware of change blindness. This lack of awareness is called **change blindness blindness.**

Choice blindness. **Choice blindness** is the failure to notice that a result is not what the person previously chose. When people believe they chose a result, when in fact they didn't, they often justify with reasons why they "chose" that unchosen result.

Meta Platforms, Inc.
Exhibit 1063
Page 171

The video overlap phenomenon. The **video overlap phenomenon** is the visual equivalent of the cocktail party effect. When observers pay attention to a video that is superimposed on another video, they can easily follow the events in one video but not both.

Change deafness. **Change deafness** is a physical change in an auditory stimulus that goes unnoticed by a listener. Attention influences change deafness.

10.3.2 Directing Attention

Whereas conscious attention is not required to get the "gist" of a scene, we do need to direct attention toward specific parts of a scene while ignoring other parts to perceive details.

Attentional Gaze

Attentional gaze (Coren et al., 1999) is a metaphor for how attention is drawn to a particular location or thing in a scene. Information is more effectively processed at the place where attention is directed. Attention is like a spotlight or zoom lens that improves processing when directed toward a specific location. Due to the ability to covertly focus attention, attentional gaze does not necessarily correspond to where the eye is looking. Attentional gaze can be directed to a local/small aspect of a scene or to a more global/larger area of a scene. We can also choose to attend to a single feature, such as texture or size, rather than the entirety of an object. However, we usually cannot be drawn to more than one location at any single point in time.

Attentional gaze can shift faster than the eye. Attentional gaze resulting from a visual stimulus starts in the primitive visual pathway (Section 8.1.1) resulting in cells firing as fast as 50 ms after a target is flashed, whereas eye saccades often take more than 200 ms to start moving toward that stimulus.

Auditory attention also acts as if it has a direction and can be drawn to particular spatial locations.

Attention Is Active

We do not passively see or hear, but we actively look or listen in order to see and hear. Our experience is largely what we agree to attend to; we don't just passively sit, letting everything enter into our minds. Instead we actively direct our attention to focus on what is important to us—our interest and goals that guide top-down processing.

Top-down processing as it relates to attention is associated with scene schemas. A **scene schema** is the context or knowledge about what is contained in a typical environment that an observer finds himself in. People tend to look longer at things

Meta Platforms, Inc.
Exhibit 1063
Page 172

that seem out of place. People also notice things where they expect them—e.g., we are more likely to look for and notice stop signs at intersections than in the middle of a city block.

Sometimes our past experience or a cue tells us where or when an important event may happen in the near future. We prepare for the event's occurrence by aligning attention with the location and time that we expect. Preparing changes one's attentional state. Auditory cues are especially good for grabbing the attention of a user to prepare for some important event.

Attention also is affected by the task being performed. We almost always look at an object before taking action on that object, and eye movements typically precede motor action by a fraction of a second, providing just-in-time information needed to interact. Observers will also change their attention based on their probabilistic estimates of dynamic events. For example, people will pay more attention to a wild animal or someone acting suspiciously.

Capturing Attention

Attentional capture is a sudden involuntary shift of attention due to salience. **Salience** is the property of a stimulus that causes it to stick out from its neighbors and grab one's attention, such as a sudden flash of light, a bright color, or a loud sound. The attentional capture reflex occurs due to the brain having a need to update as quickly as possible its mental model of the world that has been violated (Coren et al., 1999). If the same stimulus repeatedly occurs, then it becomes an expected part of the mental model and the orienting reflex is weakened.

A **saliency map** is a visual image that represents how parts of the scene stand out and capture attention. Saliency maps are created from characteristics such as color, contrast, orientation, motion, and abrupt changes that are different from other parts of the scene. Observers first fixate on areas that correlate with strong areas of the saliency map. After these initial reflexive actions, attention then becomes more active.

Attention maps provide useful information of what users actually look at (Figure 10.1). Measuring and visualizing attention maps can be extremely useful for determining what actually attracts users' attention. This can be extremely useful as part of the iterative content/interaction creation process by increasing creators' understanding of what draws users toward desired behaviors. Although eye tracking is ideal for creating quality attention maps, less precise attention maps can be created with HMDs by assuming the center of the field of view is what users are generally paying attention to.

Task-irrelevant stimuli are distracting information that is not relevant to the task with which we are involved and can result in decreased performance. Highly salient

Meta Platforms, Inc.
Exhibit 1063
Page 173



Figure 10.1 **3D attention maps show what users most pay attention to.** (From Pfeiffer and Memili (2015))

stimuli are more likely to cause distraction. Task-irrelevant stimuli have more of an impact on performance when a task is easy.

We constantly monitor the environment by shifting both overtly and covertly.

Overt orienting is the physical directing of sensory receptors toward a set of stimuli to optimally perceive important information about an event. The eyes and head reflexively turn toward (with the head following the eyes) salient stimuli. Orienting reflexes include postural adjustments, skin conductance changes, pupil dilation, decrease in heart rate, pause in breathing, and constriction of the peripheral blood vessels.

Covert orienting is the act of mentally shifting one's attention and does not require moving the sensory receptors. Overt and covert orienting together result in enhanced perception, faster identification, and improved awareness of an event. It is possible to covertly orient without an overt sign that one is doing so. Attentional hearing is more often covert than seeing.

Visual Scanning and Searching

Visual scanning is the looking from place to place in order to most clearly see items of interest on the fovea. A **fixation** is a pause on an item of interest and a **saccade** is a jerky eye movement from one fixation to the next (Section 8.1.5). Even though we are often not aware of saccadic eye movements, we make such movements about three times per second. Scanning is a form of overt orienting. **Inhibition of return** is the decreased likelihood of moving the eyes to look at something that has already been looked at.

Meta Platforms, Inc.
Exhibit 1063
Page 174

Searching is the active pursuit of relevant stimuli in the environment, scanning one's sensory world for particular features or combinations of features.

A **visual search** is looking for a feature or object among a field of stimuli and can be either a feature search or a conjunction search. A **feature search** is looking for a particular feature that is distinct from surrounding distractors that don't have that feature (e.g., an angled line). A **conjunction search** is looking for particular combinations of features. Feature searches are generally easier to do than conjunction searches.

For feature searches, the number of distracting items does not affect searching speed and the search is said to operate in parallel. In fact, feature searches sometimes result in the object being searched for to seem to "pop out" from a surrounding blur of irrelevant features. One explanation for this is that our perceptual system groups similar items together and divides scenes into figure and ground (Section 20.4) where the figure is what is being searched for.

For conjunction searches, we compare each object with what we are looking for and respond only when a match is found. Thus, conjunction searches are sometimes referred to as serial searches. Because conjuction searches are performed serially, they are slower than feature searches and are dependent on the number of items being searched.

Vigilance is the act of maintaining careful attention and concentration for possible danger, difficulties, or perceptual tasks, often with infrequent events and prolonged periods of time. After an initial time of conducting a vigilant task, sensitivity to stimuli decreases with time due to fatigue.

Flow

When in the state of **flow**, people lose track of time and that which is outside the task being performed (Csikszentmihalyi, 2008). They are at one with their actions. Flow occurs at just the proper level of difficulty: difficult enough to provide a challenge that requires continued attention, but not so difficult that it invokes frustration and anxiety.

10.4 Action

Perception not only provides information about the environment but also inspires action, thereby turning an otherwise passive experience into an active one. In fact, some researchers believe the purpose of vision is not to create a representation of what is out there but to guide actions. Taking action on information from our senses helps us to survive so we can perceive another day. This section provides a high-level

Meta Platforms, Inc.
Exhibit 1063
Page 175

overview of action and how it relates to perception. Part V focuses on how we can design VR interfaces for users to interact with.

The continuous cycle of interaction consists of forming a goal, executing an action, and evaluating the results (the perception) and is broken down in more detail in Section 25.3. Predicting the result of an action is also an important aspect of perception. See Section 7.4 for an example of how prediction and unexpected results can result in the misperception of the world moving about the viewer.

As discussed in Section 9.1.3, intended actions affect distance perception. Performance can also affect perception—for example, successful batters perceive a ball to be bigger than less successful batters, recent tennis winners report lower nets, and successful field goal kickers estimate goal posts to be further apart (Goldstein, 2014). Perception of how an object might be interacted with (Section 25.1.2) can influence our perception of it. Visual motion can also cause vection (the sense of self-motion; Section 9.3.10) and postural instability (Section 12.3.3).

10.4.1 Action vs. Perception

A large body of evidence suggests visually guided action is primarily processed through the dorsal stream and visual perception without action is primarily processed through the ventral pathway (Section 8.1.1). This results in humans using different metrics and different frames of reference depending on if they are purely observing or also taking action. For example, length estimation is biased when visually observing a Ponzo illusion (Section 6.2.4), but the hand is not fooled by the illusion when reaching out to grab the Ponzo lines (Ganel et al., 2008). That is, the illusion works for perception (the length estimation task) but not for action (the grasping task).

10.4.2 Mirror Neuron Systems

Although mirror neurons are controversial and individual mirror neurons have not yet been proven to exist in humans, the concept of mirror neurons is still useful to think about how action and perception are related. **Mirror neurons** respond in a similar way whether an action is performed or the same action is observed (Goldstein, 2014); i.e., mirror neurons seem to respond to what is happening, independently of who is performing the action. Most mirror neurons are specialized to respond to only one type of action, such as grasping or placing an object somewhere, and the specific type of object that is acted upon or observed has little effect on the neuron's response. Mirror neurons may be important for understanding actions and intentions of other people, learning new skills by imitation, and communicating emotions such as empathy.

Meta Platforms, Inc.
Exhibit 1063
Page 176

Regardless of whether or not individual mirror neurons actually exist in humans, the concept is useful to think about when designing VR interactions. For example, watching a computer-controlled character perform an action can be useful for learning new interaction techniques.

10.4.3 Navigation

Navigation is determining and maintaining a course or trajectory to an intended location.

Navigation tasks can be divided into exploration tasks and search tasks. **Exploration** has no specific target, but is used to browse the space and build knowledge of the environment (Bowman et al., 2004). Exploration typically occurs when entering a new environment to orient oneself to the world and its features. Searching (Section 10.3.2) has a specific target, whose location may be unknown (**naive search**) or previously known (**primed search**). In the extreme, naive search is simply exploration, albeit with a specific goal.

Navigation consists of wayfinding (the mental component) and travel (the motoric component), and the two are intimately tied together (Darken and Peterson, 2014).

Wayfinding is the mental element of navigation that does not involve actual physical movement of any kind but only the thinking that guides movement. Wayfinding involves spatial understanding and high-level thinking, such as determining one's current location, building a cognitive map of the environment, and planning/deciding upon a path from the current location to a goal location. Landmarks (distinct cues in the environment) and other wayfinding aids are discussed in Sections 21.5 and 22.1. Eye movements and fMRI scans show that people look at environmental wayfinding aids more often than other parts of the scene and the brain automatically distinguishes decision point location/cues to guide navigation (Goldstein, 2014).

Wayfinding works through

- **perception**: perceiving and recognizing cues along a route,
- **attention**: attending to specific stimuli that serve as landmarks,
- **memory**: using top-down information stored from past trips through the environment, and
- **combining** all this information to create cognitive maps to help relate what is perceived to where one currently is and wishes to go next.

Travel is the act of moving from one place to another and can be accomplished in many different ways (e.g., walking, swimming, driving, or for VR, pointing to fly). Travel can be seen as a continuum from passive transport to active transport.

Meta Platforms, Inc.
Exhibit 1063
Page 177

The extreme of passive transport occurs automatically without direct control from the user (e.g., immersive film). The extreme of active transport consists of physical walking and, to a lesser extent, interfaces that replicate human bipedal motion (e.g., treadmill walking or walking in place). Most VR travel is somewhere in between passive and active, such as point-to-fly or viewpoint control via a joystick. Various VR travel techniques are discussed in Section 28.3.

People rely on both optic flow (Section 9.3.3) and egocentric direction (direction of a goal relative to the body—see Sections 9.1.1 and 26.2) in a complementary manner as they travel toward a goal, resulting in robust locomotion control (Warren et al., 2001). When flow is reduced or distorted, behavior tends to be influenced more by egocentric direction. When there is greater flow and motion parallax, behavior tends to be more influenced by optical flow.

Many senses can contribute to the perception of travel including visual, vestibular, auditory, tactile, proprioceptive, and podokinetic (motion of the feet). Visual cues are the dominant modality for perceiving travel. Chapter 12 discusses how motion sickness can result when visual cues do not match other sensory modalities.

Meta Platforms, Inc.
Exhibit 1063
Page 178

Perception: Design Guidelines

Objective reality is not always how we perceive it to be and the mind does not always work in ways that we intuitively assume it works. For creating VR experiences, it is more important to understand how we perceive than it is for creating with any other digital medium. Studying human perception will help VR creators to optimize experiences, invent innovative techniques that are pleasant for humans, and troubleshoot perceptual problems as they arise.

General guidelines for applying concepts of human perception to creating VR experiences are organized below by chapter.

Objective and Subjective Reality (Chapter 6)

- Study perceptual illusions (Section 6.2) to better understand human perception, identify and test assumptions, interpret experimental results, better design VR worlds, and help understand/solve problems as they come up.
- Make sensory cues consistent in space and time across sensory modalities so that unintended illusions stranger than real-world illusions do not occur (Sections 6.2.4 and 7.2.1).

Perceptual Models and Processes (Chapter 7)

- Do not discount visceral (Section 7.7.1) and emotional (Section 7.7.4) processes. Use aesthetic intuition to drive initial attraction and positive emotions.
- Use both positive and negative feedback to minimize user frustration and to modify user behavior (Section 7.7.2).
- Make sure users end a VR experience on an emotional high—the end of the experience is what they are more likely to remember (Section 7.7.4).

Meta Platforms, Inc.
Exhibit 1063
Page 179

- Encourage users to reflect (Section 7.7.3) upon previous positive experiences by providing suggestive cues. For example, provide a review of the highlights of a user's performance at the end of the VR experience through a third-person view.
- Make interfaces intuitive by enabling complexity to be understood with the simplest mental model possible that can achieve the desired result, but not simpler (Section 7.8). This will help to minimize learned helplessness.
- Use signifying cues, feedback, and constraints to help the user form quality mental models and to make assumptions explicit (Sections 7.8 and 25.1).
- Do not fall into the trap of focusing only on one's own preferred sensory modality. Provide sensory cues across all modalities to cover a wider range of users (Section 7.9.2).
- Use consistency as users will form filters that generalize perception of events and interactions (Section 7.9.3).
- Provide cues to bring back memories in order to associate the experience with a previous real-life or virtual event (Section 7.9.3).
- Be aware that users make decisions early in a VR experience and they are more likely to stick with those decisions rather than change them (Section 7.9.3). Make it easy for users to decide that they like an experience from the very beginning.
- Define personas (Section 31.10) so that the VR experience can be designed to take advantage of the target audience's filters of general values, beliefs, attitudes, and memories (Section 7.9.3).

Perceptual Modalities (Chapter 8)

- VR creators should be very cognizant of what colors they choose as arbitrary colors can result in unintended consequences (Section 8.1.3).
- Use binaural cues with head pose to give a sense of sound location (Section 8.2.2).
- Use audio when precise timing (Section 8.2) is important and use visuals when precise location is important (Section 8.1.4).

Perception of Space and Time (Chapter 9)

- Think in terms of personal space, action space, and vista space when designing environments and interactions (Section 9.1.2).
- Consider how the relative importance of depth cues varies as a function of distance. Some depth cues are more relevant at different distances (Table 9.2).

Meta Platforms, Inc.
Exhibit 1063
Page 180

- Use a variety of depth cues to improve spatial judgments and enhance presence (Section 9.1.3).
- Use text located in space instead of 2D heads-up displays so the text has a distance from the user and occlusion is handled properly (Section 9.1.3).
- Do not place essential stimuli that must be consistently viewed close to the eye (Section 9.1.3).
- Do not forget that depth is not only a function of presented visual stimuli. Users perceive depth differently depending on their intentions and fear (Section 9.1.3).
- Be careful of having two visuals or sounds occur too closely in time as masking can cause the second stimulus to perceptually erase the earlier stimulus (Section 9.2.2).
- Be careful of sequence and timing of events. Changing the order can completely change the meaning (Section 9.2.1).

Perceptual Stability, Attention, and Action (Chapter 10)

- Include a sufficient number of distance cues and keep them consistent with each other so that size constancy, shape constancy, and position constancy are maintained (Section 10.1).
- Consider using primarily red colors and lighting for dark scenes to maintain dark adaptation while maintaining high visual acuity for foveal vision (Section 10.2.1).
- Don't expect users to notice or remember events just because they are within their field of view (Section 10.3.1). Use salience (e.g., a shiny/colorful object or a spatialized sound) to capture a person's attention (Section 10.3.2).
- Consider getting a user's attention first through spatialized audio to prepare them in advance for an event (Section 10.3.2).
- Attention can also be captured by objects that seem out of place and by putting objects where they expect them (Section 10.3.2).
- To increase performance, remove task-irrelevant stimuli (especially highly salient stimuli and when the task is easy). To make a task more challenging, consider adding distracting stimuli (Section 10.3.2).
- Collect data to build attention maps to determine what actually attracts users' attention (Section 10.3.2).
- To optimize searching, emphasize feature searching by making the searched-for objects distinct from surrounding stimuli (e.g., objects having angled lines

Meta Platforms, Inc.
Exhibit 1063
Page 181

where surrounding features don't include angled lines) so the searched-for items seem to "pop out" from a surrounding blur of irrelevant features (Section 10.3.2).

- To make a search task more challenging, emphasize conjunction searching by making the searched-for objects have particular combinations of features and including many similar objects (Section 10.3.2).
- To maximize flow, match difficulty with the individual's skill level (Section 10.3.2).
- Have computer-controlled characters perform interactions to help users learn the same interactions (Section 10.4.2).

Meta Platforms, Inc.
Exhibit 1063
Page 182

PART

ADVERSE HEALTH EFFECTS

Many users report feeling sick after experiencing VR-type experiences, and this is perhaps the greatest challenge of VR. As an example, over an eight-month period, six people over the age of 55 were taken to the hospital for chest pain and nausea after going on Walt Disney's "Mission: Space," a $100 million VR ride. This is the most hospital visits for a single ride since Florida's major theme parks agreed in 2001 to report any serious injuries to the state (Johnson, 2005). Disney also began placing vomit bags in the ride. More recently, John Carmack—legendary video game developer and CTO of Oculus—referenced his company's unwillingness to "poison the well" (i.e., to cause motion sickness on a massive scale) by introducing consumer VR hardware too quickly (Carmack, 2015).

Part III focuses on adverse health effects of VR and their causes. An **adverse health effect** is any problem caused by a VR system or application that degrades a user's health such as nausea, eye strain, headache, vertigo, physical injury, and transmitted disease. Causes include perceived self-motion through the environment, incorrect calibration, latency, physical hazards, and poor hygiene. In addition, such challenges can indirectly be more than a health problem. Users might adapt their behavior to avoid becoming sick, resulting in incorrect training for real-world tasks (Kennedy and Fowlkes, 1992). This is not just a danger for the user, as resulting aftereffects of VR usage (Section 13.4) can lead to public safety issues (e.g., vehicle accidents). We may never eliminate all the negative effects of VR for all users, but by understanding the problems and why they occur, we can design VR systems and applications to, at the very least, reduce their severity and duration.

Meta Platforms, Inc.
Exhibit 1063
Page 183

VR Sickness

Sickness resulting from VR goes by many names including motion sickness, cybersickness, and simulator sickness. These terms are often used interchangeably but are not completely synonymous, being differentiated primarily by their causes.

Motion sickness refers to adverse symptoms and readily observable signs that are associated with exposure to real (physical or visual) and/or apparent motion (Lawson, 2014).

Cybersickness is visually induced motion sickness resulting from immersion in a computer-generated virtual world. The word cybersickness sounds like it implies any sickness resulting from computers or VR (the word cyber refers to relating or involving computers and does not refer to motion), but for some reason cybersickness has been defined throughout the literature to be *only motion sickness* resulting from VR usage. That definition is quite limiting as it doesn't account for non-motion-sickness challenges of VR such as accommodation-vergence conflict, display flicker, fatigue, and poor hygiene. For example, a headache resulting from a VR experience due to accommodation-vergence conflict but no scene motion is not a form of cybersickness.

Simulator sickness is sickness that results from shortcomings of the simulation, but not from the actual situation that is being simulated (Pausch et al., 1992). For example, imperfect flight simulators can cause sickness, but such simulator sickness would not actually occur were the user in the real aircraft (although non-simulator sickness might occur). Simulator sickness most often refers to motion sickness resulting from apparent motion that does not match physical motion (rather that physical motion is actually zero motion, user motion, or motion created by a motion platform), but can also be caused by accommodation-vergence conflict (Section 13.1) and flicker (Section 13.3). Because simulator sickness does not include sickness that would result from the experience being simulated, sickness resulting from intense stressful games (e.g., that simulate life-and-death situations) is not simulator sickness. For example, acrophobia (a fear of heights that can cause anxiety) and vertigo resulting from looking down over a virtual cliff are not considered to be a form of simulator sickness (nor cybersickness). Some expert researchers also use the term simulator sickness to refer to sickness resulting from flight simulators implemented with world-fixed display and not sickness resulting from HMD usage (Stanney et al., 1997).

There is currently no generally accepted term that covers all sickness resulting from VR usage, and most users don't know or care about the specific terminology. A general term is needed that is not restricted by specific causes. Thus, this book tends to stay away from the terms cybersickness and simulator sickness, and instead uses the term **VR sickness** or simply "sickness" when discussing any sickness caused by using VR, irrespective of the specific cause of that sickness. Motion sickness is used only when specifically discussing motion-induced sickness.

Meta Platforms, Inc.
Exhibit 1063
Page 184

Overview of Chapters

Part III is divided into several chapters about the negative health effects of VR that VR creators should be aware of.

Chapter 12, Motion Sickness, discusses scene motion and how such motion can cause sickness. Several theories are described that explain why motion sickness occurs. A unified model of motion sickness is presented that ties together all the theories.

Chapter 13, Eye Strain, Seizures, and Aftereffects, discusses how non-moving stimuli can also cause discomfort and adverse health effects. Known problems include accommodation-vergence conflict, binocular-occlusion conflict, and flicker. Aftereffects are also discussed, which can indirectly result from scene motion as well as non-moving visual stimuli.

Chapter 14, Hardware Challenges, discusses physical challenges involved with the use of VR equipment. This chapter discusses physical fatigue, headset fit, injury, and hygiene.

Chapter 15, Latency, discusses in detail how unintended scene motion and motion sickness can result from latency. Since latency is a primary contributor to VR sickness, latency and its sources are covered in detail.

Chapter 16, Measuring Sickness, discusses how researchers measure VR sickness. Collecting data to determine users' level of motion sickness is important to help improve upon the design of VR applications and to make VR experiences more comfortable. Three methods of measuring VR sickness are covered: The Kennedy Simulator Sickness Questionnaire, postural stability, and physiological measures.

Chapter 17, Summary of Factors That Contribute to Adverse Effects, summarizes the primary contributors to adverse health effects resulting from VR exposure. The factors are divided into system factors, individual user factors, and application design factors. In addition, trade-offs of presence vs. motion sickness are also briefly discussed.

Chapter 18, Examples of Reducing Adverse Effects, provides some examples of specific techniques that are used to make users more comfortable and reduce negative health effects.

Chapter 19, Adverse Health Effects: Design Guidelines, summarizes the previous seven chapters and lists a number of actionable guidelines for improving comfort and reducing adverse health effects.

Meta Platforms, Inc.
Exhibit 1063
Page 185

Motion Sickness

Motion sickness refers to adverse symptoms and readily available observable signs that are associated with exposure to real and/or apparent motion (Lawson, 2014). Motion sickness resulting from apparent motion (also known as cybersickness) is the most common negative health effect resulting from VR usage. Symptoms of such sickness include general discomfort, nausea, dizziness, headaches, disorientation, vertigo, drowsiness, pallor, sweating, and, in the occasional worst case, vomiting (Kennedy and Lilienthal, 1995; Kolasinski, 1995).

Visually induced motion sickness occurs due to visual motion alone, whereas physically induced motion sickness occurs due to physical motion. Visually induced motion can be stopped by simply closing the eyes, whereas physically induced motion cannot.

Travel sickness is a form of motion sickness involving physical motion brought on by traveling in any type of moving vehicle such as a car, boat, or plane. Physically induced motion sickness can also occur from an amusement ride, spinning oneself, or playing on a swing. Although visually induced motion sickness may be similar to physically induced motion sickness, the specific causes and effects can be quite different. For example, vomiting occurs less often from simulator sickness than traditional motion sickness (Kennedy et al., 1993). Motion sickness can occur even when the user is not physically moving and results from numerous factors listed in Chapter 15.

This chapter discusses visual scene motion that occurs in VR, how vection relates to motion sickness, theories of motion sickness, and a unified model of motion sickness.

12.1 Scene Motion

Scene motion is visual motion of the entire virtual environment that would not normally occur in the real world (Jerald, 2009). A VR scene might be non-stationary due to

intentional scene motion injected into the system in order to make the virtual world behave differently than the real world (e.g., to virtually navigate through the world as discussed in Section 28.3), and

Meta Platforms, Inc.
Exhibit 1063
Page 186

unintentional scene motion caused by shortcomings of technology (where the scene motion often occurs only when moving the head), such as latency or inaccurate calibration (mismatched field of view, optical distortion, tracking error, incorrect interpupiliary distance, etc.).

Although intentional scene motion is well defined by code and unintentional scene motion is well defined mathematically (Adelstein et al., 2005; Holloway, 1997; Jerald, 2009), perception of scene motion and how it relates to motion sickness is not as well understood. Researchers do know that noticeable scene motion can degrade a VR experience by causing motion sickness, reducing task performance, lowering visual acuity, and decreasing the sense of presence.

12.2 Motion Sickness and Vection

Vection (an illusion of self-motion; Section 9.3.10) does not necessarily always cause motion sickness. For example, VR often evokes vection when the scene moves at a constant linear velocity, but such motion does not seem to be much of a factor for motion sickness. Yet, when the scene moves in other ways, motion sickness can be a major problem. Why is this the case?

Visual motion perception takes into account sensation from sensory modalities other than just the eyes. When a visual scene moves independently of how a user is physically moving, there can be mismatch between what is seen and what is physically felt. This mismatch is especially discontenting when the virtual motion accelerates because the otolith organs (Section 8.5) do not sense that same acceleration. The mismatch is not nearly as discontenting when the scene moves with constant velocity because the otolith organs do not sense velocity so there is nothing to compare the visual velocity to. Virtual rotations can be more discontenting because the semicircular canals detect both velocity and acceleration (Howard, 1986b)—one is more likely to experience motion sickness from constant angular velocity than from constant linear velocity.

Vection can be a powerful tool in VR to create a sense of travel. However, VR designers should be aware of the consequences of creating such a sense of travel—particularly linear accelerations of the entire scene and any type of angular motion. Contrary to what one might think would add to presence, adding artificial head bobbing should never be used as this can be significantly sickness inducing. Hill terrain and stairs can also be problematic, although to a lesser extent. Lateral movement can also be a problem, presumably because we do not often strafe in the real world (*Oculus Best Practices*, 2015).

Incorrect scene motion also often occurs in VR due to imperfect implementation. Although vestibular stimulation suppresses vection (Lackner and Teixeira, 1977),

Meta Platforms, Inc.
Exhibit 1063
Page 187

scene motion due to latency and head motion together can still cause vection and is a major contributor to motion sickness. Miscalibrated systems can also result in incorrect scene motion as users move their heads.

12.3 Theories of Motion Sickness

There are several theories as to why motion sickness occurs. When designing and evaluating VR experiences, it is useful to consider all these theories.

12.3.1 Sensory Conflict Theory

The **sensory conflict theory** (Reason and Brand, 1975) is the most widely accepted explanation for the initiation of motion sickness symptoms (Harm, 2002). The theory states that motion sickness may result when the environment is altered in such a way that incoming information across sensory modalities (primarily visual and vestibular) are not compatible with each other and do not match our mental model of expectations.

For VR, visual and auditory cues come from the simulation, whereas vestibular (Section 8.5) and proprioceptive (Section 8.4) cues come from body motion in the real world (motion platforms and haptics could be considered both simulation and real world). As a result, the visual and auditory synthesized cues are often in disagreement with proprioceptive and vestibular cues, and this disagreement is the primary cause of motion sickness. One of the great challenges of VR is to create visual and auditory cues that are consistent with vestibular and proprioceptive cues in order to minimize motion sickness.

The two primary conflicts at the root of motion sickness are the visual and vestibular senses—in most VR applications, the visual system senses motion whereas the vestibular does not. For example, when a user pushes forward on a game controller, it may look like she is moving forward but in reality, she is sitting in the same physical position. When these senses are inconsistent, then motion sickness may result.

12.3.2 Evolutionary Theory

The sensory conflict theory provides an explanation for motion sickness in certain situations, namely, those in which there is a conflict among the senses. However, the theory does not provide an explanation of why motion sickness occurs from an evolutionary point of view.

The **evolutionary theory** (Treisman, 1977), also known as the poison theory, offers a reason for why motion makes us sick: it is critical to our survival to properly perceive our body's motion and the motion of the world around us. If we get conflicting information from our senses, it means something is not right with our perceptual

Meta Platforms, Inc.
Exhibit 1063
Page 188

and motor systems. Our bodies have evolved to protect us by minimizing physiological disturbances produced by absorbed toxins. Protection occurs by

- discouraging movement (e.g., laying down until we recover),
- ejecting the poison via sweating and vomiting, and
- causing nausea and malaise (a slight or general feeling of not being healthy or happy) in order to discourage us from ingesting similar toxins in the future.

The emetic response associated with motion sickness may occur because the brain interprets sensory mismatch as a sign of intoxication, and triggers a nausea/vomiting defense program to protect itself.

12.3.3 Postural Instability Theory

Although the sensory conflict theory states that sickness will occur where there is a mismatch between experienced stimuli and expected stimuli, it does not accurately predict which situations will result in a mismatch or how severe sickness will be.

The **postural instability theory** predicts that sickness results when an animal lacks or has not yet learned strategies for maintaining postural stability (Riccio and Stoffregen, 1991). Riccio and Stoffregen suggest maintaining posture is one of the major goals of animals, and animals tend to become sick in circumstances where they have not learned strategies to maintain balance. They suggest people need to learn new patterns in novel situations to control their postural stability. Until this learning is completed, sickness may result. They do acknowledge there are problems with sensory stimulation in provocative situations, but argue such problems are caused by (not a cause of) postural instability; i.e., postural instability both precedes motion sickness and is necessary to produce symptoms. Furthermore, they argue the length of time one is unstable and the magnitude of that unstableness are predictors of motion sickness and the intensity of symptoms.

A person who thinks she is standing still is not completely motionless, but continuously and subconsciously adjusts muscles to provide the best balance. If the wrong muscle control is applied due to perceiving visual motion that is inconsistent with the physicality of the body, then postural instability occurs (LaViola, 2000). For example, if the visual scene is moving forward, users often lean forward to compensate (Badcock et al., 2014). Since the user is standing still instead of moving forward as perceived, then the leaning forward makes the user less stable; postural instability and motion sickness increase.

Meta Platforms, Inc.
Exhibit 1063
Page 189

Getting one's "sea legs" while traveling by boat to adjust to the boat's motion is an example of learning to cope (an adaptive mechanism) with postural stability in the real world. Similarly, VR users learn how to better control posture and balance over time, resulting in less motion sickness.

12.3.4 Rest Frame Hypothesis

Another criticism of the sensory conflict theory is there are some cue conflicts that do not provoke sickness. The rest frame hypothesis states that motion sickness does not arise from conflicting orientation and motion cues directly, but rather from conflicting stationary frames of reference implied by those cues (Prothero and Parker, 2003).

The rest frame hypothesis assumes the brain has an internal mental model of which objects are stationary and which are moving. The **rest frame** is the part of the scene that the viewer considers stationary and judges other motion relative to. In the real world, the visual background/context/ground is normally the rest frame. A room one is in is normally considered stationary, so rooms are almost always used as a basis for spatial orientation and motion. It is more intuitive to think about the motion of a ball with respect to a room than it is to think of the room moving with respect to the ball. To perceive motion, the brain first decides which objects are stationary (i.e., the rest frame). Motion of other objects and oneself is then perceived to move relative to this rest frame. As new information comes in through the senses, motion is also interpreted relative to the rest frame. When new incoming sensory motion cues do not fit the current mental model of the rest frame, motion sickness results. Motion sickness is inextricably tied to one's internal mental model of what should be stable.

Motion between an object and the self can be interpreted as either moving relative to the other. For example, an object moving to the right can also be perceived as the body moving to the left. If the object is considered the rest frame, then it is considered stable and any motion between the object and the self must be due to the self moving, and vection can result.

The rest frame hypothesis allows for the existence of sensory conflicts without causing motion sickness if those conflicting cues are not essential to the stability of the rest frame. However, if cues that are considered to be part of the rest frame are in conflict, then sickness can result. For example, when a user virtually moves through a world, the ground moves beneath his feet. If the ground is considered the rest frame that does not move, then it must be the user that is moving. Vestibular cues inform the brain that he is not actually moving over the ground, so his mental model of the world becomes confused by the conflicting information and motion sickness results.

Meta Platforms, Inc.
Exhibit 1063
Page 190

For vision, humans have a bias toward certain things, such as meaningful parts of the scene and background, as being stationary (Section 20.4.2). For example, larger backgrounds are considered stable and small stimuli less stable as demonstrated by the autokinetic effect (Section 6.2.7).

VR creators should make rest frame cues consistent whenever possible rather than being overly concerned with making all orientation and all motion consistent. The visuals in VR can be divided into two components: one component being content and the other component being a rest frame that matches the users' physical inertial environment (Duh et al., 2001). For example, since users are heavily influenced by backgrounds to serve as a rest frame, making the background consistent with vestibular cues, even if other parts of the scene move, can reduce motion sickness. Even when the background can't be made consistent with vestibular cues, then smaller foreground cues can help to serve as a rest frame. Houben and Bos (2010) created an "anti-seasickness display" for use in enclosed ship cabins where no horizon is visible. The display served as an earth-fixed rest frame that matched vestibular cues better than the rest of the room. Subjects experienced less sickness due to ship motion as a result of the display. Such a rest frame can be thought of as an inverse form of augmented reality where real-world spatially fixed visual cues (even if they are computer-generated) that match vestibular cues are brought into the virtual world (vs. adding virtual-world cues into the real world as is done in augmented reality). Such techniques can be used to reduce VR motion sickness as discussed in Section 18.2.

Motion sickness is much less of an issue for augmented reality optical-see-through HMDs (but not camera-see-through HMDS) because users can directly see the real world, which acts as a rest frame consistent with vestibular cues. Motion sickness is also reduced when the real world can be seen in the periphery of an HMD.

12.3.5 Eye Movement Theory

The eye movement theory states that motion sickness occurs due to the unnatural eye motion required to keep the scene's image stable on the retina. If the image moves differently than expected, such as often occurs in VR, then a conflict occurs between what the eyes expect and what actually occurs. The eyes then must move differently than they do in the real world in order to stabilize the image on the retina. As a result of this discrepancy, motion sickness results. See Section 7.4 for a quick demonstration of how the real world can seem to move, causing a slight sense of vection due to the eye moving by an abnormal method.

The visual and vestibular systems are strongly connected as demonstrated by eye movements that result from the vestibulo-ocular reflex (VOR) and visual suppression

Meta Platforms, Inc.
Exhibit 1063
Page 191

of the VOR through the optokinetic reflex (OKR) (Section 8.1.5). Optokinetic nystagmus has been proposed (Ebenholtz et al., 1994) to affect the vagal nerve, a mixed nerve that contains both sensory and motor neurons (Siegel and Sapru, 2014). Such innervations lead to nausea and emesis.

Studies have shown that when observers fixate on a point to reduce eye movement induced by larger moving stimuli (i.e., reduction of optokinetic nystagmus, also discussed in Section 8.1.5), then vection is enhanced. Such a stationary fixation point may also result in users selecting that point as a rest frame (Keshavarz et al., 2014). Thus, supplying fixation points for VR users to focus upon and maintain eye stability may help to reduce motion sickness.

12.4 A Unified Model of Motion Sickness

This section describes a unified model of motion perception and motion sickness that includes the true state of the world and self, sensory input, the estimated state of the world and self, actions, afferent and efferent signals, and the internal mental model of the individual (Razzaque, 2005). Integrating these various components together is not just important for perceiving motion but is critical for survival, navigation, balance/posture, and stabilizing the eyes so they function properly. The model is consistent with all the previously described theories of motion sickness and can be used to help understand how we perceive motion, whether that motion is perceived as external motion of the world or self-motion, and why motion sickness results. Figure 12.1 shows the relationship of the various components, which are discussed below.

12.4.1 The True State of the World

The "truth" of the world (which includes oneself) at any given moment can be considered a state describing everything in it. For this model, the state is the objective motion between the external world (distal stimuli) and the physical self.

12.4.2 Sensory Input

A person's perception of motion is largely a function of information coming in through multiple senses (proximal stimuli transduced into afferent nerve impulses). Humans rely on visual, auditory, vestibular, proprioceptive, and tactile information in order to maintain balance and orientation and to distinguish self-motion from external motion.

Meta Platforms, Inc.
Exhibit 1063
Page 192

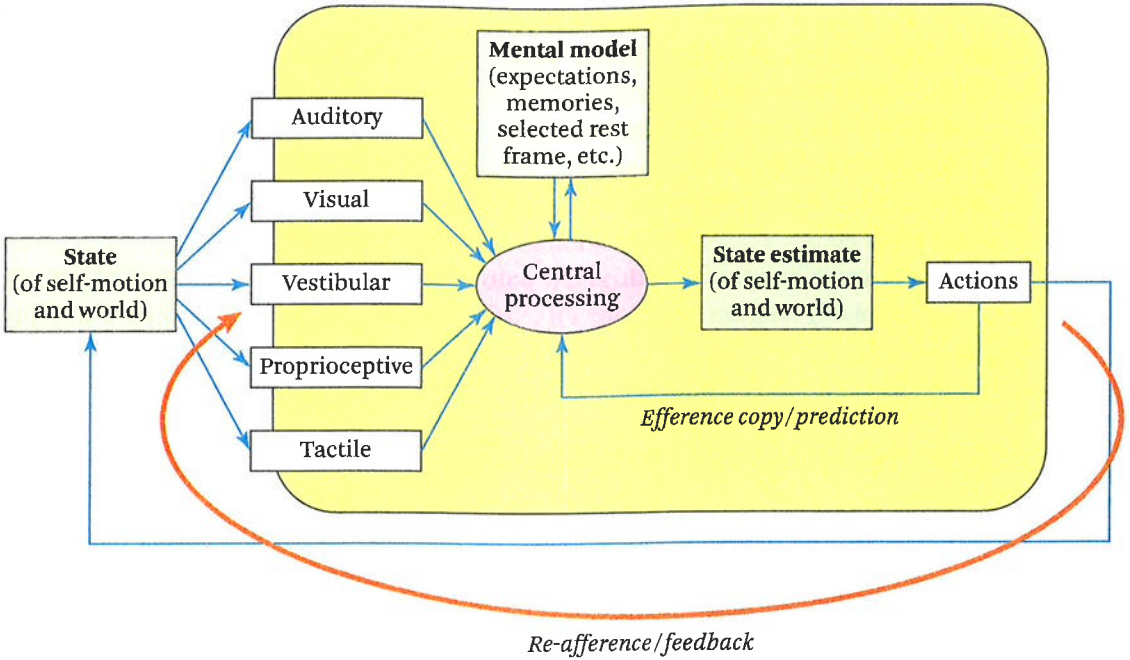


Figure 12.1 **A unified model of motion sickness.** (Adapted from Razzaque (2005)

12.4.3 Central Processing

Brain processes integrate information from multiple senses to perceive what is happening (bottom-up processing). Any single sensory modality has limitations, so the brain integrates information from multiple senses together to form a better estimate of the world and the self. In fact, even information from all the senses is not enough (and sometimes not consistent) to accurately estimate both. Thus, additional input to central processing is needed. Additional input comes from an internal mental model (top-down processing) as well as prediction of sensory input from one's own actions (efference copy).

12.4.4 State Estimation

At any given moment, the brain contains an estimate of the world and the self. Interaction among the various sensory cues helps to disambiguate whether motion is a result of the world moving or the self moving through that world. One's estimate of this motion is constantly tested and revised as a function of central processing. A state estimate that is incorrect can result in visual illusions such as vection. An

Meta Platforms, Inc.
Exhibit 1063
Page 193

inconsistent or unstable state estimate (e.g., vestibular cues do not match visual cues) can result in motion sickness. If the state estimate changes to be consistent with itself even though input has not changed, then perceptual adaptation (Section 10.2.2) has occurred. The state estimate includes not only what is currently happening but what is likely to happen.

12.4.5 Actions

The body continuously attempts to self-correct through physical actions in order to fit its state estimate. Several physical actions can occur when attempting to self-correct: (1) one attempts to stabilize himself by changing posture, (2) the eyes rotate to try to stabilize the virtual world's unstable image on the retina through the vestibulo-ocular and optokinetic reflexes, and (3) physiological responses occur, such as sweating and, in extreme cases, vomiting, to remove what the body thinks may be toxins causing the confused state estimate.

Once the mind and body have learned or adapted to new ways of taking action (e.g., a new navigation technique), then action, prediction, feedback, and the mental model become more consistent resulting in less motion sickness.

12.4.6 Prediction and Feedback

The perception of a visually stable world relies on prediction of how incoming sensory information will change due to one's action. For example, as one turns the head and eyes to the left, then copies of outgoing efferent nerve signals (Section 7.4) are sent to central processing that predict a left vestibular response, left-turning neck proprioception, and optical flow moving to the right. If this efference copy matches incoming afferent signals from the senses, then the mind perceives any changes as coming from one's own actions (the afferent signals are then called re-afference since the afference reconfirms that the intended action has taken place). If the efference copy and afference do not match, then the observer perceives the stimulus to have occurred from a change in the external world (e.g., the entire world appears to move in a strange way) instead of from one's actions.

At a higher level, drivers and pilots who control their vehicle motion rarely exhibit motion sickness although the same individuals sometimes report motion sickness when they are passengers (Rolnick and Lubow, 1991). Such results are consistent with voluntary active movement initiating perceptual processes differently than passive movements (Section 10.4) and re-afferent feedback. Similar to driving, VR users are less likely to get sick if they actively control their viewpoint.

Meta Platforms, Inc.
Exhibit 1063
Page 194

12.4.7 Mental Model of Motion

Afference and efference are not perfect and cannot completely explain position constancy and self-motion. There is noise in muscles, external forces sometimes constrain motion, eye saccades cannot be completely predicted, etc.

Each of us has an internal mental model (Section 7.8) of how the world works based on previous experience and future expectations. New incoming information is evaluated in terms of this model, rather than continuously constructing new models from scratch. When new information does not match the model, then disorientation and confusion can result. A part of a person's mental model is transient—the mental model is often reevaluated and relearned. Other parts are more permanent/hardwired assumptions. For example, larger and further-away objects are more likely to be perceived as a rest frame, which can result in induced motion (Section 9.3.5) of smaller objects when the larger, further-away objects are in fact not stationary. Another assumption is consistency—one assumes object properties (Section 10.1) do not change over a short period of time. Breaking these assumptions can lead to strong illusions, even when one cognitively knows exactly which assumptions are false—the perceptual system does not always agree with the rational thinking cortex (Gregory, 1973).

The internal mental model has expectations of how one's actions will affect the world and how stimuli will move. When one acts, the new incoming perceptual cues, efference copy, and mental model are all compared, and the mental model is verified, refined, or invalidated. This is consistent with the rest frame hypothesis—when the mental motion model is invalidated by new sensory cues, the rest frame seems to move and motion sickness can result.

Meta Platforms, Inc.
Exhibit 1063
Page 195

Eye Strain, Seizures, and Aftereffects

Non-moving visual stimuli can also cause discomfort and adverse health effects. Known problems include accommodation-vergence conflict, binocular-occlusion conflict, and flicker. Aftereffects are also discussed, which can indirectly result from scene motion as well as non-moving visual stimuli.

13.1 Accommodation-Vergence Conflict

In the real world, accommodation is closely associated with vergence (Section 9.1.3), resulting in clear vision of closely located objects. If one focuses on a near object, the eyes automatically converge and if one focuses on a distant object, the eyes automatically diverge resulting in a sharp single fused image. The HMD **accommodation-vergence conflict** occurs due to the relationship between accommodation and vergence not being consistent with what occurs in the real world. Overriding the physiologically coupled oculomotor processes of vergence and accommodation can result in eye fatigue and discomfort (Banks et al., 2013).

13.2 Binocular-Occlusion Conflict

Binocular-occlusion conflict occurs when occlusion cues do not match binocular cues, e.g., when text is visible but appears at a distance behind a closer opaque object. This can be quite confusing and uncomfortable for the user.

Many desktop first-person shooter/perspective games contain a 2D overlay/heads-up display to provide information to users (Section 23.2.2). If such games are ported to VR, then it is essential that either such information be removed or the overlay should be given a depth such that information is occluded properly when behind scene geometry.

Meta Platforms, Inc.
Exhibit 1063
Page 196

13.3 Flicker

Flicker (Section 9.2.4) is distracting and can cause eye fatigue, nausea, dizziness, heachache, panic, confusion, and in rare cases seizures and loss of consciousness (LaViola, 2000; Rash, 2004). As discussed in Section 9.2.4, the perception of flicker is a function of many variables. Trade-offs in HMD design can result in flicker becoming more perceivable (e.g., wider field of view and less display persistence), so it is essential to maintain high refresh rates so that flicker does not cause health problems. Flicker, however, is not only a function of the refresh rate of the display. Flashing lights in a virtual environment can also cause flicker, and it is important that VR creators be careful not to create such flashing. Dark scenes can also be used to increase dark adaptation, which reduces the perception of flicker.

A **photic seizure** is an epileptic event provoked by flashing lights that can occur in situations ranging from driving by a picket fence in the real world to flashing stimuli in VR, and occurs in about 0.01% of the population (Viirre et al., 2014). Those experiencing seizures generally show a brief period of absence where they are awake but do not respond. Such seizures do not always result in convulsions so are not always obvious. Repeated seizures can lead to brain injury and a lower threshold for future episodes. Reports of the frequency at which seizures most commonly occur vary widely from a range of 1 to 50 Hz, likely due to different conditions. Because the conditions that will most likely cause seizures within a specific HMD or application are not known, flickering or flashing of lights at any rate should be avoided. Anyone with a history of epilepsy should not use VR.

13.4 Aftereffects

Immediate effects during immersive experiences are not the only dangers of VR usage. Problems may persist after returning to the real world. A **negative aftereffect** (Section 10.2.3) of VR usage is any adverse effect that occurs after VR usage but was not present during VR usage. Such aftereffects include perceptual instability of the world, disorientations, and flashbacks.

About 10% of those using simulators experience negative aftereffects (Johnson, 2005). Those who experience the most sickness during VR exposure usually experience the most aftereffects. Most aftereffect symptoms disappear within an hour or two. The persistence of symptoms longer than six hours has been documented but fortunately remains statistically infrequent.

Kennedy and Lilienthal (1995) compared effects of simulator sickness to effects of alcohol intoxication using postural stability measurements after VR usage and

Meta Platforms, Inc.
Exhibit 1063
Page 197

suggested to restrict subsequent activities in a similar way that is done for alcohol intoxication. There is a danger that such effects could contribute to problems that involve more than the individual, such as vehicular accidents. In fact, many VR entertainment centers require that users not drive for at least 30–45 min after exposure (LaViola, 2000). We, as the VR community, should do the same in doing our part to not only minimize VR sickness but also educate those not familiar with VR sickness of potentially dangerous aftereffects.

13.4.1 Readaptation

As described in Section 10.2.1, a perceptual adaptation is a change in perception or perceptual-motor coordination that serves to reduce or eliminate sensory discrepancies. A **readaptation** is an adaptation back to a normal real-world situation after adapting to a non-normal situation. For example, after a sea voyage, many travelers need to readapt to land conditions, something that can last weeks or even months (Keshavarz et al., 2014).

Over time VR users adapt to shortcomings of VR. After adapting to those shortcomings, when the user returns to the real world, a readaptation occurs due to the real world being treated as something novel compared to the VR world to which the user has adapted. An example of adaptation and readaptation occurs when the rendered field of view is not properly calibrated to match the physical field of view. This causes the image in the HMD to appear distorted. Because of this distortion, when the user turns the head everything but the center of the display will seem to move in an odd way. Once a user becomes accustomed to this odd motion and the world seems stable, then adaptation has occurred. Unfortunately, after adaptation to the incorrect field of view, the inverse may very well occur when the user returns to the real world. That is, the person might perceive an inverse distortion and the world to not appear stable when turning the head. Such adaptation (called position-constancy adaptation; Section 10.2.2) and readaptation have been found to occur with minifying/magnifying glasses (Wallach and Kravitz, 1965a) and HMDs where the rendered field of view and physical field of view don't match (Draper, 1998).

Until the user has readapted to the real world, symptoms may persist, such as drowsiness, disturbed locomotor and postural control, and lack of hand-eye coordination (Kennedy et al., 2014; Keshavarz et al., 2014).

Two approaches are possible for helping with readaptation: natural decay and active readaptation (DiZio et al., 2014). **Natural decay** is the refrainment of any activity such as relaxing with little movement and the eyes closed. This approach can be less

Meta Platforms, Inc.
Exhibit 1063
Page 198

sickness inducing but can prolong the readaptation period. **Active readaptation** involves the use of real-life targeted activities aimed at recalibrating the sensory systems affected by the VR experiences, such as hand-eye coordination tasks. Such activity might be more sickness inducing but can speed up the readaptation process.

Fortunately, as stated in Section 10.2.2, users are able to adapt to multiple conditions (called dual adaptation). Expert users who frequently go back and forth between a VR experience and the real world may have less of an issue with aftereffects and other forms of VR sickness.

Meta Platforms, Inc.
Exhibit 1063
Page 199

Hardware Challenges

It is important to consider physical issues involved with the use of VR equipment. This chapter discusses physical fatigue, headset fit, injury, and hygiene.

14.1 Physical Fatigue

Physical fatigue can be a function of multiple causes including the weight of worn/held equipment, holding unnatural poses, and navigation techniques that require physical motion over an extended period of time.

Some HMDs of the 1990s weighed as much as 2 kg (Costello, 1997). Fortunately, the weights of HMDs have decreased and are not as much of an issue as they used to be. HMD weight is more of a concern when the HMD center of mass is far from the head center of mass (especially in the horizontal direction). If the HMD mass is not centered, then the neck must exert extra torque to offset gravity. In such a case, the worst problems occur when moving the head. This can tire the neck, which can lead to headaches. The offset mass not only can cause fatigue and headaches but may also modify the way in which the user perceives distance and self-motion (Willemsen et al., 2009).

Controller weight is typically not a major issue. However, some devices have their center of mass offset more than others. For such devices, trade-offs should be carefully considered before finalizing the decision on selecting controllers.

In the real world, humans naturally rest their arms when working. In many VR applications, there is not as much opportunity to rest the arms on physical objects. **Gorilla arm** is arm fatigue resulting from extended use of gestural interfaces without resting the arm(s). This typically starts to occur within about 10 seconds of having to hold the hand(s) up above the waist in front of the user. Line-of-sight issues should be considered when selecting hand-tracking devices (Section 27.1.9) so users are able to work comfortably with their hands in their lap and/or to the side of their body. Interaction techniques should be designed to not require the hands to be high and in front of the user for more than a few seconds at a time (Section 18.9).

Meta Platforms, Inc.
Exhibit 1063
Page 200

Some forms of navigation via walking (Section 28.3.1) can be exhausting after a period of time, especially when users are expected to walk long distances. Even real walking can be tiring for some users, and often they will ask for a seat after some period of time. Sitting options should be provided for longer sessions, and the experience might be optimized for sitting with only occasional need to stand up and walk around. Weight becomes more of an issue for walking interfaces.

14.2 Headset Fit

Headset fit refers to how well an HMD fits a user's head and how comfortable it feels. HMDs often cause discomfort due to pressure or tightness at contact points between the HMD and head. Such pressure points vary depending on the HMD and user but typically occur around the eye sockets and on the ears, nose, forehead, and/or back of the head. Not only can this pressure be uncomfortable on the skin, but it can cause headaches. Headset fit can especially be an issue for those who wear eyeglasses.

Loose-fitting HMDs are also a concern as anything mounted on the head (unless bolted into the skull!) will wiggle if jolted enough because the human head has skin on it and the skin is not completely taut. When the user moves her head quickly, the HMD will move slightly. Such HMD slippage can aggravate the skin and cause small scene motions. Slippage can be reduced by making lighter HMDs but will not completely go away. The small motions can be compensated for by calculating the amount of wiggle and moving the scene appropriately to remain stable relative to the eyes.

Most HMDs are able to be physically adjusted, and users should make sure to adjust the HMD at the start of use as discomfort can increase over time and users are less likely to adjust when they are fully engaged in the experience. Such adjustments do not, however, guarantee a quality fit for all users. Although basic head sizes can be used for different people, head sizes vary considerably more than along a single measurement. For optimal fit, scanning of the head can be used for customized HMD design (Melzer and Moffitt, 2011). Other options currently being investigated by Eric Greenbaum (personal communication, April 29, 2015) of Jema VR (and creator of About Face VR hygiene solutions) include pump-inflatable liners, ergonomic inserts designed to spread the force of the HMD along the most robust parts of the user's face (forehead and cheekbones), and different forms of foam.

14.3 Injury

Various types of injury are a risk in fully immersive VR due to multiple factors.

Physical trauma is an injury resulting from a real-world physical object that is an increased risk with VR due to being blind and deaf to the real world (e.g., trauma resulting from falling or hitting physical objects). Sitting is preferred from a safety

Meta Platforms, Inc.
Exhibit 1063
Page 201

perspective as standing can be more risky due to the potential of colliding with real-world walls/objects/cables and tripping or losing balance. Hanging the wires from the ceiling can reduce problems if the user does not rotate too much. A human **spotter** should closely watch a standing or walking user to help stabilize them if necessary. Harnesses, railing, or padding might also be considered. HMD cables can also be dangerous as there is a possibility of minor whiplash if a short or tangled/stuck cable pulls on the head when the user makes a sudden movement. Even if physical injury does not occur, something as simple as softly bumping against an object or a cable touching the skin can cause a break-in-presence. Active haptic devices can be especially dangerous if the forces are too strong. Hardware designers should design the device so the physical forces are not able to exceed some maximum force. Physical safety mechanisms can also be integrated into the device.

Repetitive strain injuries, such as tendonitis, fibrous tissue hyperplasia, and ligamentous injury, result from carrying out prolonged repeated activities using rapid carpal and metacarpal movements (Costello, 1997). Such injuries have been reported with using standard input devices such as mice, gamepads, and keyboards. Any interaction technique that requires continual repetitive movements is undesirable (as it is for a real-world task). Properly designed VR interfaces have the potential for less repetitive strain injuries than traditional computer mouse/keyboard input due to not being constrained to a plane. Fine 3D translation when holding down a button is stressful (Paul Mlyniec, personal communication, April 28, 2015), so button presses are sometimes performed by the non-translating hand at the cost of the interaction not being as intuitive or direct.

Noise-induced hearing loss is decreased sensitivity to sound that can be caused by a single loud sound or by lesser audio levels over a long period of time. Permanent ear damage can occur with continuous sound exposure (over 8 h) that has an average level in excess of 85 dB or with a single impulse sound in excess of 130 dB (Sareen and Singh, 2014). Ideally, sounds should not be continuously played and maximum audio levels should be limited to prevent ear damage. Unfortunately, limiting maximum levels can be difficult to guarantee as developers cannot fully control specific systems of individual users. For example, some users may have their own sound system and speakers or have their computer volume gain set differently. However, audio levels can be set to a maximum in software. For example, as users move close to a sound source, code can confirm the audio resulting from the simulation does not exceed a specified level.

14.4 Hygiene

VR hardware is a **fomite**—an inanimate physical object that is able to harbor pathogenic organisms (e.g., bacteria, viruses, fungi) that can serve as an agent for transmitting diseases between different users of the same HMD (Costello, 1997). Hygiene

Meta Platforms, Inc.
Exhibit 1063
Page 202

becomes more important as more users don the same equipment. Such challenges are not unique to VR, as traditional interfaces such as keyboards, mice, telephones, and restrooms have similar issues, albeit interfaces close to or on the face may be more of an issue.

The skin on the face produces oils and sweats. Further, facial skin is populated by a variety of beneficial and pathogenic bacteria. Heat from the electronics and lack of ventilation around the eyes can add to sweating. This may result in an unpleasant user experience and in the most extreme cases can present health risks. Kevin Williams (2014) divides hygiene into wet, dry, and unusual:

- **Wet** includes sweat, oils, and makeup/hair products.
- **Dry** includes skin, scalp, and hair.
- **Unusual** includes lice/eggs and earwax.

Some professionals clean their 3D polarized glasses with an ultraviolet sterilization process, but unfortunately such a process breaks down plastic material (known as UV degradation) and thus the process can only be used so many times before making the hardware unusable (Eric Greenbaum, personal communication, April 29, 2015). The film industry primarily uses industrial washing machines to clean their glasses. Neither of these methods is appropriate for HMDs (with possibly the exception of the cheaper cell-phone holder HMDs).

Although wiping down the system with alcohol can be helpful, this does not necessarily remove all dangerous biological agents as it can be difficult to clean cracks and porous material (material often used where the skin touches the HMD). Metal, plastic, and rubber are more resistant than porous materials and are often used, but such impermeable material is not as breathable or comfortable for users. The lenses should be wiped with a nonabrasive microfiber cloth with a mild detergent/antibacterial soap and dried before reuse. An inner cleanable cap for each individual user that prevents direct contact with the top of the head, as used by Disney, can be used to further reduce risk (Pausch et al., 1996; Mine, 2003). For audio, earbuds should not be used for more than one user. Instead larger earphones should be used that cover the entire ear and can be easily wiped clean.

Eric Greenbaum has created multiple removable-from-the HMD solutions to the hygiene problem through About Face VR. About Face VR provides several options of placing material between the HMD and face that drastically reduces transfer of biological agents. Options include personal textile pads/liners that are washed as part of one's laundry, and impermeable liners that can be wiped down with alcohol or other disinfectant for public use.

Meta Platforms, Inc.
Exhibit 1063
Page 203



Figure 14.1 **An About Face VR ergonomic insert (center gray) and removable/washable line (blue).** (Image courtesy of Jema VR)

If using controllers, provide hand sanitizer for users to sanitize their hands before and after use. Be prepared for the worst case of users becoming ill by keeping sick bags, plastic gloves, mouthwash, drinks, light snacks, air freshener, and cleaning products nearby. Keep such items out of view as such cues can cause users to think about getting sick, which can cause them to get sick.

Meta Platforms, Inc.
Exhibit 1063
Page 204

Latency

A fundamental task of a VR system is to present worlds that have no unintended scene motion even as the user's head moves. Today's VR applications often produce spatially unstable scenes, most notably due to latency.

Latency is the time a system takes to respond to a user's action, the true time from the start of movement to the time a pixel resulting from that movement responds (Jerald, 2009). Note depending on display technology different pixels can show at different times and with different rise and fall times (Section 15.4). Prediction and warping (Section 18.7) can reduce effective latency where the goal is to stabilize the scene in real-world space although the actual latency is greater than zero.

15.1 Negative Effects of Latency

For low latencies below ~100 ms, users do not perceive latency directly, but rather the consequences of latency—a static virtual scene appears to be unstable in space when users move their heads (Jerald, 2009). Latency in an HMD-based system causes visual cues to lag behind other perceptual cues (e.g., visual cues get out of phase with vestibular cues), creating sensory conflict. With latency and some head motion, the visual scene presented to a subject in an HMD moves incorrectly. This unintended scene motion (Section 12.1) due to latency is known as "swimming" and has serious usability consequences. VR latency is a major contributor to motion sickness, and it is essential for VR creators to understand latency in order to minimize it.

In addition to causing scene motion and being a primary cause of motion sickness, latency has other negative effects as described below.

15.1.1 Degraded Visual Acuity

Latency can cause degraded vision. Given some latency, as an HMD user moves her head then stops, the scene's image relative to the head is still moving when the head has stopped moving. If the image velocity on the retina is greater than 2–3°/s, then motion blur and degraded visual acuity result. Typical head motions and latencies result in scene motion greater than 3°/s. For example, a sinusoidal head motion

Meta Platforms, Inc.
Exhibit 1063
Page 205

of 0.5 Hz at ±20° and 133 ms of latency results in a peak scene motion of ±8.5°/s (Adelstein et al., 2005). It is not known if users' eyes tend to follow a lagging image of the scene, resulting in no retinal image slip, or if their eyes tend to stay stabilized in space, resulting in retinal image slip.

15.1.2 Degraded Performance

The level of latency necessary to negatively impact performance may be different from the level of latency that can be perceived. So and Griffin (1995) studied the relationship between latency and operator learning in an HMD. The task consisted of tracking a target with the head. Training did not improve performance when latency was ≥ 120 ms—subjects were unable to learn to compensate for these latencies in the task.

15.1.3 Breaks-in-Presence

Latency detracts from the sense of presence in HMDs (Meehan et al., 2003). Latency combined with head movement causes the scene to move in a way not consistent with the real world. This incorrect scene motion can distract the user, who might otherwise feel present in the virtual environment, and cause her to realize the illusion is only a simulation.

15.1.4 Negative Training Effects

A negative training effect is an unintended decrease in performance that results from training for a task. Latency has been shown to result in negative training effects with desktop displays (Cunningham et al., 2001a) and driving simulators utilizing large screens (Cunningham et al., 2001b).

15.2 Latency Thresholds

Often engineers build HMD systems with a goal of "low latency" without specifically defining what that "low latency" is. There is little consensus among researchers what latency requirements should be. Ideally, latency should be low enough so that users are not able to perceive scene motion. Latency thresholds decrease as head motion increases—latency is much easier to detect when making quick head movements.

Various experiments conducted at NASA Ames Research Center (Adelstein et al., 2003; Ellis et al., 1999, 2004; Mania et al., 2004) reported latency thresholds during quasi-sinusoidal head yaw. They found absolute thresholds to vary by 85 ms due to bias, type of head movement, individual differences, differences of experimental conditions, and other known and unknown factors. Surprisingly, they found no dif-

Meta Platforms, Inc.
Exhibit 1063
Page 206

ferences in latency thresholds for different scene complexities, ranging from single, simple objects to detailed, photorealistic rendered environments. They found just-noticeable differences to be more consistent at ~4–40 ms and that users are just as sensitive to changes in latency with a low base latency as those with a higher base latency. Consistent latency is important even when average latency is high.

After working with NASA on measuring motion thresholds during head turns (Adelstein et al., 2006), Jerald (2009) built a VR system with 7.4 ms of latency and measured HMD latency thresholds for various conditions. He found his most sensitive subject to be able to discriminate between latency differences as small as 3.2 ms. Jerald also developed a mathematical model relating latency, scene motion, head motion, and latency thresholds and then verified that model through psychophysical measurements. The model demonstrates that even though our sensitivity to scene motion decreases as head motion increases (Section 9.3.4), our sensitivity to latency increases. This is because as head motion increases, latency-induced scene motion increases more quickly than scene motion sensitivity decreases.

Note the above results apply to fully immersive VR. Optical-see-through displays have much lower latency thresholds (under 1 ms) due to users being able to directly discriminate between real-world cues and the synthesized cues (i.e., judgments are object-relative instead of subject-relative—see Section 9.3.2).

15.3 Delayed Perception as a Function of Dark Adaptation

The **Pulfrich pendulum effect** is a depth illusion that occurs when one eye is dark adapted (Section 10.2.1) by a different amount than the other eye (Gregory, 1973) or when one eye is covered with a dark filter (Arditi, 1986). A pendulum swinging in a plane orthogonal to the line of sight appears to swing in an ellipse (i.e., a greater or less distance from the observer at the bottom of its arc, when maximum velocity is reached) instead of a flat arc.

The "dark" eye trades its acuity in space and time for increased light sensitivity. The dark eye's retina integrates the incoming light over a longer period of time resulting in a time delay. The dark eye perceives the pendulum to be further in the past, and thus further behind its true position than the light eye. As the pendulum speeds up in the middle of the arc, the dark eye sees its position further and further behind the position seen by the light eye. This difference of effective position creates the illusion of an ellipse lying in depth. This illusion of depth is shown in Figure 15.1.

Delayed perception for dark adaption brings up the question of how/if we perceive latency in VR differently for different lighting conditions. The differences in delay between a light and dark environment can be up to 400 ms (Anstis, 1986). This delay

Meta Platforms, Inc.
Exhibit 1063
Page 207

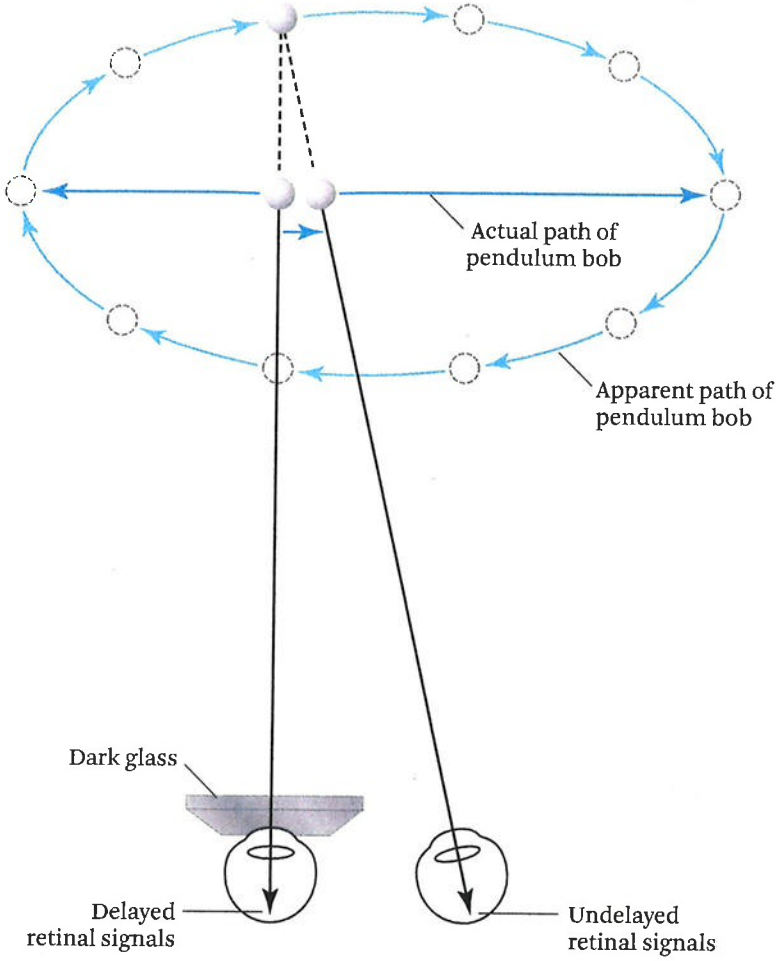


Figure 15.1 **The Pulfrich pendulum effect: A pendulum swinging in a straight arc across the line of sight appears to swing in an ellipse when one eye is dark adapted due to the longer delay of the dark-adapted eye.** (Adapted from Gregory (1973))

produces a lengthening of reaction time for automobile drivers in dim light (Gregory, 1973). Given that visual delay varies for different amounts of dark adaptation or stimulus intensity, then why do people not perceive the world to be unstable or experience motion sickness when they experience greater delays in the dark or with sunglasses as they do in delayed HMDs? Two possible answers to this question are as follows (Jerald, 2009).

Meta Platforms, Inc.
Exhibit 1063
Page 208

- The brain might recognize stimuli to be darker and calibrates for the delay appropriately. If this is true, this suggests users can adapt to latency in HMDs. Furthermore, once the brain understands the relationship between wearing an HMD and delay, position constancy could exist for both the delayed HMD and the real world; the brain would know to expect more delay when an HMD is being worn. However, the brain has had a lifetime of correlating dark adaptation and/or stimulus intensity to delay. The brain might take years to correlate the wearing of an HMD to delay.
- The delay inherent in dark adaptation and/or low intensities is not a precise, single delay. Stimuli 1 ms in duration can be perceived to be as long as 400 ms in duration (Section 9.2.2). This imprecise delay results in motion smear (Section 9.3.8) and makes precise localization of moving objects more difficult. Perhaps observers are biased to perceive objects to be more stable in such dark situations, but not for the case of brighter HMDs. If this is true, perhaps darkening the scene or adding motion blur to the scene would cause it to appear to be more stable and reduce sickness (although computing motion blur typically adds latency unless prediction is used).

15.4 Sources of Delay

Miné (1993) and Olano et al. (1995) characterize system delays in VR systems and discuss various methods of reducing latency. **System delay** is the sum of delays from tracking, application, rendering, display, and synchronization among components. Note the term system delay is used here because some consider latency to be the effective delay that can be less than system delay, accomplished by using delay compensation techniques (Section 18.7). Thus, system delay is equivalent to true latency, rather than effective latency. Figure 15.2 shows how the various delays contribute to total system delay.

Note that system delay is greater than the inverse of the update rate; i.e., a pipelined system can have a frame rate of 60 Hz but have a delay of several frames.

15.4.1 Tracking Delay

Tracking delay is the time from when the tracked part of the body moves until movement information from the tracker's sensors resulting from that movement is input into the application or rendering component of the VR system. Tracking products can include techniques that complicate delay analysis. For example, many tracking systems incorporate filtering to smooth jitter. If filters are used, the resulting output pose is only partially determined by the most recent tracker reading, so that precise delay

Meta Platforms, Inc.
Exhibit 1063
Page 209

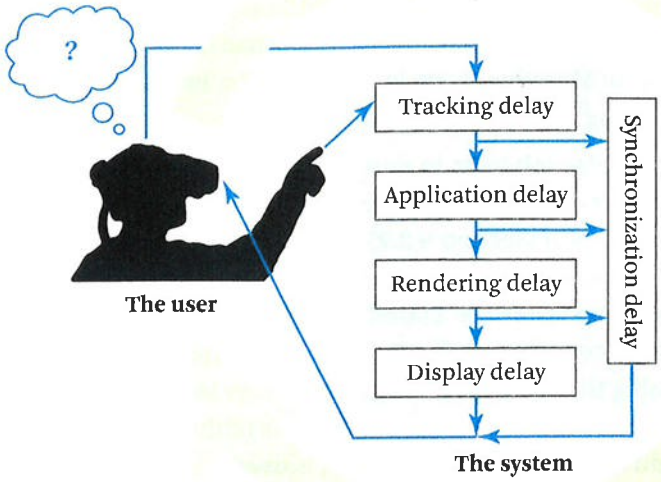


Figure 15.2 **End-to-end system delay. End-to-end system delay comes from the delay of the individual system components and from the synchronization of those components.**

is not well defined. Some trackers use different filtering models that are selected depending on the current motion estimate for different situations—delay during some movements may differ from that during other movements. For example, the 3rdTech HiBall tracking system allows the option of using multi-modal filtering. A low-pass filter is used to reduce jitter if there is little movement, whereas a different model is used for larger velocities.

Tracking is sometimes processed on a different computer from the computer that executes the application and renders the scene. In that case network delay can be considered to be a part of tracking delay.

15.4.2 Application Delay

Application delay is the time from when tracking data is received until the time data is passed onto the rendering stage. This includes updating the world model, computing the results of a user interaction, physics simulation, etc. This application delay can

Meta Platforms, Inc.
Exhibit 1063
Page 210

vary greatly depending on the complexity of the task and the virtual world. Application processing can often be executed asynchronously from the rest of the system (Bryson and Johan, 1996). For example, a weather simulation with input from remote sources could be delayed by several seconds and computed at a slow update rate, whereas rendering needs to be tightly coupled to head pose with minimal delay. Even if the simulation is slow, the user should be able to naturally look around and into the slow updating simulation.

15.4.3 Rendering Delay

A **frame** is a full-resolution rendered image. **Rendering delay** is the time from when new data enters the graphics pipeline to the time a new frame resulting from that data is completely drawn. Rendering delay depends on the complexity of the virtual world, the desired quality of the resulting image, the number of rendering passes, and the performance of the graphics software/hardware. The **frame rate** is the number of times the system renders the entire scene per second. **Rendering time** is the inverse of the frame rate, and in non-pipelined rendering systems is equivalent to rendering delay.

Rendering is normally performed on graphics hardware in parallel with the application. For simple scenes, current graphics cards can achieve frame rates of several thousand hertz. Fortunately, rendering delay is what content creators and software developers have the most control over. If geometry is reasonable, the application is optimized, and high-end graphics cards are used, then rendering will be a small portion of end-to-end delay.

15.4.4 Display Delay

Display delay is the time from when a signal leaves the graphics card to the time a pixel changes to some percentage of the intended intensity defined by the graphics card output. Various display technologies are used for HMDs. These include CRTs (cathode ray tubes), LCDs (liquid crystal displays), OLEDs (organic light-emitting diodes), DLP (digital light processing) projectors, and VRD (virtual retinal displays). Different display technologies also have different advantages and disadvantages. See Jerald (2009) for a summary of these display technologies as they relate to delay. The basics of general display delays are discussed below.

Refresh Rate

The **refresh rate** is the number of times per second (Hz) that the display hardware draws a full frame. Note this can be different from the frame rate described in Section 15.4.3. The **refresh time** (also known as stimulus onset asynchrony—see

Meta Platforms, Inc.
Exhibit 1063
Page 211

Section 9.3.6) is the inverse of the refresh rate (seconds per frame). A display with a refresh rate of 60 Hz has a refresh time of 16.7 ms. Typical displays have refresh rates from 60 to 120 Hz.

Double Buffering

In order to avoid memory access issues, the display should not read the frame at the same time it is being written to by the renderer. The frame should not scan out pixels until the rendering is complete, otherwise geometric primitives may not be occluded properly. Furthermore, rendering time can vary, depending on scene complexity, implementation, and hardware, whereas the refresh rate is set solely by the display hardware.

A solution to this problem of dual access to the frame can be solved by using a **double-buffer** scheme. The display processor renders to one buffer while the refresh controller feeds data to the display from an alternate buffer. The **vertical sync** signal occurs between frames. Most commonly, the system waits for this vertical sync to swap buffers. The previously rendered frame is then scanned out to the display while a newer frame is rendered. Unfortunately, waiting for vertical sync causes additional delay, since rendering must wait up to 16.7 ms (for a 60 Hz display) before starting to render a new frame.

Raster Displays

A **raster display** sweeps pixels out to the display, scanning out left to right in a series of horizontal scanlines from top to bottom (Whitton, 1984). This pattern is called a raster. Timings are precisely controlled to draw pixels from memory to the correct locations on the screen.

Pixels on raster displays have inconsistent intra-frame delay (i.e., different parts of the frame have different delay), because pixels are rendered from a single time-sampled viewpoint but are presented at different times; if the system waits on vertical sync, then the bottommost pixels are presented nearly a full frame time after the top-most pixels are presented.

Tearing. If the system does not wait for vertical sync to swap buffers, then the buffer swap occurs while the frame is being scanned out to the display hardware. In this case, **tearing** occurs during viewpoint or object motion and appears as a spatially discontinuous image, due to two or more frames (each rendered from a different sampled viewpoint) contributing to the same displayed image. When the system waits for vertical sync to swap buffers, no tearing is evident, because there is a single rendering for the entire frame.

Meta Platforms, Inc.
Exhibit 1063
Page 212

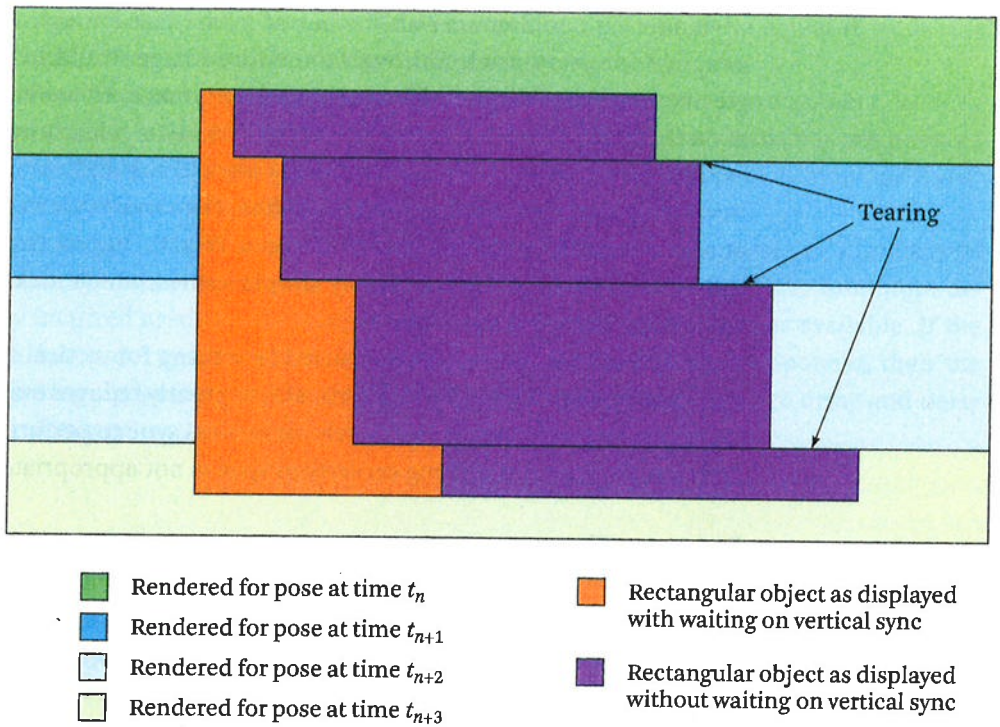


Figure 15.3 **A frame showing a rectangular object as the user is looking from right to left with and without waiting for vertical sync to swap buffers. Not waiting for vertical sync to swap buffers causes image tearing.**

Figure 15.3 shows a simulated image that would occur with a system that does not wait for vertical sync superimposed over a simulated image that would occur with a system that does wait for vertical sync. The figure shows what a static virtual block would look like when a user is turning her head from right to left. The tearing is obvious when the swap does not wait for vertical sync, due to four renderings of the images with four different head poses. Thus, most current VR systems avoid tearing at the cost of additional and variable intra-frame delay.

Just-in-time pixels. Tearing decreases with decreasing differences of head pose. As the sampling rate of tracking increases and the frame rate increases, pose coherence increases and the tearing becomes less evident. If the system were to render each pixel with the correct up-to-date viewpoint, then the tearing would occur between pixels. The tearing would be small compared to the pixel sizes, resulting in a smooth image—without perceptual tearing. Miné and Bishop (1993) call this **just-in-time pixels.**

Meta Platforms, Inc.
Exhibit 1063
Page 213

Rendering could conceivably occur at a rate fast enough that buffers would be swapped for every pixel. Although the entire image would be rendered, only a single pixel would be displayed for each rendered image. However, a 1280 × 1024 image at 60 Hz would require a frame rate of over 78 MHz—clearly impossible for the foreseeable future using standard rendering algorithms and commodity hardware. If a new image were rendered for every scanline, then the rendering would need to occur at approximately 1/1000 of that or at about 78 kHz. In practice, today's systems can render at rates up to 20 kHz for very simple scenes, which make it possible to show a new image every few scanlines.

With some VR systems, delay caused by waiting for vertical sync can be the largest source of system delay. Ignoring vertical sync greatly reduces overall delay at the cost of image tearing. Note some HMDs rely on vertical sync to perform delay compensation (Section 18.7), so not waiting on vertical sync is not appropriate for all hardware.

Response Time

Response time is the time it takes for a pixel to reach some percentage of its intended intensity. Each technology behaves differently with respect to pixel response time. For example, liquid crystals take time to turn resulting in slow response times—in some cases over 100 ms! Slow-response displays make it impossible to define a precise delay. Typically, a percentage of the intended intensity is defined, although even that is often not precise due to response time being a function of the starting and intended intensity on a per pixel basis.

Persistence

Display persistence is the amount of time a pixel remains on a display before going away. Many displays hold pixel values/intensities until the next refresh and some displays have a slow dissipation time, causing pixels to remain visible even after the system attempts to turn off those pixels. This results in delay not being well defined. Is the delay up to the time that the pixel first appears or up to the average time that the pixel is visible?

A slow response time and/or persistence on the display can appear as motion blur and/or "ghosting." Some systems only flash the pixels for some portion of the refresh time (e.g., some OLED implementations) in order to reduce motion blur and judder (Section 9.3.6).

15.4.5 Synchronization Delay

Total system delay is not simply a sum of component delays. **Synchronization delay** is the delay that occurs due to integration of pipelined components. Synchronization delay is equal to the sum of component delays subtracted from total system delay.

Meta Platforms, Inc.
Exhibit 1063
Page 214

Synchronization delay can be due to components waiting for a signal to start new computations and/or asynchrony among components.

Pipelined components depend upon data from the previous component. When a component starts a new computation and the previous component has not updated data, then old data must be used. Alternatively, the component can in some cases wait for a signal or wait for the input component to finish.

Trackers provide a good example of a synchronization problem. Commercial tracker vendors report their delays as the tracker response time—the minimum delay incurred assumes the tracker data is read as soon as it becomes available. If the tracker is not synchronized with the application or rendering component, then the tracking update rate is also a crucial factor and affects both average delay and delay consistency.

15.5 Timing Analysis

This section presents an example timing analysis and discusses complexities encountered when analyzing system delay. Figure 15.4 shows a timing diagram for a typical VR system where each colored rectangle represents a period of time that some piece of information is being created. In this example, the scanout to the display begins just after the time of vertical sync. The display stage is shown for discrete frames, even though typical displays present individual pixels at different times. The image delays are the time from the beginning of tracking until the start of scanout to the display. Response time of the display is not shown in this diagram.

As can be seen in the figure, asynchronous computations often give repeated images when no new input is available. For example, the rendering component cannot start computing a new result until the application component provides new information. If new application data is not yet available, then the rendering stage repeats the same computation.

In the figure, the display component displays frame n with the results from the most up-to-date rendering. All the component timings happen to line up fairly well for frame n, and image i delay is not much more than the sum of the individual component delays. Frame $n + 1$ repeats the display of an entire frame because no new data is available when starting to display that frame. Frame $n + 1$ has a delay of an additional frame time more than the image i delay because a newly rendered frame was not yet available when display began. Frame $n + 4$ is delayed even further due to similar reasons. No duplicate data is computed for frame $n + 5$, but image $i + 2$ delay is quite high because the rendering and application components must complete their previous computations before starting new computations.

Meta Platforms, Inc.
Exhibit 1063
Page 215

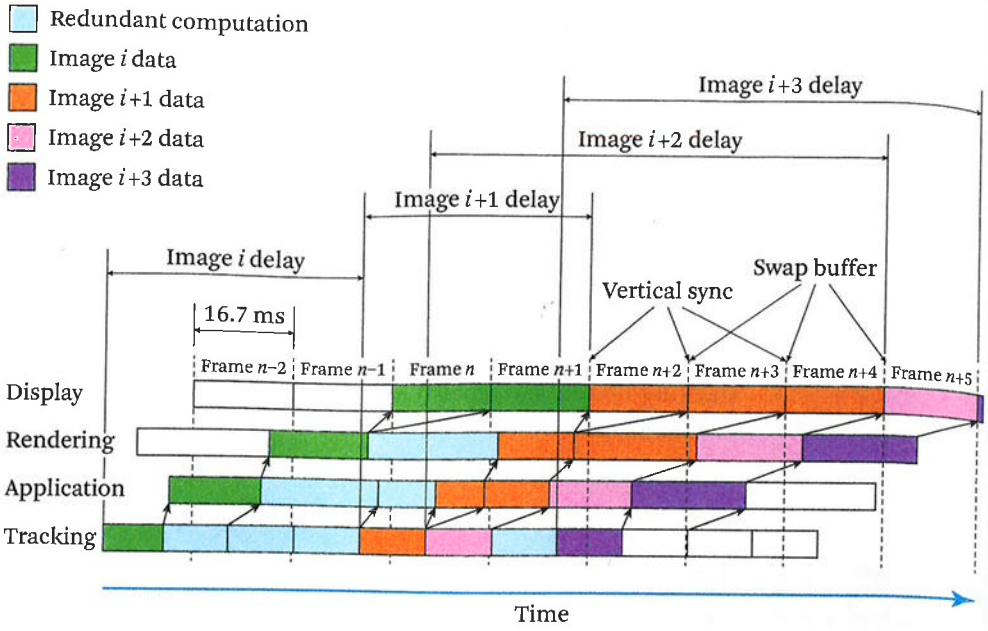


Figure 15.4 **A timing diagram for a typical VR system. In this non-optimal example, the pipelined components execute asynchronously. Components are not able to compute new results until preceding components compute new results themselves (Jerald, 2009).**

15.5.1 Measuring Delays

To better understand system delay, one can measure timings not only for the end-to-end system delay but also for sub-components of the system. Means and standard deviations can be derived from several such measurements. Latency meters can be used to measure system delay. For an example of an open-source, open-hardware latency meter, see Taylor (2015).

Timings can be further analyzed by sampling signals at various stages of the pipeline and measuring the time differences. The parallel port on PCs can be used to output timing signals. These signals are precise in time since there is no additional delay due to a protocol stack; writing to the parallel port is equivalent to writing to memory.

Synchronization delays between two adjacent components of the pipeline can also be measured indirectly. If the delays of individual components are known, then the sum of two adjacent components can be compared with the measured delay across both components. The difference is the synchronization delay between the two components.

Meta Platforms, Inc.
Exhibit 1063
Page 216

Measuring Sickness

VR sickness can be difficult to measure. One reason for this is VR sickness is polysymptomatic and cannot be measured by a single variable (Kennedy and Fowlkes, 1992). Another challenge is the large variance between individuals. When effects of VR sickness do occur, they are typically small and weak impacts that disappear quickly upon completing the experience. Because many participants eventually adapt, testing the same participants can be a problem. Thus, researchers often choose to use between-subject study designs (Section 34.4.1). These factors of large individual differences, weak effects, adaptation, and between-subject designs invariably lead to the conclusion that large sample sizes are required. A difficulty with experimenting over such a broad range of users is keeping experiment control conditions consistent.

Questionnaires or symptom checklists are the usual means of measuring VR sickness. These methods are relatively easy to do and have a long history of usage, but are subjective and rely on the participant's ability to recognize and report changes. Postural stability tests and physiological measures are occasionally used due to their objective nature.

16.1 The Kennedy Simulator Sickness Questionnaire

The Kennedy Simulator Sickness Questionnaire (SSQ) was created from analyzing data from 1,119 users of 10 US Navy flight simulators. Twenty-seven commonly experienced symptoms were identified that resulted in the creation and validation of 16 symptoms (Kennedy et al., 1993). The 16 symptoms were found to cluster into three categories: oculomotor, disorientation, and nausea. The oculomotor cluster includes eye strain, difficulty focusing, blurred vision, and headache. The disorientation cluster includes dizziness and vertigo. The nausea cluster includes stomach awareness, increased salivation, and burping. When taking the questionnaire, participants rank each of the 16 symptoms on a 4-point scale as "none," "slight," "moderate," or "severe." The SSQ results in four scores: total (overall) sickness and three subscores of oculomotor, disorientation, and nausea.

Meta Platforms, Inc.
Exhibit 1063
Page 217

The Kennedy SSQ has become a standard for measuring simulator sickness, and the total severity factor may reflect the best index of simulator sickness for a VR application. Although the three subscales are not independent, they can provide diagnostic information as to the specific nature of the sickness in order to provide clues as to what needs improvement in the VR application. These scores can be compared to scores from other similar applications, or can be used as a measure of application improvement over time.

Contrary to suggestions by Kennedy et al. (1993), the SSQ has often been given before exposure to the VR experience in order to provide a baseline to compare the SSQ given after the experience. Young et al. (2007) found that when pre-questionnaires suggest VR can make users sick, then the amount of sickness those users report after the VR experience increases. Thus, it is now standard practice to only give a simulator sickness questionnaire after the experience to remove bias, although this comes at the cost of not having a baseline to compare with and risks users not being aware that VR can cause sickness. Those not in their usual state of good health and fitness before starting the VR experience should not use VR and definitely should not be included in the SSQ analysis.

16.2 Postural Stability

Motion sickness can be measured behaviorally through postural stability tests. Postural stability tests can be separated into those testing static and dynamic postural stability. A common static test is the Sharpened Romberg Stance: one foot in front of the other with heel touching toe, weight evenly distributed between the legs, arms folded across the chest, chin up. The number of times a subject breaks the stance is the postural instability measure (Prothero and Parker, 2003). There are currently several commercial systems to assess body sway.

16.3 Physiological Measures

Physiological measures can provide detailed data over the entire VR experience. Examples of physiological data that changes as sickness occurs include heart rate, blink rate, electroencephalography (EEG), and stomach upset (Kim et al., 2005). Changes in skin color and cold sweating (best measured via skin conductance) also occur with increased sickness (Harm, 2002). The direction of change for most physiological measures of sickness is not consistent across individuals, so measuring absolute changes is recommended. Little research currently exists for physiological measures of VR sickness, and there is a need for more cost-effective and objective physiological measures of both resulting VR sickness and a person's susceptibility to it (Davis et al., 2014).

Meta Platforms, Inc.
Exhibit 1063
Page 218

Summary of Factors That Contribute to Adverse Effects

Adverse health effects due to sensory incongruences do not only occur from VR. About 90% of the population have experienced some form of travel sickness and about 1% of the population get sick from riding in automobiles. (Lawson, 2014). About 70% of naval personnel experience seasickness, with the greater problems occurring on smaller ships and rougher seas. Up to 5% of those traveling by sea fail to adapt through the duration of the voyage, and some may show effects weeks after the end of the voyage (this is known as land sickness), resulting in postural instability and perceived instability of the visual field during self-movement. VR systems of the 1990s resulted in as many as 80%–95% of users reporting sickness symptoms, and 5%–30% had symptoms severe enough to discontinue exposure (Harm, 2002). Fortunately, extreme cases of vomiting from VR is not as bad as real-world motion sickness—75% of those suffering from seasickness vomit whereas vomiting in simulators is rare at 1% (Kennedy et al., 1993; Johnson, 2005).

This chapter summarizes contributing factors to adverse health effects, several of which have been previously discussed. It is extremely important to understand these in order to create the most comfortable experiences and to minimize sickness. Contributing factors and interaction between these factors can be complex—some may not affect users until combined with other problems (e.g., fast head motion may not be a problem until latency exceeds some value). The contributing factors are divided into groups of system factors, individual user factors, and application design factors. Each group of factors is an important contributor to VR sickness, as without one of them there would be no adverse effects (e.g., turn off the system, provide no content, or remove the user), but then there would be no VR experience. These lists were created from a number of publications, the author's personal experience, discussions

Meta Platforms, Inc.
Exhibit 1063
Page 219

with others, and editing/additions from John Baker (personal communication, May 27, 2015) of Chosen Realities LLC and former chief engineer of the army's Dismounted Soldier Training System, a VR/HMD system that trained over 65,000 soldiers with over a million man-hours of use.

17.1 System Factors

System factors of VR sickness are technical shortcomings that technology will eventually solve through engineering efforts. Even small optical misalignment and distortions may cause sickness. Factors are listed in approximate order of decreasing importance.

Latency. Latency in most VR systems causes more error than all other errors combined (Holloway, 1997) and is often the greatest cause of VR sickness. Because of this, latency is discussed in extensive detail in Chapter 15.

Calibration. Lack of precise calibration can be a major cause of VR sickness. The worst effect of bad calibration is inappropriate scene motion/scene instability that occurs when the user moves her head. Important calibration includes accurate tracker offsets (the world-to-tracker offset, tracker-to-sensor offset, sensor-to-display offset, display-to-eye offset), mismatched field-of-view parameters, misalignment of optics, incorrect distortion parameters, etc. Holloway (1995) discusses error due to miscalibration in detail.

Tracking accuracy. Low-accuracy head trackers result in seeing the world from an incorrect viewpoint. Inertial sensors used in phones result in yaw drift over time, causing heading to become less accurate over time. A magnetometer can sense the earth's magnetic field to reduce drift, but rebar in floors, I-beams in tall buildings, computers/monitors, etc. can cause distortion to vary as the user changes position. Error in tracking the hands does not cause motion sickness, although such error can cause usability problems.

Tracking precision. Low precision results in jitter, which for head tracking is perceived as shaking of the world. Precision/jitter can be reduced via filtering at the cost of latency.

Lack of position tracking. VR systems that only track orientation (i.e., not position) result in the virtual world moving with the user when the user translates. For example, as the user leans down to pick something off the floor, the floor extends downward. Users can learn to minimize changing position, but even subtle changes in position can contribute to VR sickness. If no position tracking is

Meta Platforms, Inc.
Exhibit 1063
Page 220

available, calibrate the sensor-to-eye offset to estimate motion parallax as the user rotates her head.

Field of view. Displays with wide field of view result in more motion sickness due to (1) users being more sensitive to vection in the periphery, (2) the scene moving over a larger portion of the eyes, and (3) scene motion (either unintentional due to other shortcomings or intentional by design—e.g., when navigating through a world). A perfect wide field-of-view VR system with no scene motion would not cause motion sickness.

Refresh rate. High display refresh rates can reduce latency, judder, and flicker. Ideally refresh rates should be as fast as possible.

Judder. Judder is the appearance of jerky or unsmooth visual motion and is discussed in Section 9.3.6.

Display response time and persistence. Display response time is the time it takes for a pixel to reach some percentage of its intended intensity (Section 15.4.4). Persistence is the time that pixels remain on the display after reaching their intended intensity. Such timings have trade-offs of judder, motion smear/blur, flicker, and latency—all of which are a problem, so these should all be considered when choosing hardware.

Flicker. Flicker (Sections 9.2.4 and 13.3) is distracting, induces eye fatigue, and can cause seizures. As luminance increases, more flicker is perceived. Flicker is also more perceivable with wider-field-of-view HMDs. Displays that have longer response times and longer persistence (e.g., LCDs) can reduce flicker at the cost of adding motion blur, latency, and/or judder.

Vergence/accommodation conflict. For conventional HMDs, accommodative distance remains constant whereas vergence does not (Section 13.1). Visuals should not be placed close to the eye, and if they are then only for short periods of time to reduce strain on the eyes.

Binocular images. HMDs can present monocular (one image for a single eye), biocular (identical images for each eye), or binocular images (two different images for each eye that provide a sense of depth) (Section 9.1.3). Incorrect binocular images can result in no stereopsis and eye strain.

Eye separation. For many VR systems, the inter-image distance, inter-lens distance, and interpupillary distance are often in conflict with each other. This conflict can give rise to visual discomfort symptoms and muscle imbalances in users' eyes (Costello, 1997).

Meta Platforms, Inc.
Exhibit 1063
Page 221

Real-world peripheral vision. HMDs that are open in the peripherey where users can see the real world reduce vection due to the real world acting as a stable rest frame (Section 12.3.4). Theoretically, this factor would not make a difference for an HMD that is perfectly calibrated.

Headset fit. Improperly fitting HMDs (Section 14.2) can cause discomfort and pressure, which can lead to headaches. For those who wear eyeglasses, HMDs that push up against the glasses can be very uncomfortable or even painful. Improper fit also results in larger HMD slippage and corresponding scene motion.

Weight and center of mass. A heavy HMD or HMD with an offset center of mass (Section 14.1) can tire the neck, which can lead to headaches and also modify the way distance and self-motion are perceived.

Motion platforms. Motion platforms can significantly reduce motion sickness if implemented well (Section 18.8). However, if implemented badly then motion sickness can increase due to adding physical motion cues that are not congruent with visual cues.

Hygiene. Hygiene (Section 14.4) is especially important for public use. Bad smells are a problem as nobody wants a stinky HMD and bad smells can cause nausea.

Temperature. Reports of discomfort go up as temperature increases beyond room temperature (Pausch et al., 1996). Isolated heat near the eyes can occur due to lack of ventilation.

Dirty screens. Dirty screens can cause eye strain due to not being able to see clearly.

17.2 Individual User Factors

Susceptibility to VR sickness is polygenic (a function of multiple genes, although specific genes contributing to sickness are not yet known), and perhaps the greatest factor for VR sickness is the individual. In fact, Lackner (2014) claims the range of vulnerability varies by a factor of 10,000 to 1. Some people immediately exhibit all the signs of VR sickness within moments of using a VR system, others exhibit only a few signs after some period of use, and some exhibit no signs of sickness after extended use for all but the most extreme conditions. Almost anyone exposed to provocative physical body motion or who is exposed to dramatic inconsistencies between the vestibular and visual systems can to some extent be made sick. The only exception is those who have a total loss of vestibular function (Lawson, 2014).

Not only is there variability to susceptibility of VR sickness between individuals, but individuals are affected differently by different causes. For example, lateral (e.g.,

Meta Platforms, Inc.
Exhibit 1063
Page 222

strafing) motions make some users uncomfortable (*Oculus Best Practices,* 2015) where other users have no problem whatsoever with such motions, even though they are susceptible to other provocative motions. Other users seem more susceptible to problems caused by accommodation-vergence conflict—perhaps this is related to the decreased ability to focus with age. Furthermore, new users are almost always more susceptible to motion sickness than more experienced users. Because of these differences, different users often have very different opinions of what is important for reducing sickness. One way to deal with this is by providing configuration options so that users can choose between different configurations to optimize their comfort. For example, a smaller field of view can be more appropriate for new users.

Even though there are dramatic differences between individuals, three key factors seem to affect how much individuals get sick (Lackner, 2014):

- sensitivity to provocative motion,
- the rate of adaptation, and
- the decay time of elicited symptoms.

These three key factors may explain why some individuals are better able to adapt to VR over a period of time. For example, a person with high sensitivity but with a high rate of adaptation and short decay time can eventually experience less sickness than an individual with moderate sensitivity but a low rate of adaptability and long decay time.

Listed below are descriptions of more specific individual factors that correlate to VR sickness in approximate order of decreasing importance.

Prior history of motion sickness. In the behavioral sciences, past behavior is the best predictor of future behavior. Evidence suggests it is also the case that an individual's history of motion sickness (physically induced or visually induced) predicts VR sickness (Johnson, 2005).

Health. Those not in their usual state of health and fitness should not use VR as unhealthiness can contribute to VR sickness (Johnson, 2005). Symptoms that make individuals more susceptible include hangover, flu, respiratory illness, head cold, ear infection, ear blockage, upset stomach, emotional stress, fatigue, dehydration, and sleep deprivation. Those under the influence of drugs (legal or illegal) or alcohol should not use VR.

VR experience. Due to adaptation, increased experience with VR generally leads to decreased incidence of VR sickness.

Meta Platforms, Inc.
Exhibit 1063
Page 223

Thinking about sickness. Users who think about getting sick are more likely to get sick. Suggesting to users that they may experience sickness can cause them to dwell on adverse effects, making the situation worse than if they were not informed (Young et al., 2007). This causes an ethical dilemma as not warning users can be a problem as well (e.g., users should be aware of sickness so they can make an informed decision and know to stop the experience if they start to feel ill). Users should be casually warned that people experience some "discomfort" and not to try to "tough it out" if they start to experience any adverse effects, but do not overemphasize sickness so they are less likely to become overly concerned.

Gender. Females are as much as three times more susceptible to VR sickness than males (Stanney et al., 2014). Possible explanations include hormonal differences, field-of-view differences (females have a larger field of view which is associated with greater VR sickness), and biased self-report data (males may tend to underreport symptoms).

Age. Susceptibility to physically induced motion sickness is greatest at ages 2–12 (kids are the ones who get sick in a car), decreases rapidly from ages 12–21, and then decreases more slowly thereafter (Reason and Brand, 1975). VR sickness, however, increases with age (Brooks et al., 2010). The reason for the difference of age effects on physically induced motion sickness vs. VR sickness is not clear, but it could be due to confounding factors such as experience (flight hours for flight simulators and video game experience for immersive entertainment), excitement, lack of the ability to accommodate, concentration level, balance capability, etc.

Mental model/expectations. Users are more likely to get sick if scene motion does not match their expectations (Section 12.4.7). For example, gamers accustomed to first-person shooter games may have different expectations than what occurs with commonly used VR navigation methods. Most first-person shooter games have the view direction, weapon/arm, and forward direction always pointing in the same direction. Because of this, gamers often expect their bodies to go whereever their head is pointing, even if walking on a treadmill. When navigation direction does not match viewing/head direction, their mental model can break down and confusion along with sickness can result (sometimes called Call of Duty Syndrome).

Interpupillary distance. The distance between the eyes of most adults ranges from 45 to 80 mm and goes as low as 40 mm for children down to five years

Meta Platforms, Inc.
Exhibit 1063
Page 224

of age (Dodgson, 2004). VR systems should be calibrated for each individual's interpupillary distance. Otherwise, eye strain, headaches, general discomfort, and associated problems can result.

Not knowing what looks correct. Simply telling someone to put an HMD on doesn't mean that it is properly secured on the head and in the correct position. The user's eyes may not be in the sweet spots, the headset can be at an odd angle, the headset can be loose, etc. Educating users how to put on the headset and asking the user to look at objects and stating "adjust the headset until comfortable and the objects in the display are crisp and clear" can help reduce negative effects.

Sense of balance. Postural instability (Section 12.3.3) correlates well with motion sickness. Pre-VR postural stability is most strongly associated with the nausea and disorientation subscale of the Kennedy SSQ (Kolasinski, 1995).

Flicker-fusion frequency threshold. A person's flicker-fusion frequency threshold is the flicker frequency where flicker becomes visually perceptible (Section 9.2.4). There is a wide individual variability in flicker-fusion threshold across individuals due to time of day, gender, age, and intelligence (Kolasinski, 1995).

Real-world task experience. More experienced pilots (i.e., those who have more real flight hours) are more susceptible to simulator sickness (Johnson, 2005). This is likely due to having higher sensitivity to expectations of how the real world should behave for the task being simulated.

Migraine history. Individuals with a history of migraines have a tendency to experience higher levels of VR sickness (Nichols et al., 2000).

There are likely several other individual user factors that contribute to VR sickness. For example, it has been hypothesized that ethnicity, concentration, mental rotation ability, and field independence may correlate with susceptibility to VR sickness (Kolasinski, 1995). However, there has been little conclusive research so far to support these claims.

17.3 Application Design Factors

Even if all technical issues were perfectly solved (e.g., zero latency, perfect calibration, and unlimited computing resources), content can still cause VR sickness. This is due to the human physiological hardware of our bodies and brains instead of the technical limitations of human-made hardware. Note users can drive content via head motions

Meta Platforms, Inc.
Exhibit 1063
Page 225

and navigation techniques so those actions are included here. Factors are listed in approximate order of decreasing importance.

Frame rate. A slow frame rate contributes to latency (Section 15.4.3). Frame rate is listed here as an application factor vs. a system factor as frame rate depends on the application's scene complexity and software optimizations. A consistent frame rate is also important. A frame rate that jumps back and forth between 30 Hz and 60 Hz can be worse than a consistent frame rate of 30 Hz.

Locus of control. Actively being in control of one's navigation reduces motion sickness (Section 12.4.6). Motion sickness is often less for pilots and drivers than for passive co-pilots and passengers (Kolasinski, 1995). Control allows one to anticipate future motions.

Visual acceleration. Virtually accelerating one's viewpoint can result in motion sickness (Sections 12.2 and 18.5), so the use of visual accelerations should be minimized (e.g., head bobbing that simulates walking causes oscillating visual accelerations and should never be used). If visual accelerations must be used, the period of acceleration should be as short as possible.

Physical head motion. For imperfect systems (e.g., a system with latency or inaccurate calibration), motion sickness is significantly reduced when individuals keep their head still, due to those imperfections only causing scene motion when the head is moved (Section 15.1). If the system has more than ~30 ms of latency, content should be designed for infrequent and slow head movement. Although head motion could be considered an individual user factor, it is categorized here since content can suggest head motions.

Duration. VR sickness increases with the duration of the experience (Kolasinski, 1995). Content should be created for short experiences, allowing for breaks between sessions.

Vection. Vection is an illusion of self-motion when one is not actually physically moving (Section 9.3.10). Vection often, but not always, causes motion sickness.

Binocular-occlusion conflict. Stereoscopic cues that do not match occlusion clues can lead to eye strain and confusion (Section 13.2). 2D heads-up displays that result from porting from existing games often have this problem (Section 23.2.2).

Virtual rotation. Virtual rotations of one's viewpoint can result in motion sickness (Section 18.5).

Gorilla arm. Arm fatigue can result from extended use of above-the-waist gestural interfaces when there is not frequent opportunity to rest the arms (Section 14.1).

Meta Platforms, Inc.
Exhibit 1063
Page 226

Standing/walking vs. sitting. VR sickness rates are greater for standing users vs. sitting users, which correlates with postural stability (Polonen, 2010). This is in congruence with the postural stability theory (Section 12.3.3). Due to being blind and deaf to the real world, there is also more risk of injury due to falling down and collisions with real-world objects/cables when standing. Requiring users to always stand or navigate excessively via some form of walking also results in fatigue (Section 14.1).

Rest frames. Humans have a strong bias toward worlds that are stationary. If some visual cues can be presented that are consistent with the vestibular system (even if other visual cues are not), motion sickness can be reduced (Sections 12.3.4 and 18.2).

Height above the ground. Visual flow is directly related to velocity and inversely related to altitude. Low altitude in flight simulators has been found to be one of the strongest contributions to simulator sickness due to more of the visual field containing a moving scene (Johnson, 2005). Some users have reported that being placed visually at a different height than their real-world physical height causes them to feel ill.

Excessive binocular disparity. Objects that are too close to the eyes result in seeing a double image due to the eyes not being able to properly fuse the left and right eye images.

VR entrance and exit. The display should be blanked or users should shut their eyes when putting on or taking off the HMD. Visual stimuli that are not part of the VR application should not be shown to users (i.e., do not show the operating system desktop when changing applications).

Luminance. Luminance is related to flicker. Dark conditions may result in less problems for displays with little persistence (e.g., CRTs or OLEDs) and a low refresh rate (Pausch et al., 1992).

Repetitive strain injuries. Repetitive strain injury results from carrying out prolonged repeated activities (Section 14.3).

17.4 Presence vs. Motion Sickness

The rest frame hypothesis (Section 12.3.4) claims the sense of presence (Section 4.2) correlates with the stability of rest frames (Prothero and Parker, 2003). For example, a perfectly calibrated system with the entire scene acting as a rest frame that is truly stable relative to the real world (i.e., no scene motion) can indeed have high presence.

Meta Platforms, Inc.
Exhibit 1063
Page 227

However, this claim does not hold under all conditions. For example, adding non-realistic real-world stabilized cues to the environment as one flies forward in order to reduce sickness (Section 18.2) can reduce presence, especially if the rest of the environment is realistic.

Factors that enhance presence (e.g., wider field of view) can also enhance vection and in many cases vection is a VR design goal. Unfortunately, some of these same presence-enhancing variables can also increase motion sickness. Likewise, design techniques can be used to reduce the sense of self-motion and motion sickness. Such trade-offs should be considered when designing VR experiences.

Meta Platforms, Inc.
Exhibit 1063
Page 228

Examples of Reducing Adverse Effects

This chapter provides several examples of reducing sickness based off of previous chapters. The intention is not only that these examples be used, but that they serve as motivation for understanding the theoretical aspects of VR sickness so that new methods and techniques can be created.

18.1 Optimize Adaptation

Since visual and perceptual-motor systems are modifiable, sickness can be reduced through dual adaptation (Sections 10.2 and 13.4.1). Fortunately, most people are able to adapt to VR, and the more one has adapted to VR the less she will get sick (Johnson, 2005). Of course, adaptation does not occur immediately and adaptation time depends on the type of discrepancy (Welch, 1986) and the individual. Individuals who adapt quickly might not experience any sickness whereas those who are slower to adapt may become sick and give up before completely adapting (McCauley and Sharkey, 1992). Incremental exposure and progressively increasing the intensity of stimulation over multiple exposures is a very effective way to reduce motion sickness (Lackner, 2014). To maximize adaptation, sessions should be spaced 2–5 days apart (Stanney et al., 2014; Lawson, 2014; Kennedy et al., 1993). In addition to verbally and textually telling novice users to make only slow head motions, they can also be more subtly encouraged to do so through relaxing/soothing environments. Latency should also be kept constant to maximize the chance of temporal adaptation (Section 10.2.2).

18.2 Real-World Stabilized Cues

Seasickness and car sickness are triggered by the motion of the vehicle. When one looks at something that is stable within the vehicle, such as a book when reading, visual cues imply no movement, but the motion of the vehicle stimulates the vestibular

Meta Platforms, Inc.
Exhibit 1063
Page 229

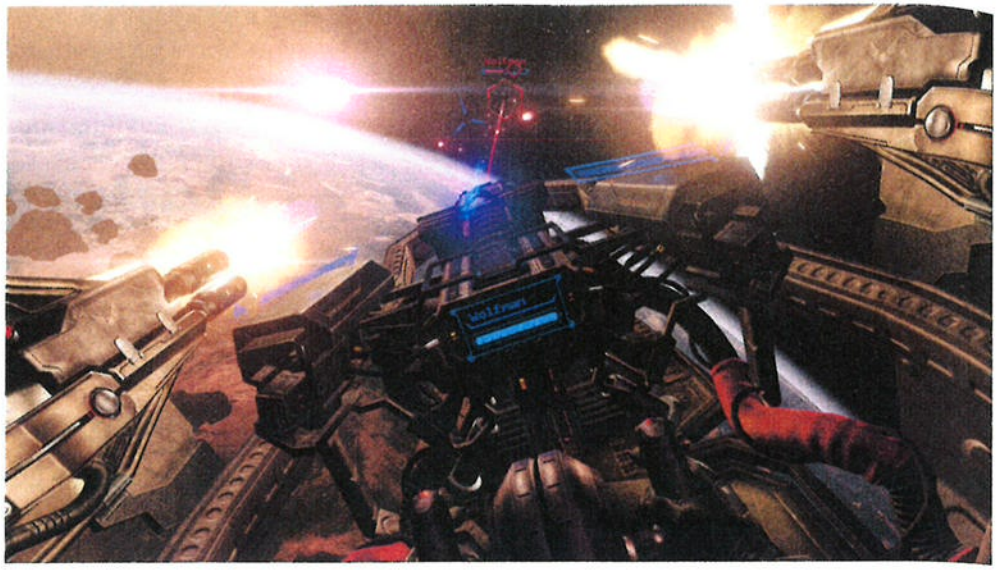


Figure 18.1 **The cockpit from EVE Valkyrie helps users to feel stabilized and reduces motion sickness.** (Image courtesy of CCP Games)

system implying there is movement. Vection can be thought of as the inverse of this—the eyes see movement, but the vestibular system is not stimulated. Although inverses of each other, visual and vestibular cues are in conflict in both cases, and as a result similar sickness often occurs.

Looking out the window in a car, or at the horizon in a boat, results in seeing visual motion cues that match vestibular cues. Since vestibular-ocular cues are no longer in conflict, motion sickness is dramatically reduced. Inversely, motion sickness can be reduced by adding stable visual cues to virtual environments (assuming low latency and the system is well calibrated) that act as a rest frame (Section 12.3.4). A cockpit that is stable relative to the real world substantially reduces motion sickness. The virtual world outside the cockpit might change as one navigates their vehicle through space, but the local vehicle interior is stable relative to the real world and the user (Figure 18.1). Visual cues that stay stable relative to the real world and the user but are not a full cockpit can also be used to help the user feel more stabilized in space and reduce motion sickness. Figure 18.2 shows an example—the blue arrows are stable relative to the real world no matter how the user virtually turns or walks through the environment.

Meta Platforms, Inc.
Exhibit 1063
Page 230



Figure 18.2 **For a well-calibrated system with low latency, the blue arrows are stationary relative to the real world and are consistent with vestibular cues even as the user virtually turns or walks around. In this case, the blue arrows also serve to indicate which way the forward direction is.** (Image courtesy of NextGen Interactions)

18.3 Manipulate the World as an Object

Although quite different than how the real world works, 3D Multi-Touch (Section 28.3.3) enables users to push, pull, turn, and stretch the viewpoint with their hands. Such viewpoint control can be thought about and perceived in two subtly different but important ways:

- *self-motion*, where the user pulls himself through the world, or
- *world motion*, where the user pushes and pulls the world as if it were an object.

While these two alternatives are technically equivalent in terms of the relative movement with respect to the world they affect, anecdotal evidence suggests the latter way of thinking can alleviate motion sickness. When the user thinks about viewpoint movement in terms of manipulating the world as an object from a stationary vantage point (i.e., the user considers gravity and the body to be the rest frame), the mind less expects the vestibular system to be stimulated. In fact, Paul Mlyniec (personal communication, April 28, 2015) claims "at least in the absence of roll and pitch, there is no more anticipation of viewpoint acceleration than if a user picked up a large object in the real world." Maintaining a mental model of the world being an object

Meta Platforms, Inc.
Exhibit 1063
Page 231

can be a challenge for some users, so visual cues that help with that mental model are important. For example, the center point of rotation and/or scaling at the midpoint between the hands can be important to help users more easily predict motion and give them a sense of direct and active control. Constraints can also be added for novice users. For example, keeping the world vertically upright can help to increase control and reduce disorientation.

18.4 Leading Indicators

When a pre-planned travel path must be used, adding cues that help users to reliably predict how the viewpoint will change in the near future can help to compensate for disturbances. This is known as a **leading indicator.** For example, adding an indicator object that leads a passive user by 500 ms along a motion trajectory of a virtual scene (similar to a driver following another car when driving) alleviates motion sickness (Lin et al., 2004).

18.5 Minimize Visual Accelerations and Rotations

Constant visual velocity (e.g., virtually moving straight ahead at a constant speed) is not a major contributor to motion sickness due to the vestibular organs not being able to detect linear velocity. However, visual accelerations (e.g., transitioning from no virtual movement to virtually moving in the forward direction) can be more of a problem and should be minimized (assuming a motion platform is not being used). If the design requires visual acceleration, the acceleration should be infrequent and occur as quickly as possible. For example, when transitioning from a stable position to some constant velocity, the period of acceleration to the constant velocity should occur quickly. Change in direction while moving, even when speed is kept constant, is a form of acceleration although it's not clear if change in direction is as sickness inducing as forward linear acceleration.

Minimizing accelerations and rotations is especially important for passive motion where the user is not actively controlling the viewpoint (e.g., immersive film where the capture point is moving). Slower orientation changes do not seem to be nearly as much of a problem as faster orientation changes. For real-world panoramic capture (Section 21.6.1), unless the camera can be carefully controlled and tested for small comfortable motions, the camera should only be moved with constant linear velocity, if moved at all. Likewise, for computer-generated scenes, the user's viewpoint should not be attached to the user's own avatar head when it is possible for the system to affect the avatar's head pose (e.g., from a physics simulation).

Meta Platforms, Inc.
Exhibit 1063
Page 232

Visual accelerations are less of a problem when the user actively controls the viewpoint (Section 12.4.6) but can still be nauseating. When using the analog sticks on a gamepad or using other steering technique (Section 28.3.2), it is best to have discrete speeds and the transition between those speeds to occur quickly.

Unfortunately, there are no published numbers as to what is acceptable for VR. VR creators should perform extensive testing across multiple users (particularly users new to VR) to determine what is comfortable with their specific implementation.

18.6 Ratcheting

Denny Unger (personal communication, March 26, 2015) has found that discrete virtual turns (e.g., instantaneous turns of 30°) he calls **ratcheting** result in far fewer reports of sickness compared to smooth virtual turns. This technique results in breaks-in-presence and can feel very strange when one is not accustomed to it, but can be worth the reduction in sickness. Interestingly, not just any angles work. Denny had to experiment and fine-tune the rotation until he reached 30°. Rotation by other amounts turned out to be more sickness inducing.

18.7 Delay Compensation

Because computation is not instantaneous, VR will always have delay. **Delay compensation techniques** can reduce the deleterious effects of system delay, effectively reducing latency (Jerald, 2009). Prediction and some post-rendering techniques are described below. These can be used together by first predicting and then correcting any error of that prediction with a post-rendering technique.

18.7.1 Prediction

Head-motion prediction is a commonly used delay compensation technique for HMD systems. Prediction produces reasonable results for low system delays or slow head movements. However, prediction increases motion overshoot and amplifies sensor noise (Azuma and Bishop, 1995). Displacement error increases with the square of the angular head frequency and the square of the prediction interval. Predicting for more than 30 ms can be more detrimental than doing no prediction for fast head motions. Furthermore, prediction is incapable of compensating for rapid transients.

18.7.2 Post-rendering Techniques

Post-rendering latency reduction techniques first render an image larger than the final display and then, late in the display process, select the appropriate subset to be presented to the user.

Meta Platforms, Inc.
Exhibit 1063
Page 233

The simplest post-rendering latency reduction technique is accomplished via a **2D warp** (what Oculus calls time warping; Carmack, 2015). As done in standard rendering, the system first renders the scene to a single image plane. The pixels are then selected from the larger image plane and presented at the corrected location just before the pixels are displayed (Jerald et al., 2007).

A cylindrical panorama, as in Quicktime VR (Chen, 1995), or spherical panorama can be used, instead of a single image plane, to remove error due to rotation. However, rendering to a panorama or sphere is computationally expensive, because standard graphics pipelines are not optimized to render to such surfaces.

A compromise between a single image plane and a spherical panorama is a **cubic environment map.** This technique extends the image plane technique by rendering the scene onto the six sides of a large cube (Greene, 1986). Head rotation simply alters what part of display memory is accessed—no other computation is required.

Regan and Pose (1994) took environment mapping further by projecting geometry onto concentric cubes surrounding the viewpoint. Larger cubes that contain projected geometry far from the viewpoint do not require re-rendering as often as smaller cubes that are close to the viewpoint. However, in order to minimize error due to large translations or close objects, a full 3D warp (Mark et al., 1997) or a pre-computed light field (Regan et al., 1999) is required, although such techniques do produce other types of visual artifacts.

18.7.3 Problems of Delay Compensation

None of the delay compensation techniques described above are perfect. For example, 2D warping a single image plane minimizes error at the center of the display, but error increases toward the periphery in the direction of the head turn (Jerald et al., 2007). Ultra-wide-field-of-view HMDs may especially be a problem due to this peripheral error that results in incorrect motion as the head turns and the mind's sensitivity to motion in the periphery. Cubic environment mapping helps by reducing error in the periphery, but such techniques only compensate for rotations, and natural head rotations include some motion parallax. Thus, visual artifacts are worse the closer geometry is to the viewpoint. Furthermore, many latency compensation techniques assume the scene is static; object motion is often not corrected for. Regardless, most latency compensation techniques, if implemented well, can reduce error with visual artifacts being less detrimental than with no correction.

18.8 Motion Platforms

Motion platforms (Section 3.2.4), if implemented well, can help to reduce simulator sickness by reducing the inconsistency between vestibular and visual cues. How-

Meta Platforms, Inc.
Exhibit 1063
Page 234

ever, motion platforms do add a risk of increasing motion sickness by (1) causing incongruency between the physical motion and visual motion (the primary cause), and (2) increasing physically induced motion sickness independent of how visuals move (a lesser cause). Thus motion platforms are not a quick fix for reducing motion sickness—they must be carefully integrated as part of the design.

Even passive motion platforms (Section 3.2.4) can reduce sickness (Max Rheiner, personal communication, May 3, 2015). Birdly by Somniacs (Figure 3.15) is a flying simulation where the user becomes a bird flying above San Francisco. The user controls the tilt of the platform by leaning. The active control of flying with the arms also helps to reduce motion sickness. Motion sickness is further reduced because the user is lying down, so less postural instability occurs than if the user were standing.

18.9 Reducing Gorilla Arm

Using VR devices can become tiresome after extended use due to holding the controllers out and in front of the user (Section 14.1). However, when the interaction is designed so users can mostly interact with their hands in a relaxed manner at their sides and/or in their laps, then they can interact for several hours at a time without experiencing gorilla arm (Jerald et al., 2013).

18.10 Warning Grids and Fade-Outs

For a user approaching the edge of tracking, a real-world physical object not visually represented in the virtual world (e.g., a wall), or the edge of a safe zone, a **warning grid** (Figure 18.3) or the real world can fade in to provide a cue to the user that he should back off. Whereas such feedback can cause a break-in-presence, it is better than the alternative of the loss of quality tracking that results in a worse break-in-presence and injury.

Once tracking is no longer solid or latency exceeds some value, the system should quickly fade out the display to a single color to prevent motion sickness. Such problems are most easily detected by detecting tracking jitter or a decreased frame rate. Alternatively, for systems with low-latency video-see-through capability, the system can quickly fade in the real world.

18.11 Medication

Anti-motion-sickness medication can potentially enhance the rate of adaptation by allowing progressive exposure to higher levels of stimulation without symptoms being elicited. Unfortunately, many of the various medications that have been developed to minimize motion sickness lead to serious side effects that dramatically restrict their field of application (Keshavarz et al., 2014).

Meta Platforms, Inc.
Exhibit 1063
Page 235

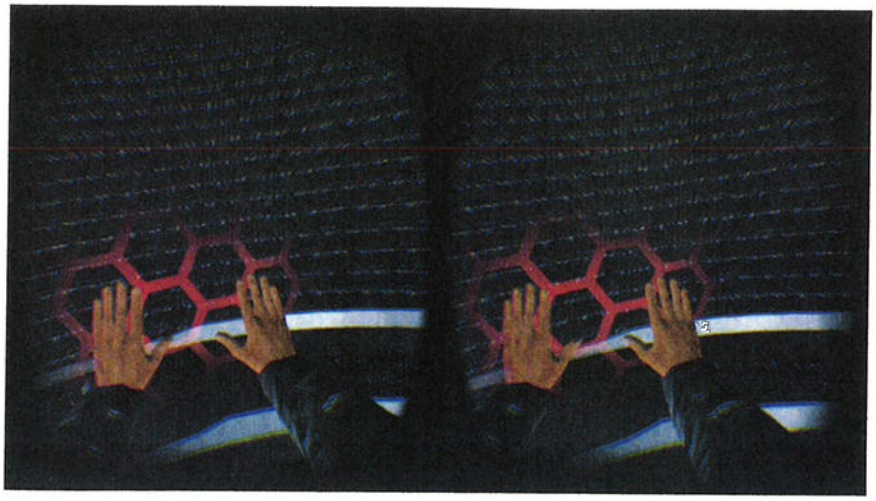


Figure 18.3 **A warning grid can be used to inform the user he is nearing the edge of tracked space.** (Image courtesy of Sixense)

Ginger has also been touted as a remedy, but its effects are marginal (Lackner, 2014). Relaxation feedback training has been used in conjunction with incremental exposure to increasingly provocative stimulation as a way of decreasing susceptibility to VR sickness.

In drug studies of motion sickness, there are usually large benefits of placebo effects, typically in the 10%–40% range of the drug effects. Wrist acupressure bands and magnets sold to "alleviate" or prevent motion sickness could potentially provide placebo benefits for some people.

Meta Platforms, Inc.
Exhibit 1063
Page 236

Adverse Health Effects: Design Guidelines

The potential for adverse health effects of VR is the greatest risk for individuals as well as for VR achieving mainstream acceptance. It is up to each of us to take any adverse health effects seriously and do everything possible to minimize problems. The future of VR depends on it.

Practitioner guidelines that summarize chapters from Part III are organized below as hardware, system calibration, latency reduction, general design, motion design, interaction design, usage, and measuring sickness.

Hardware

- Choose HMDs that have no perceptible flicker (Section 13.3).
- Choose HMDs that are light, have the weight centered just above the center of the head, and are comfortable (Section 14.1).
- Choose HMDs without internal video buffers, with a fast pixel response time, and with low persistence (Section 15.4.4).
- Choose trackers with high update rates (Section 15.4.1).
- Choose HMDs with tracking that is accurate, precise, and does not drift (Section 17.1).
- Use HMD position tracking if the hardware supports it (Section 17.1). If position tracking is not available, estimate the sensor-to-eye offset so motion parallax can be estimated as users rotate their heads.
- If the HMD does not convey binocular cues well, then use biocular cues, as they can be more comfortable even though not technically correct (Section 17.1).
- Use wireless systems if possible. If not, consider hanging the wires from the ceiling to reduce tripping and entanglement (Section 14.3).

Meta Platforms, Inc.
Exhibit 1063
Page 237

- Choose haptic devices that are not able to exceed some maximum force and/or that have safety mechanisms (Section 14.3).
- Add code to prevent audio gain from exceeding a maximum value (Section 14.3).
- For multiple users, use larger earphones instead of earbuds (Section 14.4).
- Use motion platforms whenever possible in a way that vestibular cues correspond to visual motion (Section 18.8).
- Choose hand controllers that do not have line-of-sight requirements so that users can work with their hands comfortably to their sides or in their laps with only occasional need to reach up high above the waist (Section 18.9).

System Calibration

- Calibrate the system and confirm often that calibration is precise and accurate (Section 17.1) in order to reduce unintended scene motion (Section 12.1).
- Always have the virtual field of view match actual field of view of the HMD (Section 17.1).
- Measure for interpupillary distance and use that information for calibrating the system (Section 17.2).
- Implement options for different users to configure their settings differently. Different users are prone to different sources of adverse effects (e.g., new users might use a smaller field of view and experienced users may not mind more complex motions) (Section 17.2).

Latency Reduction

- Minimize overall end-to-end delay as much as possible (Section 15.4).
- Study the various types of delay in order to optimize/reduce latency (Section 15.4) and measure the different components that contribute to latency to know where optimization is most needed (Section 15.5).
- Do not depend on filtering algorithms to smooth out noisy tracking data (i.e., choose trackers that are precise so there is no tracker jitter). Smoothing tracker data comes at the cost of latency (Section 15.4.1).
- Use displays with fast response time and low persistence to minimize motion blur and judder (Section 15.4.4).

Meta Platforms, Inc.
Exhibit 1063
Page 238

- Consider not waiting on vertical sync (note that this is not appropriate for all HMDs) as reducing latency can be more important than removing tearing artifacts (Section 15.4.4). If the rendering card and scene content allow it, render at an ultra-high rate that approaches just-in-time pixels to reduce tearing.
- Reduce pipelining as much as possible as pipelining can add significant latency (Sections 15.4 and 15.5). Likewise, beware of multi-pass rendering as some multi-pass implementations can add delay.
- Be careful of occasional dropped frames and increases in latency. Inconsistent latency can be as bad as or worse than long latency and can make adaptation difficult (Section 18.1).
- Use latency compensation techniques to reduce effective/perceived latency, but do not rely on such techniques to compensate for large latencies (Section 18.7).
- Use prediction to compensate for latencies up to ~30 ms (Section 18.7.1).
- After predicting, correct for prediction error with a post-rendering technique (e.g., 2D image warping; Section 18.7.2).

General Design

- Minimize visual stimuli close to the eyes as this can cause vergence/accommodation conflict (Section 13.1).
- For binocular displays, do not use 2D overlays/heads-up displays. Position overlays/text in 3D at some distance from the user (Section 13.2).
- Consider making scenes dark to reduce the perception of flicker (Section 13.3).
- Avoid flashing lights 1 Hz or above anywhere in the scene (Section 13.3).
- To reduce risk of injury, design experiences for sitting. If standing or walking is used, provide protective physical barriers (Section 14.3).
- For long walking or standing experiences, also provide a physical place to sit, with visual representation in the virtual world so users can find it (Section 14.1).
- Design for short experiences (Section 17.3).
- Fade in a warning grid when the user approaches the edge of tracking, a real-world physical object not visually represented in the virtual world, or the edge of a safe zone (Section 18.10).
- Fade out the visual scene if latency increases or head-tracking quality decreases (Section 18.10).

Meta Platforms, Inc.
Exhibit 1063
Page 239

Motion Design

- If the highest-priority goal is to minimize VR sickness (e.g., when the intended audience is non-experienced VR users), then do not move the viewpoint in any way whatsoever that deviates from actual head motion of the user (Chapter 12).
- If latency is high, do not design tasks that require fast head movements (Section 18.1).

Both Active and Passive Motion

- Be careful of adding visual motion that results in users trying to incorrectly physically balance themselves in order to compensate for the visual motion (Section 12.3.3).
- Design for the user to be seated or lying down to reduce postural instability (Section 12.3.3).
- Focus on making rest frame cues consistent with vestibular cues rather than being overly concerned with making all orientation and all motion consistent (Section 12.3.4).
- Consider dividing visuals into two components, one component that represents content and one component that represents the rest frame (Section 12.3.4).
- When a large portion of the background is visible and realism is not required, consider making the entire background a real-world stabilized rest frame that matches vestibular cues (Section 12.3.4).
- Use a stable cockpit for vehicle experiences or consider using non-realistic world-stabilized cues for non-vehicle experiences (Section 18.2).
- Perform extensive testing across multiple users (particularly users new to VR) to determine what motions are comfortable with the specific implementation or experience (Chapter 16).
- Do not put the viewpoint near the ground when the viewpoint moves in any way other than from direct head motion (Section 17.3).

Passive Motion Design

- Only passively change the viewpoint when absolutely required (Section 12.4.6).
- If passive motion is required, then minimize any motion other than linear velocity (Sections 12.2 and 18.5).
- Never add visual head bobbing (Section 12.2).

Meta Platforms, Inc.
Exhibit 1063
Page 240

- Do not attach the viewpoint to the user's own avatar head if the system can affect the pose of the avatar's head (Section 18.5).
- For real-world panoramic capture, unless the camera can be carefully controlled and tested for small comfortable motions, the camera should only be moved with constant linear velocity, if moved at all (Section 18.5).
- Slower orientation changes are not as much of a problem as faster orientation changes (Section 18.5).
- If virtual acceleration is required, ramp up to a constant-speed velocity quickly (e.g., instantaneous acceleration is better than gradual change in velocity) and minimize the occurrence of these accelerations (Section 18.5).
- Use leading indicators so the user can know what motion is about to occur (Section 18.4).

Active Motion Design

- Visual accelerations are less of a problem when the user actively controls the viewpoint (Section 12.4.6) but can still be nauseating.
- When using analog sticks on a gamepad to navigate or when using other steering technique, it is best to have discrete speeds and the transition between those speeds to occur quickly (Section 18.5).
- Design for physical rotation instead of virtual rotation whenever possible (Section 12.2). Consider using a swivel chair (although this can be troublesome for wired systems).
- Considering offering a non-realistic ratchet mode for angular rotations (Section 18.6).
- Be careful of the viewpoint moving up and/or down due to hilly terrain, stairs, etc. (Section 12.2).
- For 3D Multi-Touch, add cues that imply the world is an object to be manipulated instead of moving oneself through the environment (Section 18.3).

Interaction Design

- Design interfaces so users can work with their hands comfortably at their sides or in their laps with only occasional need to reach up high above the waist (Section 18.9).
- Design interactions to be non-repetitive to reduce repetitive strain injuries (Section 14.3).

Meta Platforms, Inc.
Exhibit 1063
Page 241

Usage

Safety

- Consider forcing users to stay within safe areas via physical constraints such as padded circular railing (Section 14.3).
- For standing or walking users, have a human spotter carefully watch the users and help stabilize them when necessary (Section 14.3).

Hygiene

- Consider providing to users a clean cap to wear underneath the HMD and removable/cleanable padding that sits between the face and the HMD (Section 14.4).
- Wipe down equipment and clean the lenses after use and between users (Section 14.4).
- Keep sick bags, plastic gloves, mouthwash, drinks, light snacks, and cleaning products nearby, but keep these items hidden so they don't suggest to users that they might get sick (Section 14.4).

New Users

- For new users, be especially conservative with presenting any cues that can induce sickness.
- Consider decreasing the field of view for new users (Section 17.2).
- Unless using gaze-directed steering (Section 28.3.2), tell gamers something like "Forget your video games. If you think you're going to move where your head is pointing, then think again. This isn't Call of Duty." (Section 17.3).
- Encourage users to start with only slow head motions (Section 18.1).
- Begin with short sessions and/or use timeouts/breaks excessively. Do not allow prolonged sessions until the user has adapted (Sections 17.2 and 17.3).

Sickness

- Do not allow anyone to use a VR system that is not in their usual state of health (Section 17.2).
- Inform users in a casual way that some users experience discomfort and warn them to stop at the first sign of discomfort, rather than "tough it out." Do not overemphasize adverse effects (Sections 16.1 and 17.2).
- Respect a user's choice to not participate or to stop the experience at any time (Section 17.2).

Meta Platforms, Inc.
Exhibit 1063
Page 242

- Do not present visuals while the user is putting on or taking off the HMD (Section 17.2).
- Do not freeze, stop the application, or switch to the desktop without first having the screen go blank or having the user close her eyes (Section 17.3).
- Carefully pay attention for early warning signs of VR sickness, such as pallor and sweating (Chapter 12).

Adaptation and Readaptation

- For faster readaptation to the real world after VR usage, use active readaptation by providing activities targeted at recalibrating the user's sensory systems (Section 13.4.1).
- For reducing readaptation discomfort after VR usage, use natural decay by having the user gradually readapt to the real world—encourage the user to sit and relax with little movement and eyes closed (Section 13.4.1).
- After VR usage, inform users not to drive, operate heavy machinery, or participate in high-risk behavior for at least 30–45 min and after all aftereffects have passed (Section 13.4).
- To maximize adaptation, use VR at spaced intervals of 2–5 days (Section 18.1).

Measuring Sickness

- For easiest data collection, use symptom checklists or questionnares. The Kennedy Simulator Sickness Questionnaire is the standard (Section 16.1).
- Postural stability tests are also quite easy to use if the person administrating the test is trained (Section 16.2).
- For objectively measuring sickness, consider using physiological measures (Section 16.3).

Meta Platforms, Inc.
Exhibit 1063
Page 243

PART

CONTENT CREATION

We graphicists choreograph colored dots on a glass bottle so as to fool eye and mind into seeing desktops, spacecraft, molecules, and worlds that are not and never can be.

—Frederick P. Brooks, Jr. (1988)

Humans have been creating content for thousands of years. Cavemen painted inside caves. The ancient Egyptians created everything from pottery to pyramids. Such creations vary widely, from the magnificent to the simple, across all cultures and ages, but beauty has always been a goal. Philosophers and artists argue that delight is equal to usefulness, and this is even truer for VR.

VR is a relatively new medium and is not yet well understand. Most other disciplines have been around much longer than VR and although many aspects of other crafts are very different from VR, there are still many elements of those disciplines that we can learn from. For content creation, studying fields such as architecture, urban planning, cinema, music, art, set design, games, literature, and the sciences can add great benefit to VR creations. The chapters within Part IV utilize some concepts from these disciplines that are especially pertinent to VR. Different pieces from different disciplines all must come together in VR to form the experience—the story, the action and reaction, the characters and social community, the music and art. Integrating these different pieces together, along with some creativity, can result in experiences that are more than the sum of the individual pieces.

Part IV consists of five chapters that focus on creating the assets of virtual worlds.

Meta Platforms, Inc.
Exhibit 1063
Page 244

Chapter 20, High-Level Concepts of Content Creation, discusses core concepts of content creation common to the most engaging VR experiences: the story, the core experience, conceptual integrity, and gestalt principles.

Chapter 21, Environmental Design, discusses design of the environment ranging from different types of geometry to non-visual cues. Designing the environment is more than about just creating pleasing stimuli—the design of the environment provides clues to affordances, constrains interaction, and enhances wayfinding.

Chapter 22, Affecting Behavior, discusses how content creation can affect the states and actions of users. Influencing behavior is especially important for VR because the experience cannot be directly controlled by the system like many other mediums often do. The chapter includes sections on personal wayfinding aids, directing attention, how hardware choice relates to content creation and behavior, and social aspects of VR.

Chapter 23, Transitioning to VR Content Creation, discusses how creating for VR is different than creating for other mediums. Ideally, the VR content should be designed from the beginning for VR. However, when that is not possible, several tips are provided to help port existing content to VR.

Chapter 24, Content Creation: Design Guidelines, summarizes the previous four chapters and lists a number of actionable guidelines for those creating VR content.

Meta Platforms, Inc.
Exhibit 1063
Page 245

High-Level Concepts of Content Creation

Virtual worlds would be very lonely places without content. This chapter discusses general high-level tips on creating content—creating a compelling story, focusing on the core experiences, keeping the integrity of the world consistent, and gestalt principles of perceptual organization.

20.1 Experiencing the Story

We need to consider a user's experiential arc through time. Virtual worlds can be intense or contemplative, ideally experiences oscillate between these, thus giving the author the ability to punctuate story elements with different design elements of VR.

—Mark Bolas (personal communication, June 13, 2015)

In traditional media ranging from plays to books to film to games, stories are compelling because they resonate with the audience's experiences. VR is extremely experiential and users can become part of a story more so than with any other form of media. However, simply telling a story using VR does not automatically qualify it to be an engaging experience—it must be done well and in a way that resonates with the user.

To convey a story, all of the details do not need to be presented. Users will fill in the gaps with their imaginations. People give meaning to even the most basic shapes and motions. For example, Heider and Simmel (1944) created a two-and-a-half-minute video of some lines, a small circle, a small triangle, and a large triangle (Figure 20.1). The circle and triangles moved around both inside the lines and outside the lines, and sometimes touched each other. Participants were simply told "write down what happened in the picture." One participant gave the following story.

Meta Platforms, Inc.
Exhibit 1063
Page 246

> A man has planned to meet a girl and the girl comes along with another man. The first man tells the second to go; the second tells the first, and he shakes his head. Then the two men have a fight, and the girl starts to go into the room to get out of the way and hesitates and finally goes in. She apparently does not want to be with the first man. The first man follows her into the room after having left the second in a rather weakened condition leaning on the wall outside the room. The girl gets worried and races from one corner to the other in the far part of the room . . .

All 34 participants created stories with 33 participants interpreting the actions of the geometric shapes as animate beings. Only one participant created a story consisting of geometric shapes. What is most surprising, however, is how similar some parts of their stories were. The 33 subjects who had stories about animate beings all stated similar things for specific points in the animation: two characters had a fight, one character was shut in some structure and tries to get out, and one character chased another.

VR creators will never be able to take the subjectivity out of a story experience. Users will always distort any stimuli presented to them based on their filters of the values, beliefs, memories, etc. (Section 7.9.3). However, the above experiment suggests that in some cases, even simple cues can provide fairly consistent interpretations. Although we can't control exactly how users will make up all the details of the story in their own mind, content creators can lead the mind through suggestion. Content creators

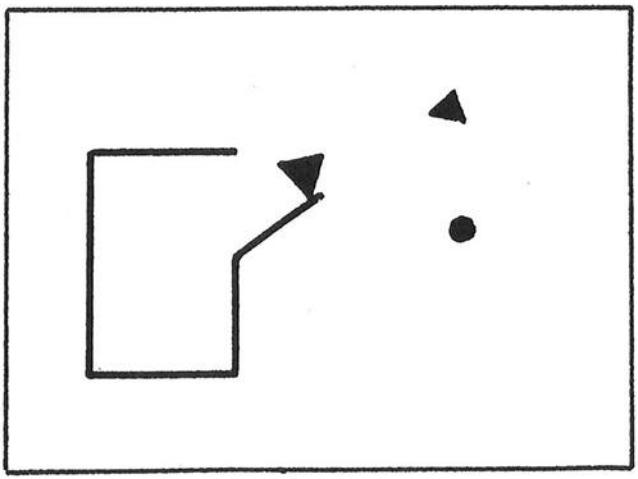


Figure 20.1 A frame from a simple geometric animation. Subjects came up with quite elaborate stories based on the animation. (From Heider and Simmel (1944))

Meta Platforms, Inc.
Exhibit 1063
Page 247

should focus on the most important aspects of the story and attempt to make them consistently interpreted across users.

Experiential fidelity is the degree to which the user's personal experience matches the intended experience of the VR creator (Lindeman and Beckhaus, 2009). Experiential fidelity can be increased by better technology, but is also increased through everything that goes with the technology. For example, priming users before they enter the virtual world can structure anticipation and expectation in a way that will increase the impact of what is presented. Prior to the exposure, give the user a backstory about what is about to occur. The mind provides far better fidelity than any technology can, and the mind must be tapped to provide quality experiences. Providing perfect realism is not necessary if enough clues are provided for them to fill in the details with their imagination. Note experiential fidelity for all parts of the experience is not always the goal. A social free-roaming world can allow users to create their own stories—perhaps the content creator simply provides the meta-story that users create their own personal stories within.

Skeuomorphism is the incorporation of old, familiar ideas into new technologies, even though the ideas no longer play a functional role (Norman, 2013). This can be important for a wide majority of people who may be more resistant to change than early adopters. If designed to match the real world, VR experiences can be more acceptable than other technologies because of the lower learning curve—one interacts in a similar way to the real world rather than having to learn a new interface and new expectations. Then real-world metaphors can be incorporated into the story, gradually teaching how to interact and move through the world in new ways. Eventually, even conservative users will be open to more magical interactions (Section 26.1) that have less relationship to the old, through skeuomorphic designs that help with the gradual transition.

In June of 2008, approximately 50 leading VR researchers at the "Designing the Experience" session at the Dagstuhl Seminar on Virtual Realities identified four primary attributes that characterize great experiences (Lindeman and Beckhaus, 2009) as described below. To achieve the best experiences, shoot for the following concepts through powerful storytelling and sensory input.

Strong emotion can cover a range of emotions and feeling of extremes. These include any strong feeling such as joy, excitement, and surprise that occurs without conscious effort. Strong emotions result from anticipating an event, achieving a goal, or fulfilling a childhood dream. Most people are more aroused by and interested in emotional stories than logical stories. Stories aimed at emotions distract from the limits of technology.

Meta Platforms, Inc.
Exhibit 1063
Page 248

Deep engagement occurs when users are in the moment and experience a state of flow as described by Csikszentmihalyi (2008). Engaged users are extremely focused on the experience, losing track of space and time. Users become hypersensitive, with a heightened awareness without distraction.

Massive stimulation occurs when there is a sensory input across all the modalities, with each sense receiving a large amount of stimulation. The entire body is immersed and feels the experience in multiple ways. The stimulation should be congruent with the story.

Escape from reality is being psychologically removed from the real world. There is a lack of demands of the real world and the people in it. Users fully experience the story when not paying attention to sensory cues from the real world or the passage of time.

VR does not exist in isolation. Disney understands this extremely well. Over a period of 14 months in the mid-1990s, approximately 45,000 users experienced the Aladdin attraction, which used VR as a medium to tell stories (Pausch et al., 1996). The authors came to the following conclusions.

- Those not familiar with VR are unimpressed with technology. Focus on the entire experience instead of the technology.
- Provide a relatable background story before immersing users into the virtual world.
- Make the story simple and clear.
- Focus on content, why users are there, and what there is to do.
- Provide a specific goal to perform.
- Focus on believability instead of photorealism.
- Keep the experience tame enough for the most sensitive users.
- Minimize breaks-in-presence such as interpenetrating objects and characters not responding to the user.
- Like film, VR appeals to everyone but content may segment the market, so focus on the targeted audience.

20.2 The Core Experience

The **core experience** is the essential moment-to-moment activity of users making meaningful choices resulting in meaningful feedback. Even though the overall VR

Meta Platforms, Inc.
Exhibit 1063
Page 249

experience might contain an array of various contexts and constraints, the core experience should remain the same. Because the core experience is so essential, it should be the primary focus of VR creators to make that core experience pleasurable so users stay engaged and want to come back for more.

A hypothetical example is a VR ping-pong game. A traditional ping-pong game very much has a core experience of continuously competing via simple actions that can be improved upon over time. A VR game might go well beyond a traditional real-world ping-pong game via adding obstacles, providing extra points for bouncing the ball differently, increasing the difficulty, presenting different background art, rewarding the player with a titanium racquet, etc. However, the core experience of hitting the ping-pong ball still remains. If that core experience is not implemented well so that it is enjoyable, then no number of fancy features will keep users engaged for more than a couple of minutes.

Even for a serious non-game application, making the core experience pleasurable is just as crucial for users to think of it as more than just an interesting demo, and to continue using the system. Gamification is, at least in part, about taking what some people consider mundane tasks and making the core of that task enjoyable, challenging, and rewarding.

Determining what the core experience is and making it enjoyable is where creativity is required. However, that creativity rarely results in the core experience being enjoyable on a first implementation. Instead VR creators must continuously improve upon the core experience, build various prototype experiences focused on that core, and learn from observing and collecting data from real users. These concepts are covered in Part VI.

20.3 Conceptual Integrity

No style will appeal to everyone; a mishmash of styles delights no one . . . Somehow consistency brings clarity, and clarity brings delight.

—Frederick P. Brooks, Jr. (2010)

Brooks argues in the *Mythical Man-Month* (Brooks, 1995) "conceptual integrity is the most important consideration in system design." This is even more so with VR content creation. **Conceptual integrity** is also known as coherence, consistency, and sometimes uniformity of style (Brooks, 2010).

The basic structure of the virtual world should be directly evident and self-explanatory so users can immediately understand and start experiencing and using the world.

Meta Platforms, Inc.
Exhibit 1063
Page 250

There might be complexity added as users become experts, but the core conceptual model should be the same.

Extraneous content and features that are immaterial to the experience, even if good, should be left out. It is better to have one good idea that carries through the entire experience than to have several loosely related good ideas. If there are many but incompatible ideas where the sum of those ideas is better than the existing primary concept, then the primary concept should be reconsidered to provide an overarching theme that encompasses all of the best ideas in a coherent manner.

The best artists and designers create conceptual integrity with consistent conscious and subconscious decisions. How can conceptual integrity be enforced when the project is complex enough to require a team? Empower a single individual director who controls the basic concepts and has a clear passionate vision to direct the project's vision and content.

The **director** primarily acts as an agent to serve the users of the experience, but also serves to fulfill the project goals and to direct the team in focusing on the right things. The director takes input and suggestions from the team but is the final decision maker on the concepts and what the experience will be. She leads many steps in defining the project (Chapter 31). However, the director is not a dictator of how implementation will occur (Chapter 32) or how feedback will be collected from users (Chapter 33), but instead suggests to and takes suggestions from those working in those stages. The director should always be prepared to show an example implementation, even if not an ideal implementation, of how his ideas might be implemented. Even if the makers and data collectors accept the director's suggestions, then the director should give credit to them.

Makers and data collectors should have immediate access to the director to make sure the right things are being implemented and measured. Questions and communication should be highly encouraged. This communication should be maintained throughout the project as the project definition will change and assumptions may drift apart as the project evolves.

20.4 Gestalt Perceptual Organization

Gestalt is a German word that roughly translates into English as *configuration*. The central principle of **gestalt psychology** states that perception depends on a number of organizing principles, which determine how we perceive objects and the world. This principle maintains that the human mind considers objects in their entirety before, or in parallel with, perception of their individual parts; suggesting the whole is different than the sum of its parts. Gestalt psychology is even more important for VR

Meta Platforms, Inc.
Exhibit 1063
Page 251

as compared to pictures alone as the best VR applications bring together all the senses into a cohesive experience that could not be obtained by one sensory modality alone.

The gestalt effect is the capability of our brain to generate whole forms, particularly with respect to the visual recognition of global figures, instead of just collections of simpler and unrelated elements such as points, lines, and polygons. As demonstrated in Section 6.2 and with every VR experience we have ever tried, we are pattern-recognizing machines and consistently see things that do not exist.

Perceptual organization involves two processes: grouping and segregation. Grouping is the process by which stimuli are put together into units or objects. Segregation is the process of separating one area or object from another.

20.4.1 Principles of Grouping

Gestalt groupings are rules of how we naturally perceive objects as organized patterns and objects, enabling us to form some degree of order from the chaos of individual components.

The principle of **simplicity** states that figures tend to be perceived in their simplest form instead of complicated shapes. As can be seen in Figure 20.2, perception of an object being 3D varies depending on the simplicity of its 2D projection; the simpler the 2D interpretation, the more likely that figure will be perceived as 2D. At a higher level, we tend to recognize 3D objects even if they are not realistic. The hydrant in Figure 20.3 is intentionally bent in an unrealistic manner to provide character, but we still perceive it as a simple recognizable object instead of abstract individual components.

The principle of **continuity** states aligned elements tend to be perceived as a single group or chunk, and are interpreted as being more related than unaligned elements. Points that form patterns are seen as belonging together, and the lines tend to be seen in such a way as to follow the smoothest path. An "X" symbol or two crossing lines (Figure 20.4) are perceived as two lines that cross instead of angled parts. Our

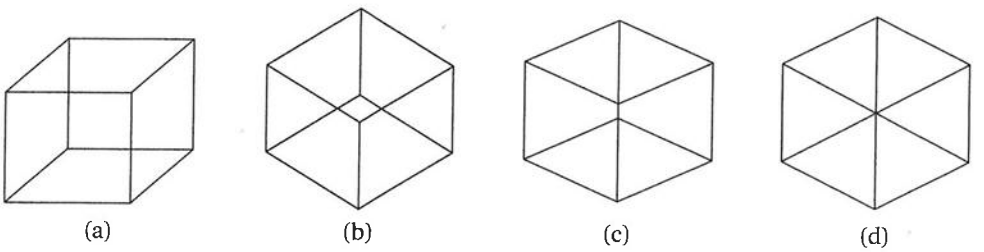


Figure 20.2 **The principle of simplicity states that figures tend to be perceived in their simplest form, whether that is 3D (a) or 2D (d).** (From Lehar (2007))

Meta Platforms, Inc.
Exhibit 1063
Page 252

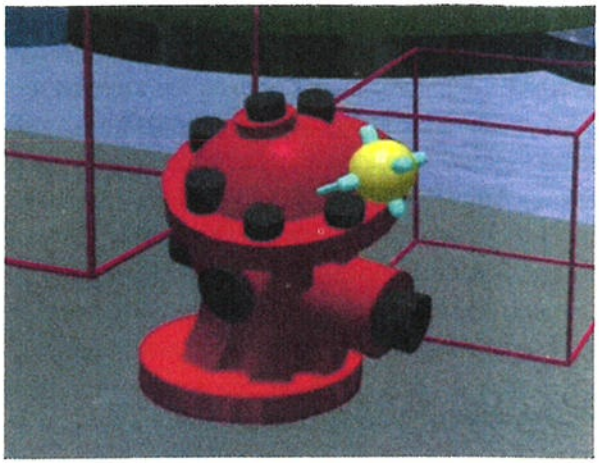


Figure 20.3 **A MakeVR user applies the principle of simplicity when creating (the yellow/blue geometry is the user's 3D cursor) a fire hydrant.** (Image courtesy of Sixense)

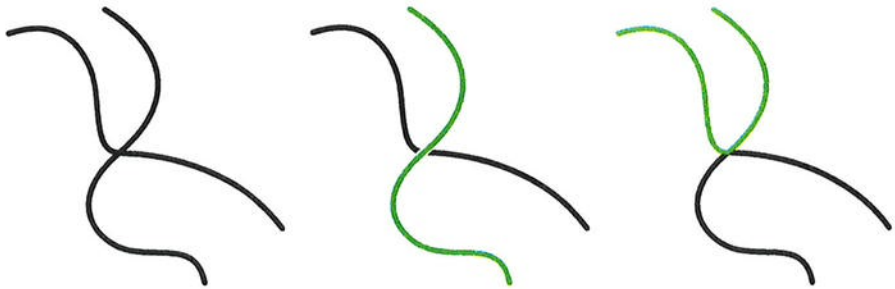


Figure 20.4 **The principle of continuity states that elements tend to be perceived as following the most aligned or smoothest path.** (Based on http://braveux.com/the-origins-of-ux-part-2-of-2; accessed November 15, 2014)

brains automatically and subconsciously integrate cues as whole objects that are most likely to be in reality. Objects that are partially covered by other objects are also seen as continuing behind the covering object, as shown in Figure 20.5.

The principle of **proximity** states that elements that are close to one another tend to be perceived as a shape or group (Figure 20.6). Even if the shapes, sizes, and objects are radically different from each other, they will appear as a group if they are close

Meta Platforms, Inc.
Exhibit 1063
Page 253

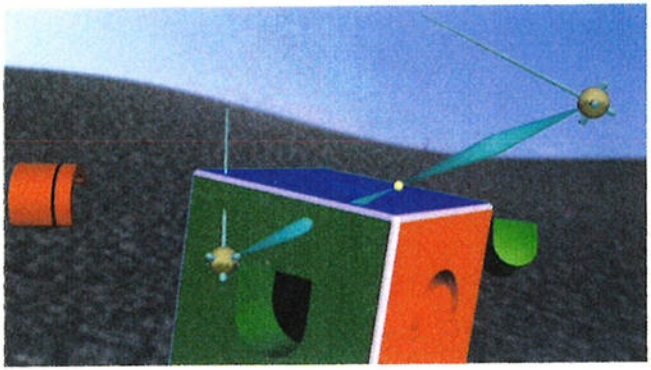


Figure 20.5 **The blue geometry connecting the two yellow 3D cursors are perceived as being connected even when occluded by the cubic object. The "poker" part of the left cursor is also perceived as continuing through the cube (Yoganandan et al., 2014).**

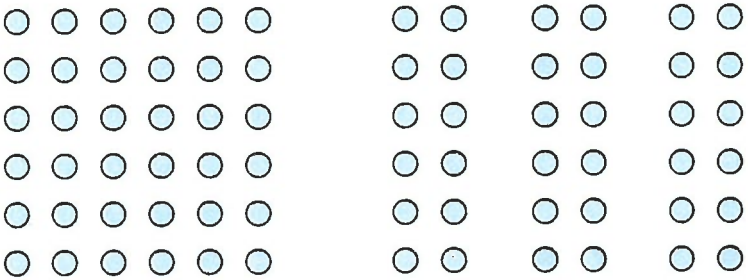


Figure 20.6 **The principle of proximity states that elements close together tend to be perceived as a shape or group.** (Based on http://en.wikipedia.org/wiki/Principles_of_grouping; accessed on November 15, 2014)

together. Society evolved by using proximity to create written language. Words are groups of letters, which are perceived as words because they are spaced close together in a deliberate way. While learning to read, one recognizes words not only by individual letters but by the grouping of those letters into larger recognizable patterns. Proximity is also important when creating graphical user interfaces, as shown in Figure 20.7.

The principle of **similarity** states that elements with similar properties tend to be perceived as being grouped together (Figure 20.8). Similarity depends on form, color, size, and brightness of the elements. Similarity helps us organize objects into groups in realistic scenes or abstract art scenes, as shown in Figure 20.9.

Meta Platforms, Inc.
Exhibit 1063
Page 254



Figure 20.7 **The panel in MakeVR uses the principles of proximity and similiarity to group different user interface elements together (drop-down menu items, tool icons, boolean operations, and apply-to options).** (Image courtesy of Sixense)

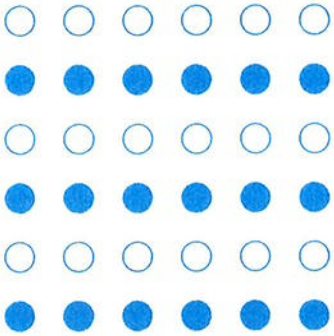


Figure 20.8 **The principle of similarity states that elements with similar properties tend to be perceived as being grouped together.** (Based on http://en.wikipedia.org/wiki/Principles_of_grouping; accessed on May 12, 2015)

The principle of **closure** states that when a shape is not closed, the shape tends to be perceived as being a whole recognizable figure or object (Figure 20.10). Our brains do everything to perceive incomplete objects and shapes as whole recognizable objects, even if that means filling in empty space with imaginary lines such as in the Kanizsa illusion (Section 6.2.2). When the viewer's perception completes a shape, closure occurs.

Temporal closure is used with film, animation, and games. When a character begins a journey and then suddenly in the next scene the character arrives at his

Meta Platforms, Inc.
Exhibit 1063
Page 255

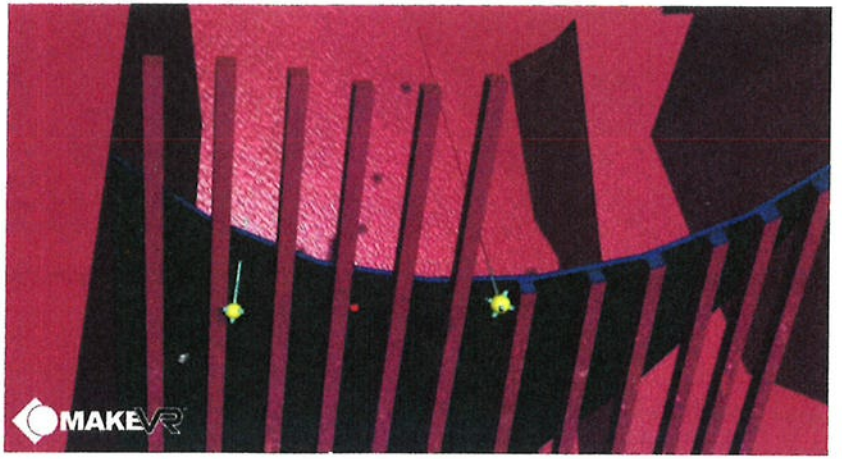


Figure 20.9 **A MakeVR user using the principle of similarity to build abstract art.** (Image courtesy of Sixense)

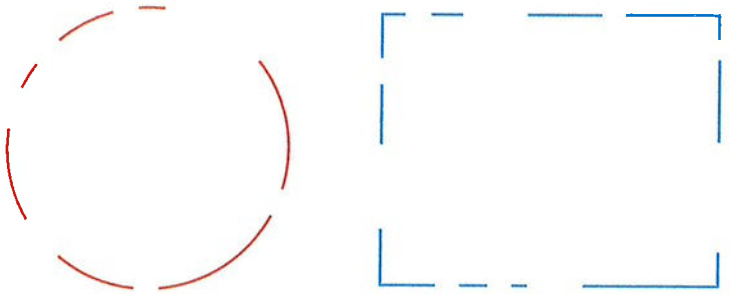


Figure 20.10 **The principle of closure states that when a shape is not closed, the shape tends to be perceived as being a whole recognizable figure or object.** (Based on http://en.wikipedia.org/wiki/Principles_of_grouping; accessed on May 12, 2015)

destination, then the mind fills the time between those scenes—we don't need to watch the entire journey. A similar concept is used in immersive games and film. At shorter timescales, apparent motion can be perceived when in fact no motion occurred or the screen is blank between displayed frames (Section 9.3.6).

The principle of **common fate** states that elements moving together tend to be perceived as grouped, even if they are in a larger, more complex group (Figure 20.11). Common fate applies not only to groups of moving elements but also to the form of

Meta Platforms, Inc.
Exhibit 1063
Page 256

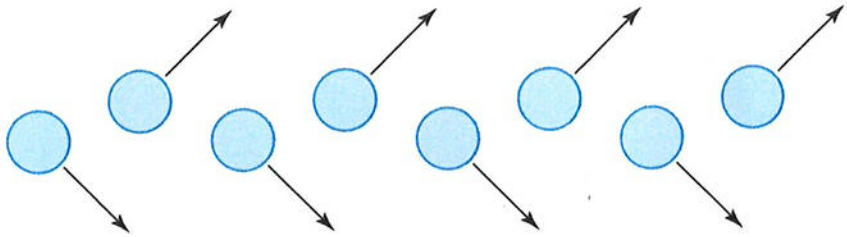


Figure 20.11 **The principle of common fate states that elements moving in a common way (signified by arrows) tend to be perceived as being grouped together.**

objects. For example, common fate helps to perceive depth of objects when moving the head through motion parallax (Section 9.1.3).

20.4.2 Segregation

The perceptual separation of one object from another is known as **segregation.** The question of what what is the foreground and what is the background is often referred to as the figure-ground problem.

Objects tend to stand out from their background. In perceptual segregation terms, a **figure** is a group of contours that has some kind of object-like properties in our consciousness. The **ground** that is the background is perceived to extend behind a figure. Borders between figure and ground appear to be part of the figure (known as ownership).

Figures have distinct form whereas ground has less form and contrast. Figures appear brighter or darker than equivalent patches of light that form part of the ground. Figures are seen as richer, are more meaningful, and are remembered more easily. A figure has the characteristics of a "thing," more apt to suggest meaning, whereas the ground has less conscious meaning but acts as a stable reference that serves as a rest frame for motion perception (Section 12.3.4). Stimuli perceived as figures are processed in greater detail than stimuli perceived as ground.

Various factors affect whether observers view stimuli as figure or ground. Objects tend to be convex and thus convex shapes are more likely to be perceived as figure. Areas lower in the scene are more likely to be perceived as ground. Familiar shapes that represent objects are more likely to be seen as figure. Depth cues such as partial occlusions/interposition, linear perspective, binocular disparity, motion parallax, etc. can have significant impact on what we perceive as figure and ground. Figures are more likely to be perceived as objects and signifiers (Section 25.2.2), giving us clues that the form can be interacted with.

Meta Platforms, Inc.
Exhibit 1063
Page 257

Environmental Design

The environment users find themselves in defines the context of everything that occurs in a VR experience. This chapter focuses on the virtual scene and its different aspects that environmental designers should keep in mind when creating virtual worlds.

21.1 The Scene

The **scene** is the entire current environment that extends into space and is acted within.

The scene can be divided into the background, contextual geometry, fundamental geometry, and interactive objects (similar to that defined by Eastgate et al., 2014). Background, contextual geometry, and fundamental geometry correspond approximately to vista space, action space, and personal space as described in Section 9.1.2.

The background is scenery in the periphery of the scene located in far vista space. Examples are the sky, mountains, and the sun. The background can simply be a textured box since non-pictorial depth cues (Section 9.1.3) at such a distance are non-existent.

Contextual geometry helps to define the environment one is in. Contextual geometry includes far landmarks (Section 21.5) that aid in wayfinding and are typically located in action space. Contextual geometry has no affordances (i.e., can't be picked up; Section 25.2.1). Contextual geometry is often far enough away that it can consist of simple faked geometry (e.g., 2D billboards/textures often used to portray more complex geometry such as trees).

Fundamental geometry consists of nearby static components that add to the fundamental experience. Fundamental geometry includes items such as tables, instructions, and doorways. Fundamental geometry often has some affordances, such as preventing users from walking through walls or providing the ability to set an object upon something. This geometry is most often located in personal space and action space. Because of its nearness, artists should focus on funda-

Meta Platforms, Inc.
Exhibit 1063
Page 258

mental geometry. For VR, 3D details are especially important at such distances (Section 23.2.1).

Interactive objects are dynamic items that can be interacted with (discussed extensively in Part V). These typically small objects are in personal space when directly interacted with and most commonly in action space when indirectly interacted with.

All objects and geometry scaling should be consistent relative to each other and the user. For example, a truck in the distance should be scaled appropriately relative to a closer car in 3D space; the truck should not be scaled smaller because it is in the distance. A proper VR rendering library should handle the projection transformation from 3D space to the eye; artists need not be concerned with such detail. For realistic experiences, include familiar objects with standard sizes that are easily seen by the user (see relative/familiar size in Section 9.1.3). Most countries have standardized dimensions for sheets of paper, cans of soda, money, stair height, and exterior doors.

21.2 Color and Lighting

Color is so pervasive in the world that we often take it for granted. However, we constantly perceive and interact with colors throughout the day whether getting dressed or driving. We associate color to emotional reactions (purple with rage, green with envy, feeling blue) and special meanings (red signifies danger, purple signifies royalty, greens signifies ecology).

Other factors influence our perception of color. For example, the surrounding context/background can influence our perception of an object's color. Exaggerated coloring of the entire scene or extreme lighting conditions can result in loss of lightness and color constancy (Section 10.1.1) and objects seen as being different than they would be under normal conditions. Unless intentional for special situations, content creators should use multiple colors and only slight variations of white light so users perceive the intended color of objects and maintain color constancy.

Color enables us to better distinguish between objects. For example, bright colors can capture users' attention through salience (salience refers to physical properties of a scene that reflexively grabs our attention; Section 10.3.2). Figure 21.1 (left) shows an example of how non-relevant background can be made a dull gray to draw attention to more relevant objects. Figure 21.1 (right) shows how color is turned on, signifying that brain lobes can be grabbed after the system provides audio instructions.

Consider that real-world painters choose from only a little over 1,000 colors (the Pantone Matching System has about 1,200 color choices). This does not mean we only need 1,200 pixel options for VR. Much of the subtle changes in color come from

Meta Platforms, Inc.
Exhibit 1063
Page 259



Figure 21.1 **Color is used as salience to direct attention. The user's eyes are drawn toward the colored items in the scene (left). After the system provides audio instructions, the brain lobes and table are colored to draw attention and to signify they can be interacted with (right).**

gradual changes over surfaces—for example, change in lighting intensity across a surface. However, ~1,000 colors is fine for creating content before lighting is added to the scene (unless one is attempting to perfectly match a color from outside the 1,000-color set).

21.3 Audio

As discussed in Section 8.2, sound is quite complex with many factors affecting how we perceive it. Auditory cues play a crucial role in everyday life as well as VR, including adding awareness of surroundings, adding emotional impact, cuing visual attention, conveying a variety of complex information without taxing the visual system, and providing unique cues that cannot be perceived through other sensory systems. Although deaf people have learned to function quite well, they have a lifetime of experience learning techniques to interact with a soundless world. VR without sound is equivalent to making someone deaf without the benefit of having years of experience learning to cope without sound.

The entertainment industry understands audio well. For example, George Lucas, known for his stunning visual effects, has stated sound is 50% of the motion picture experience. Like a great movie, music is especially good at evoking emotion. Subtle **ambient sound effects** such as birds chirping along with rustling of trees by the wind, children playing in the distance, or the clanking of an industrial setting can provide a surprisingly strong effect on a sense of realism and presence.

When sounds are presented in an intelligent manner, they can be informing and extremely useful. Sounds work well for creating situational awareness and can obtain

Meta Platforms, Inc.
Exhibit 1063
Page 260

attention independent of visual properties and where one is looking. Although sound is important, sound can be overwhelming and annoying if not presented well. It is also difficult to ignore sounds like what can be done with vision (e.g., closing the eyes). Aggressive warning sounds that are infrequent with short durations are designed to be unpleasant and attention-getting, and thus are appropriate if that is the intention. However, if overused such sounds can quickly become annoying rather than useful.

Vital information can be conveyed through spoken language (Section 8.2.3) that is either synthesized, prerecorded by a human, or live from a remote user. The spoken information might be something as simple as an interface responding with "confirmed" or could be as complex as a spoken interface where the system both provides information and responds to the user's speech. Other verbal examples include clues to help users reach their goals, warnings of upcoming challenges, annotations describing objects, help as requested from the user, or to add personality to computer-controlled characters.

At a minimum, VR should include audio that conveys basic information about the environment and user interface. Sound is a powerful feedback cue for 3D interfaces when haptic feedback is not available, as discussed in Section 26.8.

For more realistic audio and where auditory cues are especially important, **auralization**—the rendering of sound to simulate reflections and binaural differences between the ears—can be used. The result of auralization is **spatialized audio**—sound that is perceived to come from some location in 3D space (Section 8.2.2). Spatialized audio can be useful to serve as a wayfinding aid (Section 22.1), to provide cues of where other characters are located, and to give feedback of where the element of a user interface is located. A **head-related transfer function** (HRTF) is a spatial filter that describes how sound waves interact with the listener's body, most notably the outer ear, from a specific location. Ideally, HRTFs model a specific user's ears, but in practice a generic ear is modeled. Multiple HRTFs from different directions can be interpolated to create an HRTF from any direction relative to the ear. The HRTF can then modify a sound wave from a sound source, resulting in a realistic spatialized audio cue to the user.

21.4 Sampling and Aliasing

Aliasing is an artifact that occurs due to approximating data with discrete sampling. With computer graphics, aliasing occurs due to approximating edges of geometry and textures with discretely sampled/rendered pixels. This occurs because the geometry or texture representing the point sample either does or does not project onto the pixel, i.e., a pixel either represents a piece of geometry or color of a texture or it does not.

Meta Platforms, Inc.
Exhibit 1063
Page 261

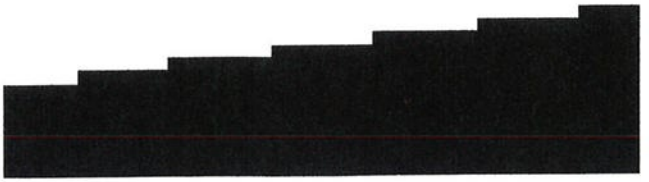


Figure 21.2 **Jaggies/staircasing as seen on a display that represents the edge of an object. Such artifacts result due to discrete sampling of the object.**

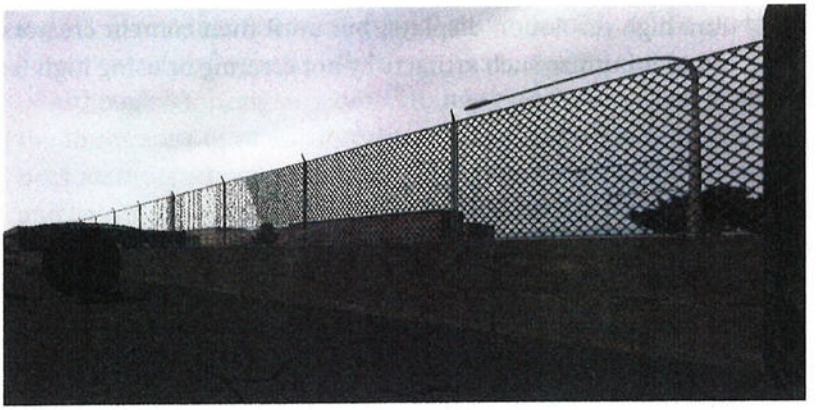


Figure 21.3 **Aliasing artifacts can clearly be seen at the left-center area of the chain-link fence. Such artifacts are even worse in VR due to the artifact patterns continuously moving as a result of viewpoint movement as small as normally imperceptible head motion.** (Image courtesy of NextGen Interactions)

Edges in the environment that are projected onto the display cause discontinuities called "jaggies" or "staircasing," as shown in Figure 21.2. Other artifacts also occur, such as moire patterns, as seen in Figure 21.3. Such artifacts are worse in VR due to the artifact patterns continuously fluctuating as a result of viewpoint movement, each eye being rendered from different views, and the pixels being spread out over a larger field of view. Even subconscious small head motion that is normally imperceptible can be distracting due to the eye being drawn toward the continuous motion of the jaggies and other artifact patterns.

Such artifacts can be reduced through various anti-aliasing techniques, such as mipmapping, filtering, rendering, and jittered multi-sampling. Although such techniques can reduce aliasing artifacts, they unfortunately cannot completely obviate all

Meta Platforms, Inc.
Exhibit 1063
Page 262

artifacts from every possible viewpoint. Some of these techniques also add significantly to rendering time, which can add latency. In addition to using anti-aliasing techniques, content creators can help to reduce such artifacts by not including high-frequency repeating components in the scene. Unfortunately, it is not possible to remove all aliasing artifacts. For example, linear perspective causes far geometry to have high spatial frequency (although artifacts can be somewhat reduced by implementing fog or atmospheric attenuation). Swapping in/out models with different levels of detail (e.g., removing geometry with higher-frequency components when it becomes further from the viewpoint) can help but often occurs at the cost of geometry "popping" into and out of view, which can cause a break-in-presence. The ideal solution is to have ultra-high resolution displays, but until then content creators should do what they can to minimize such artifacts by not creating or using high-frequency components.

21.5 Environmental Wayfinding Aids

Wayfinding aids (Darken and Sibert, 1996) help people form cognitive maps and find their way in the world (Section 10.4.3). Wayfinding aids help to maintain a sense of position and direction of travel, to know where goals are located, and to plan in the mind how to get to those goals. Examples of wayfinding aids include architectural structures, markings, signposts, paths, compasses, etc. Wayfinding aids are especially important for VR because it is very easy to get disoriented with many VR navigation techniques. Virtual turns can be especially disorienting (e.g., turning with a hand controller but not physically turning the body) due to the lack of vestibular cues and other physical sensations of turning. The lack of physically stepping/walking can also cause incorrect judgments of distance.

Fortunately, there is much opportunity for creating wayfinding aids in VR that can't be done in the real world. Wayfinding aids are most often visual but do not need to be. Floating arrows, spatialized audio, and haptic belts conveying direction are examples of VR wayfinding aids that would be difficult to implement in the real world. The aids may or may not be consciously noticed by users but are often useful in either case. Like the real world, there is far too much sensory information in VR for users to consciously take in information from every single element in the environment. Knowing the most important information can be useful for designers to know what to put where to help users find their way.

The remaining portion of this section focuses on environmental wayfinding aids. Section 22.1 focuses on personal wayfinding aids.

Meta Platforms, Inc.
Exhibit 1063
Page 263

Environmental wayfinding aids are cues within the virtual world that are independent of the user. These aids are primarily focused on the organization of the scene itself. They can be overt such as signs along a road, church bells ringing, or maps placed in the environment. They can also be subtle such as buildings, characters traveling in some direction, and shadows from the sun. The sound of vehicles on a highway can be a useful cue when occlusions, such as trees, are common to the environment. Although not typical for VR, smell can also be a strong cue that a user is in a general area. Well-constructed environments do not happen by accident and good level designers will consciously include subtle environmental wayfinding aids even though many users may not consciously notice. There is much that can be done in designing VR spaces that enhances spatial understanding of the environment so users can comprehend and operate effectively.

Although VR designers can certainly create more than what can be done within the limitations of physical reality, we can learn much from non-VR designers. Just because more is possible with VR does not mean we should not understand real spaces and construct space in a meaningful way. Architectural designers and urban planners have been dealing with wayfinding aids for centuries through the relationship between people and the environment. Lynch (1960) found there are similarities across cities as described below.

Landmarks are disconnected static cues in the environment that are unmistakable in form and dissimilar in form from the rest of the scene. They are easy to recognize and help users with spatial understanding. Global landmarks can be seen from about anywhere in the environment (such as a tower), whereas local landmarks provide spatial information closer to the user.

The strongest landmarks are strategically placed (e.g., on a street corner) and often have the most salient properties (highly distinguishable characteristics), such as a brightly colored light or pulsing glowing arrows showing a path. However, landmarks do not need to be explicit. They might be more subtle like colors or lighting on the floor. In fact, landmarks that are overdominating may cause users to move from one location to another without paying attention to other parts of the scene, discouraging acquisition of spatial knowledge (Darken and Sibert, 1996). The structure and form of a well-designed natural environment provides strong spatial cues without clutter.

Landmarks are sometimes directional, meaning they might be perceived from one side but not another. Landmarks are especially important when a user enters a new space, because they are the first things users pay attention to when getting oriented. Landmarks are important for all environments ranging from walking on the ground to

Meta Platforms, Inc.
Exhibit 1063
Page 264

sea travel to space travel. Familiar landmarks, such as a car, aid in distance estimation due to their familiarity and the size of surrounding objects.

Regions (also known as districts and neighorhoods) are areas of the environment that are implicitly or explicitly separated from each other. They are best differentiated perceptually by having different visual characteristics (e.g., lighting, building style, and color). **Routes** (also known as paths) are one or more travelable segments that connect two locations. Routes include roads, connected lines on a map, and textual directions. Directing users can be done via channels. **Channels** are constricted routes, often used in VR and video games (e.g., car-racing games) that give users the feeling of a relatively open environment when it is not very open at all. **Nodes** are the interchanges between routes or entrances to regions. Nodes include freeway exits, intersections, and rooms with multiple doorways. Signs giving directions are extremely useful at nodes. **Edges** are boundaries between regions that prevent or deter travel. Examples of edges are rivers, lakes, and fences. Being in a largely void environment is uncomfortable for many users, and most people like to have regular reassurance they are not lost (Darken and Peterson, 2014). A visual "handrail" is a linear feature of an environment, such as a side of a building or a fence, which is used to guide navigation. Such handrails are used to psychologically constrain movement, keeping the user to one side and traveling along it for some distance.

How these parts of the scene are classified often depends on user capabilities. For walkers, paths are routes and highways tend to be edges, whereas for drivers, highways are routes. For flying users, paths and routes might only serve as landmarks. Often these distinctions are also in the mind of the user, and geometry and cues can be perceived differently by different users or even the same user under different conditions (e.g., when one starts driving a vehicle/aircraft after walking).

Constructing geometric structure is not enough. The user's understanding of overarching themes and structural organization are important for wayfinding. Users should know the structure of the environment for best results, and making such structure explicit in the minds of users is important as it can help give cues meaning, affect what strategies users employ, and improve navigation performance. Street numbers or names in alphabetical order make more sense after understanding streets follow a grid pattern and are named in numerical and alphabetical order.

Once the person builds a mental model of the meta-structure of the world, violations of the metaphor used to explain the model can be very confusing and it should be made explicit when such violations do occur. A person who lived his entire life in Manhattan (a grid-like structure of city blocks) traveling to Washington, DC, will become very confused until the person understands the hub-and-spoke structure of many of

Meta Platforms, Inc.
Exhibit 1063
Page 265

the city's streets, at which point wayfinding becomes easier. However, due to having many more violations of the metaphor, Washington, DC, is much more confusing than New York for most people.

Adding concepts of city-like structures to abstract data such as information or scientific visualization can help users better understand and navigate that space (Ingram and Benford, 1995). Adding cues such as a simple rectangular grid or radial grid can improve performance. Adding paths suggests leading to somewhere useful and interesting. Sectioning data into regions can emphasize different parts of the dataset. However, developers should be careful of imposing inappropriate structure onto abstract or scientific data as users will grasp onto anything they perceive as structure, and that may result in perceiving structure in the data that does not exist in the data itself. Subject-matter experts familiar with the data should be consulted before adding such cues.

21.5.1 Markers, Trails, and Measurement

Markers are user-placed cues. If using a map (Section 22.1.1), markers can be placed on the map (e.g., as colored pushpins) as well as in the environment. This helps users remember which preexisting landmarks are important. **Breadcrumbs** are markers that are dropped often by the user when navigating through an environment (Darken and Sibert, 1993). A **trail** is evidence of a path traveled by a user and informs the person who left the path as well as other users that they have already been to that place and how they traveled to and from there. Trails consisting of individual directional cues (such as footprints) are generally better than non-directional cues. Multiple trails inform of well-traveled paths. Unfortunately, many trails can result in visual clutter. Footprints can fade away after some time to reduce clutter. Sometimes it is better to identify areas searched instead of paths followed.

Data understanding and spatial comprehension via interactive analysis often requires markup and quantification of dataset features within the environment. Different types of markup tools such as paintbrushes and skewers that contain symbolic, textual, or numerical information can be provided to interactively mark, count, and measure features of the environment. Users may want to simply indicate areas of interest for later investigation, or to count the number of vessels exiting a mass (e.g., a tumor) by having the system automatically increment a counter as markers are placed. Linear segments, surface area, and angles can also be measured and the resulting values placed in the environment with a line attached to the markups. Figure 21.4 shows examples of drawing on the world and placing skewers. Figure 21.5 shows an example of measuring aspects the circumference of an opening within a medical dataset.

Meta Platforms, Inc.
Exhibit 1063
Page 266

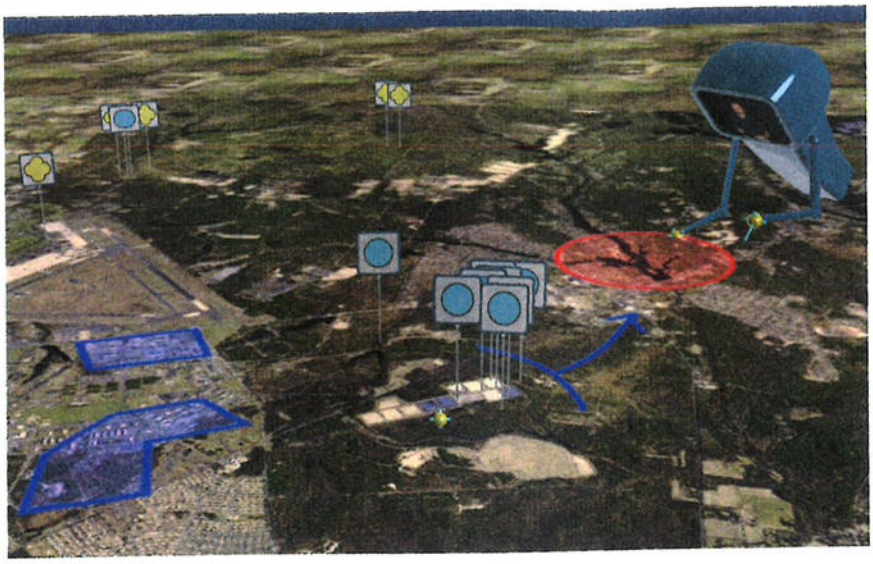


Figure 21.4 **A user observes a colleague marking up a terrain.** (Image courtesy of Digital ArtForms)

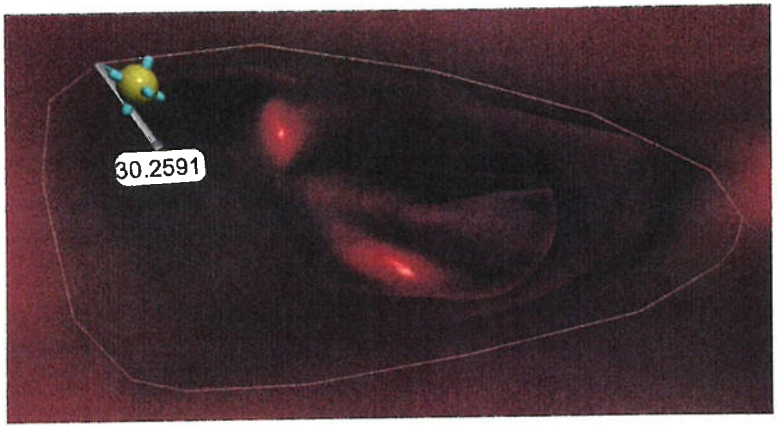


Figure 21.5 **A user measures the circumference of an opening in a CT medical dataset.** (Image courtesy of Digital ArtForms)

21.6 Real-World Data

Data does not necessarily need to be created by artists. One way of creating VR environments is not to build content but to reuse what the real world already provides. Real-world data capture can result in quite compelling experiences.

Meta Platforms, Inc.
Exhibit 1063
Page 267



Figure 21.6 **A 360° image from the VR experience *Strangers with Patrick Watson*.** (© Strangers with Patrick Watson / Félix and Paul Studios)

21.6.1 360° Cameras

Specialized panoramic cameras can be used to capture the world in 360° from one or more specific viewpoints (Figure 21.6). Capturing the world in this way is forcing filmmakers to rethink what it means to experience content. Instead of watching a movie through a "window," viewers of immersive film are in and part of the scene. An example of capturing 360° data that is very different from traditional filming is having to carefully place equipment in a way so that it is not seen by the camera in all directions. In a similar way, all individuals that are not part of the story must leave the set or hide behind objects so they do not unintentionally become part of the story. Those capturing 360° should do so from static poses unless they are extremely careful in controlling the motion of the camera (Section 18.5).

Stereoscopic Capture

Creating stereoscopic 360° content is technically challenging due to the cameras capturing content from only a limited number of views. Such content can work reasonably well with VR when the viewer is seated in a fixed position and primarily only makes yaw (left-to-right or right-to-left) type of head motions. However, if the viewer pitches the head (looks down) or rolls the head (twists the head while still looking straight ahead), then the assumptions for presenting the stereoscopic cues no longer hold and the viewer will perceive strange results. A common method of dealing with the pitch problem is to have the disparity between the left and right eyes in the upper and lower parts of the scene converge. This reduces discomfort at the cost of everything above the viewer and below the viewer appearing far away. Another method is to smooth out the ground beneath the viewer to a single color such that it contains no depth cues. Or a virtual object can be placed beneath the viewer to prevent seeing the real-world data

Meta Platforms, Inc.
Exhibit 1063
Page 268

at that location. These methods all work well enough when the content of interest is not above or below the viewer.

Another solution for the stereo problem is simply not to show the captured content in stereo. This makes content capture a lot easier and reduces artifacts. Computer-generated content that has depth information can be added to the scene but will always appear in front of the captured content.

21.6.2 Light Fields and Image-Based Capture/Rendering

Capture from a single location cannot properly convey a fully immersive VR experience when users move their heads—for example, when leaning left or right. A **light field** describes light that flows in multiple directions through multiple points in space. An array of cameras can be coupled with light field capture and rendering techniques to portray multiple viewpoints from different locations, including those not originally captured (Gortler et al., 1996; Levoy and Hanrahan, 1996). Examples include a system that captures a static scene with an outward-looking spherical capture device (Debevec et al., 2015) and a system for portraying animated stop motion puppets by capturing sequences of a 360° ring of images around each puppet (Bolas et al., 2015).

21.6.3 True 3D Capture

Depth cameras or laser scanners capture true 3D data and can result in more presence-inducing experiences at the cost of the scene not seeming quite as real due to artifacts such as gaps or skins (interpolating color where the camera can't see). Data used from multiple depth cameras can be used to help reduce some of these artifacts. Data acquired from 3rdTech laser scanners is used for crime scenes where investigators can go back and take measurements after data has been collected (Figure 21.7). When 3rdTech CEO Nick England (personal communication, June 4, 2015) viewed himself in an HMD lying dead on the floor in a mock murder scene, he witnessed quite a jarring out-of-body experience.

21.6.4 Medical and Scientific Data

Figure 21.8 shows an image from iMedic—immersive medical environment for distributed interactive consultation (Mlyniec et al., 2011). iMedic enables real-time interactive exploration of volumetric datasets by "crawling" through the data with a 3D multi-touch interface (Section 28.3.3). The mapping of voxel (3D volumetric pixels) source density values to visual transparency can also be controlled via a virtual panel held in the non-dominant hand, enabling real-time changes of what structures can be seen in the dataset.

Meta Platforms, Inc.
Exhibit 1063
Page 269