Second Edition

HANDBOOK OF VIRTUAL ENVIRONMENTS

Design, Implementation, and Applications

Edited by

Kelly S. Hale
Kay M. Stanney

CRC Press
Taylor & Francis Group

Second Edition

HANDBOOK OF VIRTUAL ENVIRONMENTS

Design, Implementation, and Applications

Human Factors and Ergonomics Series

PUBLISHED TITLES

Around the Patient Bed: Human Factors and Safety in Health Care
Y. Donchin and D. Gopher

Conceptual Foundations of Human Factors Measurement
D. Meister

Content Preparation Guidelines for the Web and Information Appliances: Cross-Cultural Comparisons
H. Liao, Y. Guo, A. Savoy, and G. Salvendy

Cross-Cultural Design for IT Products and Services
P. Rau, T. Plocher and Y. Choong

Data Mining: Theories, Algorithms, and Examples
Nong Ye

Designing for Accessibility: A Business Guide to Countering Design Exclusion
S. Keates

Handbook of Cognitive Task Design
E. Hollnagel

The Handbook of Data Mining
N. Ye

Handbook of Digital Human Modeling: Research for Applied Ergonomics and Human Factors Engineering
V. G. Duffy

Handbook of Human Factors and Ergonomics in Health Care and Patient Safety Second Edition
P. Carayon

Handbook of Human Factors in Web Design, Second Edition
K. Vu and R. Proctor

Handbook of Occupational Safety and Health
D. Koradecka

Handbook of Standards and Guidelines in Ergonomics and Human Factors
W. Karwowski

Handbook of Virtual Environments: Design, Implementation and Applications, Second Edition,
K. S. Hale and K M. Stanney

Handbook of Warnings
M. Wogalter

Human–Computer Interaction: Designing for Diverse Users and Domains
A. Sears and J. A. Jacko

Human–Computer Interaction: Design Issues, Solutions, and Applications
A. Sears and J. A. Jacko

Human–Computer Interaction: Development Process
A. Sears and J. A. Jacko

Human–Computer Interaction: Fundamentals
A. Sears and J. A. Jacko

The Human–Computer Interaction Handbook: Fundamentals Evolving Technologies, and Emerging Applications, Third Edition
J. A. Jacko

PUBLISHED TITLES (CONTINUED)

Human Factors in System Design, Development, and Testing
D. Meister and T. Enderwick

Introduction to Human Factors and Ergonomics for Engineers, Second Edition
M. R. Lehto

Macroergonomics: Theory, Methods and Applications
H. Hendrick and B. Kleiner

Practical Speech User Interface Design
James R. Lewis

The Science of Footwear
R. S. Goonetilleke

Skill Training in Multimodal Virtual Environments
M. Bergamsco, B. Bardy, and D. Gopher

Smart Clothing: Technology and Applications
Gilsoo Cho

Theories and Practice in Interaction Design
S. Bagnara and G. Crampton-Smith

The Universal Access Handbook
C. Stephanidis

Usability and Internationalization of Information Technology
N. Aykin

User Interfaces for All: Concepts, Methods, and Tools
C. Stephanidis

FORTHCOMING TITLES

Cognitive Neuroscience of Human Systems Work and Everyday Life
C. Forsythe and H. Liao

Computer-Aided Anthropometry for Research and Design
K. M. Robinette

Handbook of Human Factors in Air Transportation Systems
S. Landry

Variability in Human Performance
T. Smith, R. Henning, and M. Wade

This page intentionally left blank

Second Edition

HANDBOOK OF VIRTUAL ENVIRONMENTS

Design, Implementation, and Applications

Edited by

Kelly S. Hale
Kay M. Stanney

CRC Press
Taylor & Francis Group
Boca Raton London New York

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

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20140619

International Standard Book Number-13: 978-1-4665-1185-9 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

This page intentionally left blank

Perspective by Jaron Lanier

The public remains as entranced by a romantic ideal of virtual reality (VR) as ever. VR the ideal, as opposed to the real technology, weds the nerdy thing with the hippie mystic thing; it's high tech and like a dream or an elixir of unbounded experience, all at the same time.

I wish I could fully convey what it was like in the early days. There was a feeling, in the early 1980s, of opening up a new plane of experience. Inhabiting the first immersive avatars, seeing others as avatars, experiencing one's body for the first time as a nonrealistic avatar; these things transfixed us. Everything else in the world seemed dull in comparison.

The irony is that I cannot use VR to share what that experience was like with you. VR, for all it can do, is not a medium of internal states. There might never be such a medium. There is little need for me to make this point in an introduction to a book for VR professionals, but it's a clarification that I'm sure many of the contributors and readers of this book have been called upon to give many times. What the public seems to want from VR is telepathic conjuring of arbitrary reality. It can be difficult to explain that VR is wonderful for what it is, precisely because it isn't magic.

Eventually a new kind of culture, a massive tradition of tricks of the VR trade, might arise and that culture might allow me to convey something to you about how early VR felt, using VR-borne metaphors. I have spent many hours daydreaming about what a mature culture of expression would be like in VR. A cross between cinema, jazz, and programming, I used to say.

Even though we don't know how expressive VR might eventually become, there is always that little core of thrill in the very idea of VR. Arbitrary experience, shared with other people, conversationally, under our control. An approach to a holistic form of expression. Shared lucid dreaming. A way out of the dull persistence of physicality. This thing we seek, it's a way of being that isn't tied just to our circumstances in this world.

Every few years a new wave of public VR mania rises, yet somehow very few people seem to realize how often it's happened before. Long ago, a scrappy little startup called VPL Research seemed for a moment like a much larger company, a real force. The same kind of thing can still happen; it is happening once again at the time of this writing with a startup selling an HMD called Oculus Rift.

The actual experience of high-quality VR remains elusive, however. You really still have to have access to a top lab to experience the good stuff. By "the good stuff" I mean a system good enough to fool you, to engage your whole body, to include others as avatars with you in there, to be usable in the long term, and one that gives you enough to do to outlast the first few demos. There are remarkably few places on Earth today where one can experience anything close to the whole package.

However, such places exist. This is well known to lucky students who work with Jeremy Bailenson at Stanford, with Henry Fuchs or Fred Brooks at the University of North Carolina, Chapel Hill, or with Carolina Cruz-Neira at the University of Louisiana, Lafayette—or at any of the other fine VR research labs. Then there is the even less accessible world of training and design centers that use VR. To my knowledge, no one has ever thought to maintain an atlas of VR systems, but there are still probably only a few thousand decent VR systems in the world as I write this Perspective in 2013.

The VR known to researchers is sometimes a little like what the public dreams of, but there is a colossal difference, because in research labs, VR is a practical, if expensive, tool. Researchers continue to learn more about the human organism by using VR as an experimental device to test hypotheses in cognitive science, biomechanics, and other disciplines. VR research continues to spin off useful technologies, such as motion capture suits, surgical simulators, and vehicle design simulators. From the point of view of pop culture, the realization of the dream remains elusive, while to practitioners, our progress has been satisfying.

Nonetheless, the cost of the parts we use gradually lowers, and a horizon comes into view. Over that horizon lies the long-anticipated popularization of VR, meaning a form of VR that would be complete enough to feel like a protean dream, at least a little like The Matrix, if you squint. The world waits for this.

Much of these could have been written at any time in the last thirty years. In earlier times, we might have mentioned Star Trek's Holodeck instead of The Matrix, but the pattern was similar. We've been at work, while the public frets that what we do hasn't reached its potential yet. The popular circumstances in which we work have remained about the same even as we have learned more, and successive generations of researchers have come to the fore. Journalists periodically write about how VR has "failed to deliver on its promises." But this last year has been different.

Consumer devices related to VR are appearing with enough frequency that they no longer feel exceptional. The last few years have seen Microsoft's Kinect success in the marketplace and Google's Glass public trials, for instance. Neither device is a VR system, but both are consumer gadget versions of components that are cousins to what would be needed in full-on VR systems. Both have changed the public's expectations of information technology.

It does appear that we are about to arrive at the long-anticipated threshold. Even so, it could still take some years for the details to come into alignment. Thirty years ago, my colleagues and I used to guess that a reasonably comprehensive form of VR would be popularized around 2020, based on Moore's Law and related trends, and perhaps we will turn out to have been a little pessimistic.

Those of us who have worked on VR for many years wonder what it will be like once VR becomes both cheap and well-wrought, and the novelty factor has eroded to nothing. Will it be banal—like plumbing or ever new—like books? This is always the question with a new layer of digital capability.

So, a drama of anticipation builds. It is like wondering how a child will turn out. Have we been building something banal or beautiful? Reality can only be known empirically. We don't yet really know what we're building. We don't know to what degree it will be a stream of vital passion and poetry, a banal waste of time, or a cruel gauntlet for manipulation and bullying. Any human medium will be all these things, but in what proportions?

I am currently feeling confident in our hardware and not so confident in our software. Hardware first.

Unsolved problems remain in VR hardware. Some of the problems are shared in common with other technologies. Most approaches to VR involve wearable components, but batteries still weigh too much, for instance. Therefore, VR components tend to either be heavy or need to be charged too often. Trends in battery technology are positive, so we expect this problem to be addressed.

But VR also poses some unique hardware challenges. What is the right haptic feedback design for general popularization? What is the best general scheme for reconciling the specific limited physical space a user can explore at a given time with the unbounded, beckoning virtual spaces that might be presented? Researchers have presented hundreds of good ideas for answers to both questions, but there is no consensus because there haven't yet been adequate, widespread tests of the most promising ideas.

VR intrinsically presents conundrums. The best known one might be this: To fill the human visual field with visual information, it is common to place a source of photons near the eyes, in the form of an HMD. But for people to see each other as avatars, the facial expressions should be measured so as to have an influence on one's avatar. After all, we present ourselves visually using our faces most of all. But HMDs often obscure or interfere with facial expressions. To resolve this difficulty, a clever combination of optics, industrial design, and graphics/machine vision will probably be needed.

VR also intrinsically pushes requirements, for the simple reason that the human organism is remarkable. We are sensitive to minute latencies, and in some cases to tiny inconsistencies in our sensory data. Fortunately, as everyone who works with it knows, VR is based on finding ways to present what should by all rights be inadequate equipment in a way that somehow meets the expectations of the human nervous system. Our field can be considered as the study of highly advanced stage magic without the stage.

Even so, we sometimes find a challenge we must meet head-on. For instance, to obscure items in the real world in order to show virtual content instead (in a mixed- or augmented-reality presentation), presents a healthy challenge to both optics engineers, machine vision scientists, and systems engineers doing battle with latency.

Fortunately, the public's fascination with our work creates an interesting external laboratory that is not available in many other research fields. The marketplace can support widespread testing of hardware components in the form of entertainment products. For instance, real-time human motion sensing in a form that would be suitable for VR systems has also been packaged independently in gaming devices like Matte's Power Glove and Microsoft's Kinect. With each market test we learn more about what a popular, complete VR system might be like. Fortunately, these tests have yielded positive results in many cases.

The hardware side of VR progresses reliably towards fulfilling the collective dream of VR, it seems. Each year, new and better optical and display strategies appear; same for new sensors and actuators. The mega-reorientation of consumer computing towards portable devices has resulted in vast improvements to components like microdisplays and inertial motion units, so VR researchers have lately been showered with hardware gifts.

But no one is helping us with our software, and there are two huge reasons for concern that our software might not turn out as well as our hardware. The first reason is that previous attempts to define software architectures for VR have all turned out to be too specialized to support a broad community. The second is that as we approach our long-anticipated threshold, the world at large is entering into what might be called a surveillance economy that could render our dreams into creepy nightmares.

One problem with VR software architecture is that VR can hypothetically align with a huge variety of human experiences. Software on a personal computer was stuck on the screen, and the person using it was stuck on a desk, with a hand on a mouse. Every program would be reliably circumscribed by those conditions, so it was possible to declare a significant commonality between programs that otherwise had no relationship. Both Mathematica and Angry Birds can be scrolled in a window and pricked by a mouse. More recently, approximately the same kinds of rules have applied to mobile devices; now programs assume multitouch, for instance. But the background in which programs are defined is still the class of device, not the world at large.

VR isn't defined on an artifact. The background is the human body and—if the system is presented for mixed or augmented reality—the physical environment. Neither are human inventions, and both are still under study. We will never have complete knowledge of our physical environment or of our bodies.

So the first job of a VR architecture is to ignore almost everything about reality in order to choose a few hooks that can be used to align virtuality. But what should be the preferred blindness?

Different VR software architectures choose quite different emphases. A major divide can be observed between attempts to interface individuals as well as possible and attempts to interface large numbers of people.

Designs that emphasize increasing the number of participants typically do so at the expense of the interaction model. Examples are military ground combat simulators and popular network sites with a VR flavor, like Second Life. In these systems, users move about, talk to each other, lob projectiles, and so on, but there is little use for detailed hand tracking. You couldn't conduct a dance lesson in such a system.

Then there are designs that take great care to measure and integrate human activity carefully. These include, on a professional level, surgical simulators as well as entertainment systems that use depth cameras or other means to measure human motion comprehensively in real time, like Kinect.

Designers of these systems have their hands full, so to speak, and don't deal with the many issues that come up with massive numbers of users distributed across large distances.

These two schools are not the end of it. There are also designs that emphasize representing the physical world and presenting that to users, but are poor at measuring the users or allowing users to interact. This stream of design might even be the largest. It started long ago with forgotten attempts at standards like VRML and Java3D, and continues today with cloud services that integrate street view, satellite, and other image sources to generate streaming models of the world. It is never elegant to inhabit a virtual world driven by an architecture that is focused obsessively on data rather than on the experience of data.

Here we find another example of how our field can sometimes seem to stand still for decades, even though we're all working very hard. For decades there has been endless talk about combining the different approaches to networked VR software architecture. Wouldn't it be grand to be able to have dance lessons over a service resembling Second Life, using something like Kinect, in a simulation of a historic dance studio, gathered by something like Photosynth?

I have been involved in all three projects, and the simple truth is that just to get those out the door well involved total devotion, total exhaustion, really, for the teams who had to get the jobs done. There was no further energy or time left to figure out dreams of integration or generalization. We don't know how hard it will be to "get it all" in one network VR architecture, but we know it will be hard.

I hope we do it in the next few years. I don't want to have to be carrying four different headsets with me in ten years, each specialized to a different sort of task.

This brings up a correlate to the unbounded nature of VR. Everyone understands that not every Android app will run on an iPhone. But are we ready for a world where certain customers can see certain augmentations of physical reality while others cannot?

Will we enter into a situation in which people walk down the street and see entirely different augmentations of the world depending on which vendor they bought their headsets from? Sherry Turkle wrote a book called "Alone Together" about how people can become oddly insulted by using network technology, which was certainly not the intent of the designers. I worry that mixed reality systems used in public could result in an "Apart Together" situation, in which people are adjacent, but not able to experience much in common, even when they are trying to.

This eventuality could be the result of cloud service rivalries, carrier wars, or—and this becomes our problem—architectural specializations that corral users into disjoint communities. Quests for perfection in computer architectures tend to backfire, but that doesn't mean we have to resign ourselves to worst case scenarios. If the research community can prove that an architecture can come into existence that supports a wide range of scenarios, then the political fight over whether the various companies will play nice with each other can at least have a chance to play out.

Finally, the rise of creepiness must be addressed. VR is the most intimate communication technology. In order to work well, VR instrumentation senses the saccade of the eyes and the turn of the lip. What we do is interface intimately with the human organism.

It happens, however, that as we approach the day when the fruits of our labors become inexpensive enough for everyone to enjoy, another commercial trend is on the rise that would use these data against our users, our customers.

It is now commonplace for people to accept constant surveillance in exchange for so-called bargains or convenience. Most people have accounts on social networks and trust companies to keep track of many of their life memories, for free. We have also recently learned that governments routinely piggyback on these arrangements to maintain their own surveillance. The data derived from surveillance are used commercially to power big data systems that, in turn, micromanage some of the options placed in front of people in the form of paid links. Some feel this is a helpful economic arrangement, while I do not. In any case, everyone should be concerned that the surveillance economy is creating astonishing potential for abuses.

It is one thing for remote corporations or governments to have access to the metadata about who one has been in touch with or one's recent purchases. But it is another entirely for remote powers to compile information about how your body responds to stimuli. There is then a potential for neo-Pavlovian manipulation, for instance, where a person might not realize that a manipulation had even taken place.

Many new consumer products will not function if a user does not "click through" an agreement sending their experiences to the company's servers. In general, all tech companies, and all governments, have come to expect people to click through such agreements and accept surveillance.

I had always thought I could keep politics and VR research separate, but that is no longer true. Our community has to consider how to protect our eventual users. Should sensors of the most important signals in the human body encrypt their results as part of the measurement? Should we be baking in privacy at the lowest level to make man-in-the-middle attacks harder? I'd rather not have to think about such questions, but I do, we all do.

This is a wonderful, but challenging time for the VR community. I have lately become more excited than ever to meet new students who have decided to enter our field because I know their careers will be filled with delightful surprises and thrills. I also worry. Will they come to feel that they were part of the problem instead of a solution to the mounting wave of cyber-creepiness?

The story unfolds. Our time has arrived.

Jaron Lanier
Interdisciplinary Scientist
Microsoft Research

This page intentionally left blank

Preface

The first edition of the *Handbook of Virtual Environments* was published in 2000. A lot has changed since that time. Virtual environments (VEs) have always presented infinite possibilities for extending the interactivity and engagement of an ever-expanding variety of applications in the areas of entertainment, education, medicine, and beyond. But now, with the technology advancing at rates exceeding conventional laws, the possibility to create VEs that provide true-to-life experiences has never been more attainable. We thus felt it was time to assemble a second edition of the handbook to summarize the current cutting-edge research and technology advances that are bringing these truly compelling virtual worlds to life.

This second edition includes nine new as well as 41 updated chapters that reflect the progress made in basic and applied research related to the creation, application, and evaluation of VEs. Leading researchers and practitioners from multidisciplinary domains have contributed to this second edition to provide a wealth of both theoretical and practical information, resulting in a complete toolbox of theories and techniques that researchers, designers, and developers of VEs can rely on to develop more captivating and effective virtual worlds.

Section I consists of two chapters. Chapter 1 provides a thorough summary of VEs in the twenty-first century, while Chapter 2 provides an updated glossary of terms to promote common language throughout the community. Section II focuses on system requirements, both hardware and software. Individual chapters are provided for various modalities of presentation, highlighting known techniques for eliciting and manipulating visual (Chapter 3), auditory (Chapter 4), and haptic perception (Chapter 5), as well as perception of body motion (Chapter 7). This second edition also includes a new contribution examining olfactory perception and display design (Chapter 6). In addition, interaction methods that support optimized human–system collaboration are reviewed, including eye tracking (Chapter 8), gesture (Chapter 9), and locomotion systems (Chapter 10). These chapters highlight advances in hardware design and capabilities to support natural interaction that increases presence and human performance within VEs. Four chapters in this section focus on software design of VEs, including development of VE models (Chapter 11), designing effective interaction techniques from the software perspective (Chapter 12), simulating and optimizing VE realism where and when needed (Chapter 13), and design of embodied autonomous agents (Chapter 14).

Section III focuses on design approaches and implementation strategies for VEs. Each chapter in this section has been updated to reflect advances made in the last decade and provides a solid foundation for new and experienced designers. Chapter 15 summarizes the VE development structure, a standardized process for designing and developing VE technology. Design implications for human cognition (Chapter 16), modeling multimodal interaction (Chapter 17), illusory self-motion (Chapter 18), spatial orientation and wayfinding (Chapter 19), technology management for user acceptance (Chapter 20), and product liability (Chapter 21) round out this informative section regarding design of VEs.

Section IV focuses on health and safety issues, which have been observed since the earliest VE systems were created. The chapters in this section examine the direct effects of VEs on users (Chapter 22) and provide a thorough review of motion sickness symptomatology and origins (Chapter 23) and methods to measure motion sickness (Chapter 24) within a VE. The section also includes a review of adaptation effects and their implications (Chapter 25), as well as discussions of visually induced motion sickness (Chapter 26) and the social impact of VE technology (Chapter 27).

Section V focuses on evaluation of VE systems. It begins with a review of usability evaluation techniques for VEs (Chapter 28), summarizing how traditional usability methods may be adapted and extended to evaluate 3D spaces and experiences. Chapter 29 summarizes methods for measuring human performance within VEs, while the addition of Chapter 30 provides guidance on evaluating the effectiveness of VE training systems. VE usage protocols are summarized in Chapter 31, which outlines methods to minimize negative impacts of VE exposure. Methods to evaluate visual aftereffects from VE exposure (Chapter 32), proprioceptive adaption and aftereffects (Chapter 33), presence experienced while in a VE (Chapter 34), and a new chapter highlighting the potential of utilizing an augmented cognition system to evaluate VE systems (Chapter 35) round out this section of the handbook.

Section VI, which focuses on the application of VE technology, has been substantially updated to reflect the tremendous progress made over the last decade in applying VE technology to a growing number of domains. Chapter 36 provides an overview of application areas, while follow-on chapters discuss applications in national defense (Chapter 37) and various training and learning domains, including team training (Chapter 38), perceptual skills training (Chapter 39), conceptual learning (Chapter 40), experimental, STEM, and health sciences learning (Chapter 41), special needs education (Chapter 42), and cultural training (Chapter 46). Additionally, reviews of VEs in the application areas of assisted teleoperation (Chapter 43), human–robot communication (Chapter 44), clinical settings (Chapter 45), geology (Chapter 47), information visualization of big data (Chapter 48), and entertainment (Chapter 49) are provided.

The handbook concludes with Section VII, which provides a review of VE research and advances made since the inception of the technology, as well as a list of references that can provide additional information for VE professionals (Chapter 50).

We are very pleased to have the opportunity to bring together the vast expertise and knowledge provided in this handbook. We trust you will find this compendium a valuable resource in the advancement of VE applications as you take them from the laboratory to the real-world lives of people everywhere.

Kelly S. Hale
Kay M. Stanney

Editors

Dr. Kelly S. Hale is senior vice president of technical operations at Design Interactive, Inc., a woman-owned small business focused on human–systems integration. Her research and development efforts in human–systems integration have examined virtual environment design and evaluation, augmented cognition, multimodal interaction and haptic interfaces, and training sciences.

Over the past 10 years, Dr. Hale has guided Design Interactive, Inc. to be a leader in developing advanced performance metrics and diagnostic capabilities, including behavioral, physiological, and neural metric suites for capturing traditionally *unobservable* behavioral data to evaluate perceptual skills and cognitive processing in real time by synchronizing data with system events/areas of interest dynamically throughout simulated scenarios. She holds a patent for a tactile display language and serves on the editorial boards for the *International Journal of Human–Computer Interaction* and *Theoretical Issues in Ergonomics Science*. Dr. Hale received her BSc in kinesiology (ergonomics option) from the University of Waterloo in Ontario, Canada, and her master's degree and PhD in industrial engineering, with a focus on human factors engineering, from the University of Central Florida.

Dr. Kay Stanney founded Design Interactive, Inc. (DI) in 1998 and serves as president. She provides executive leadership and strategic direction, formulating and driving key business strategies across DI's three business units. Dr. Stanney is recognized as a world leader in virtual environment (VE) technology, especially as it relates to training. During her more than 25 years of carrying out research in the area of VE training, she has led numerous efforts involved in furthering adaptive VE training techniques. She is the recipient of the 2013 Women Who Mean Business Award, 2012 and 2011 Top Simulation & Training Company for innovation, and the IEEE VGTC Virtual Reality Technical Achievement Award.

Dr. Stanney has an MS and PhD in human factors engineering from Purdue University and a BS in industrial engineering from SUNY Buffalo and is a Certified Human Factors Practitioner.

This page intentionally left blank

Contributors

David R. Badcock
School of Psychology
University of Western Australia
Crawley, Western Australia, Australia

Daniel Barber
Institute for Simulation and Training
University of Central Florida
Orlando, Florida

Angelos Barmpoutis
Digital Worlds Institute
University of Florida
Gainesville, Florida

Kathleen M. Bartlett
DSCI MESH Solutions, LLC
Orlando, Florida

Cagatay Basdogan
College of Engineering
Koc University
Istanbul, Turkey

Elizabeth Sheldon Biddle
Human Systems Department
Naval Air Warfare Center Training Systems Division
Orlando, Florida

Mark Billinghurst
University of Canterbury
Christchurch, New Zealand

Richard A. Blade
IPI Press
Aurora, Colorado

James P. Bliss
Department of Psychology
Old Dominion University
Norfolk, Virginia

Christoph W. Borst
Emerging Markets and Technologies Division
Center for Advanced Computer Studies
University of Louisiana at Lafayette
Lafayette, Louisiana

Doug A. Bowman
Center for Human-Computer Interaction
Virginia Tech
Blacksburg, Virginia

Grigore (Greg) Burdea
Department of Electrical and Computer Engineering
Rutgers, The State University of New Jersey
Piscataway, New Jersey

Sandra L. Calvert
Department of Psychology
Georgetown University
Washington, DC

Meredith Carroll
Defensive Solutions Division
Design Interactive, Inc.
Oviedo, Florida

Roberto K. Champney
Medical Innovations Division
Design Interactive, Inc.
Oviedo, Florida

Eric T. Chancey
Department of Psychology
Old Dominion University
Norfolk, Virginia

Bradley Chase
Department of Engineering
University of San Diego
San Diego, California

Aashish Chaudhary
Scientific Computing Division
Kitware, Inc.
Clifton Park, New York

Ryad Chellali
Istituto Italiano di Tecnologia
Genova, Italy

Dustin Chertoff
Intelligent Automation, Inc.
Atlanta, Georgia

Sue Cobb
Virtual Reality Applications Research Team
University of Nottingham
Nottingham, United Kingdom

Joseph V. Cohn
Human and Bioengineered Systems Division
Office of Naval Research
Arlington, Virginia

Philippe Coiffet
National Academy of Technologies of France
Paris, France

Daniel S. Coming
Google, Inc.
Mountain View, California

Jeffrey L. Cowgill, Jr.
Department of Psychology
Wright State University
Dayton, Ohio

Peter R. Culmer
School of Mechanical Engineering
University of Leeds
Leeds, United Kingdom

Rudolph P. Darken
Department of Computer Science
Naval Postgraduate School
Monterey, California

Mirabelle D'Cruz
Virtual Reality Applications Research Team
University of Nottingham
Nottingham, United Kingdom

Sara Dechmerowski
Medical Innovations Division
Design Interactive, Inc.
Oviedo, Florida

Patricia Denbrook
DCS Inc.
Washington, DC

Benjamin DeVane
Learning Sciences
University of Iowa
Iowa City, Iowa

Nuray Dindar
College of Engineering
Koc University
Istanbul, Turkey

Paul DiZio
Ashton Graybiel Spatial Orientation Laboratory
Volen Center for Complex Systems
Brandeis University
Waltham, Massachusetts

Tripp Driskell
Institute for Simulation and Training
University of Central Florida
Orlando, Florida

Richard M. Eastgate
Virtual Reality Applications Research Team
University of Nottingham
Nottingham, United Kingdom

Joseph R. Fanfarelli
Institute for Simulation and Training
University of Central Florida
Orlando, Florida

Andrew Feng
Institute for Creative Technologies
University of Southern California
Playa Vista, California

Neal M. Finkelstein
US Army Research Laboratory
Orlando, Florida

Joseph L. Gabbard
Grado Department of Industrial and Systems Engineering
Virginia Tech
Blacksburg, Virginia

Robert H. Gilkey
Department of Psychology
Wright State University
Dayton, Ohio

Adams Greenwood-Ericksen
Department of Game Studies
Full Sail University
Orlando, Florida

Tami Griffith
Simulation and Training Technology Center
US Army Research Laboratory
Orlando, Florida

David Gross
Lockheed Martin
Marietta, Georgia

Kelly S. Hale
Design Interactive, Inc.
Oviedo, Florida

Frank P. Hannigan
Design Interactive, Inc.
Oviedo, Florida

Tessa Hawkins
Virtual Reality Applications Research Team
University of Nottingham
Nottingham, United Kingdom

Heiko Hecht
Department of Psychology
Johannes Gutenberg University
Mainz, Germany

Lawrence J. Hettinger
Center for Behavioral Sciences
Liberty Mutual Research Institute for Safety
Hopkinton, Massachusetts

Charles E. Hughes
Department of Electrical Engineering and Computer Science
University of Central Florida
Orlando, Florida

Matthew Johnston
Emerging Markets and Technologies Division Center for Advanced Computer Studies
Design Interactive, Inc.
Oviedo, Florida

David L. Jones
Medical Innovations Division
Design Interactive, Inc.
Oviedo, Florida

Brenna Kelly
MESH Solutions, LLC
Orlando, Florida

Kristyne E. Kennedy
CSK Legal
Orlando, Florida

Robert C. Kennedy
Department of Game Studies
Full Sail University
and
RSK Assessments, Inc.
Orlando, Florida

Robert S. Kennedy
RSK Assessments, Inc.
Orlando, Florida

Behrang Keshavarz
Research Department (iDAPT)
Toronto Rehabilitation Institute
Toronto, Ontario, Canada

G. Drew Kessler
SRI International
Princeton, New Jersey

Abderrahmane Kheddar
CNRS-AIST Joint Robotics Laboratory
Centre National de la Recherche Scientifique
National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan

and

Interactive Digital Human Group
Centre National de la Recherche Scientifique
Montpellier, France

Gary L. Kinsland
School of Geosciences
University of Louisiana at Lafayette
Lafayette, Louisiana

Sebastian Koenig
Institute for Creative Technologies
University of Southern California
Playa Vista, California

Regis Kopper
Pratt School of Engineering
Duke University
Durham, North Carolina

Stephanie Lackey
Institute for Simulation and Training
University of Central Florida
Orlando, Florida

James R. Lackner
Ashton Graybiel Spatial Orientation Laboratory
Volen Center for Complex Systems
Brandeis University
Waltham, Massachusetts

Belinda Lange
Institute for Creative Technologies
University of Southern California
Playa Vista, California

Ben D. Lawson
US Army Aeromedical Research Laboratory
Fort Rucker, Alabama

Margaux Lhommet
Institute for Creative Technologies
University of Southern California
Playa Vista, California

Albert Y.M. Lin
Qualcomm Institute
University of California, San Diego
La Jolla, California

Robert W. Lindeman
Worcester Polytechnic Institute
Worcester, Massachusetts

Robb Lindgren
College of Education
University of Illinois at Urbana-Champaign
Champaign, Illinois

J. Peter A. Lodge
Department of Hepatobiliary and Transplant Surgery
St James's University Hospital
Leeds, United Kingdom

Valerie Lugo
Medical Innovations Division
Design Interactive, Inc.
Oviedo, Florida

Stacy Marsella
Institute for Creative Technologies
University of Southern California
Playa Vista, California

Douglas Maxwell
Simulation and Training Technology Center
US Army Research Laboratory
Orlando, Florida

James G. May
Department of Psychology
University of New Orleans
Lakefront, New Orleans

Ryan P. McMahan
Department of Computer Science
University of Texas at Dallas
Richardson, Texas

Laura Millen
Virtual Reality Applications Research Team
University of Nottingham
Nottingham, United Kingdom

Betty J. Mohler
Max Planck Institute for Biological Cybernetics
Tübingen, Germany

Mark Mon-Williams
School of Psychology
University of Leeds
Leeds, United Kingdom

J. Michael Moshell
School of Visual Arts and Design
University of Central Florida
Orlando, Florida

Allen Munro
University of Southern California
Los Angeles, California

Denise Nicholson
DSCI MESH Solutions, LLC
Orlando, Florida

Margaret Nolan
DSCI MESH Solutions, LLC
Orlando, Florida

Razia Oden
AMTIS, Inc.
Oviedo, Florida

James C. Oliverio
Digital Worlds Institute
University of Florida
Gainesville, Florida

Eric Ortiz
Institute for Simulation and Training
University of Central Florida
Orlando, Florida

Mary Lou Padgett
Life Senior Member, IEEE
Marietta, Georgia

Robert Page
US Naval Research Laboratory
Washington, DC

Stephen Palmisano
School of Psychology
University of Wollongong
Wollongong, News South Wales, Australia

Jim Patrey
Human Systems Department
Naval Air Warfare Center Aircraft Division
Orlando, Florida

Barry Peterson
Department of History
University of Nevada, Reno
Reno, Nevada

William Pike
Simulation and Training Technology Center
US Army Research Laboratory
Orlando, Florida

Nicholas F. Polys
Interdisciplinary Center for Applied Mathematics
Virginia Tech
Blacksburg, Virginia

George V. Popescu
National Institute for Laser, Plasma and Radiation Physics
Bucharest, Romania

B.J. Price
Department of Engineering
Walt Disney World
Orlando, Florida

Alexandra B. Proaps
Department of Psychology
Old Dominion University
Norfolk, Virginia

Lauren Reinerman-Jones
Institute for Simulation and Training
University of Central Florida
Orlando, Florida

Bernhard E. Riecke
Simon Fraser University
Surrey, British Columbia, Canada

Albert "Skip" Rizzo
Institute for Creative Technologies
University of Southern California
Playa Vista, California

Eduardo Salas
Department of Psychology
Institute for Simulation and Training
University of Central Florida
Orlando, Florida

Sae Schatz
DSCI MESH Solutions, LLC
Orlando, Florida

Tarah N. Schmidt-Daly
Medical Innovations Division
Design Interactive, Inc.
Oviedo, Florida

Dylan Schmorrow
Soar Technologies, Inc.
Vienna, Virginia

Jurgen P. Schulze
Qualcomm Institute
University of California, San Diego
La Jolla, California

Lee W. Sciarini
Operations Research Department
Naval Postgraduate School
Monterey, California

Ari Shapiro
Institute for Creative Technologies
University of Southern California
Playa Vista, California

William R. Sherman
Pervasive Technology Institute
Indiana University
Bloomington, Indiana

Barbara Shinn-Cunningham
Department of Biomedical Engineering
and
Center for Computational Neuroscience and Neural Technology
Boston University
Boston, Massachusetts

Brian D. Simpson
Air Force Research Laboratory
Human Effectiveness Directorate
Wright-Patterson Air Force Base, Ohio

Shawn Stafford
Department of Game Studies
Full Sail University
Orlando, Florida

Kay M. Stanney
Design Interactive, Inc.
Oviedo, Florida

Robert J. Stone
School of Electronic, Electrical and Computer Engineering
University of Birmingham
Birmingham, United Kingdom

Simon Su
Ball Aerospace & Technologies Corp.
Dayton, Ohio

Glen Surpris
Defensive Solutions Division
Design Interactive, Inc.
Oviedo, Florida

A. Murat Tekalp
College of Engineering
Koc University
Istanbul, Turkey

James Templeman
US Naval Research Laboratory
Washington, DC

Helmuth Trefftz
Virtual Reality Lab
Universidad Escuela de Administración, Finanzas y Technologia
Medellín, Colombia

Matthew Turk
Department of Computer Science
University of California, Santa Barbara
Santa Barbara, California

William B. Vessey
Enterprise Advisory Services, Inc.
Johnson Space Center
Houston, Texas

Erik Viirre
Division of Neurosciences
School of Medicine
Perlman Ambulatory Care Center
University of California, San Diego
La Jolla, California

Michael Vorländer
Institute of Technical Acoustics
RWTH Aachen University
Aachen, Germany

Xuezhong Wang
Engineering Division
Design Interactive, Inc.
Oviedo, Florida

Jingjing Wang-Costello
Yahoo, Inc.
Sunnyvale, California

John P. Wann
Department of Psychology
Royal Holloway, University of London
Surrey, United Kingdom

Christine Wasula
CSK Legal
Orlando, Florida

Philip Weber
Qualcomm Institute
University of California, San Diego
La Jolla, California

Janet M. Weisenberger
Department of Speech and Hearing Sciences
The Ohio State University
Columbus, Ohio

Robert B. Welch
Ames Research Center
National Aeronautics and Space Administration
Moffett Field, California

Alan D. White
Department of Hepatobiliary and Transplant Surgery
St James's University Hospital
Leeds, United Kingdom

Eric Whiting
Idaho National Laboratory
Idaho Falls, Idaho

Richard M. Wilkie
School of Psychology
University of Leeds
Leeds, United Kingdom

John R. Wilson
Virtual Reality Applications Research Team
University of Nottingham
Nottingham, United Kingdom

Brent Winslow
Design Interactive, Inc.
Oviedo, Florida

Michael Zyda
Department of Computer Science
USC GamePipe Laboratory
Los Angeles, California

This page intentionally left blank

Section I

Introduction

This page intentionally left blank

1 Virtual Environments in the Twenty-First Century

Kay M. Stanney, Kelly S. Hale, and Michael Zyda

CONTENTS

1.1 INTRODUCTION

> Whatever I have up till now accepted as most true I have acquired either from the senses or through the senses. But from time to time I have found that the senses deceive.
>
> ***Meditations on First Philosophy*, Descartes**

As Descartes suggested, there is no definitive truth; reality emanates from that which is present to our senses, and these senses we trust to distinguish reality from illusion can be deceived. It is this capacity to fool the sensory systems that we attempt to capitalize on when building a virtual world. More than cinema, which can evoke emotion even without relying on identification with psychological characters (Eisenstein, 1987), virtual worlds can provide a direct and egocentric perceptive of an imaginary world within which it is difficult to distinguish the virtual from the real. The goal of this experience is to *deceive* us and represent a *truth* that can educate, train, entertain, and inspire. In their ultimate form, virtual environments (VEs) immerse users in an alternate *reality* that stimulates multiple senses, providing vibrant experiences that are so veridical they fundamentally transform those exposed (e.g., via training, educating, marketing, or entertaining).

When the first edition of the VE Handbook was assembled over a decade ago, visions such as *The Matrix* (1999) IMDb, had elevated the status of VE to the level of pop iconography, and some of those associated with the technology arguably had risen to star status (e.g., Jaron Lanier [2013], who coined the term *virtual reality* and is still influencing the technology of tomorrow):

> Virtual Reality was, by the end of the 20th Century, destined to have helped computer users abandon the keyboard, mouse, joystick and computer display in favour of interfaces exploiting a wide range of natural human skills and sensory characteristics. They would be able to interact intuitively with virtual objects, virtual worlds and virtual actors whilst "immersed" within a multi-sensory, 3D computer-generated world. As is evident today, this brave new world simply did not happen.
>
> **R. Stone (2012, p. 24)**

Technology has been invested in, academic centers of excellence have been stood up, national initiatives have been launched, yet widespread adoption of VE technology has proven elusive. So why embark on the second edition of the VE Handbook? Because the promise is still alive! This promise is evidenced by the fact that the National Academy of Engineering (NAE) has identified the goal of enhancing virtual reality as one of the 14 grand challenges in need of solutions in the twenty-first century (National Academy of Engineering, 2008). As we journey forward, however, rather than being seduced by the technology, we need to take a deliberate look at what has been achieved and which hard problems still stand in the way of providing veridical virtual experiences that can transform those exposed, leading to gains in education, training, entertainment, and more. In this regard, the agenda set forth by Durlach and Mavor (1995) in their seminal National Research Council (NRC) report *Virtual Reality: Scientific and Technological Challenges* still provides an appropriate yardstick by which to judge the current level of maturity of VE technology and associated applications. This report developed a set of recommendations that, if heeded, should assist in realizing the full potential of VE technology. In the first edition, when reviewing this agenda, we found that much progress had been made with regard to improved computer generation of multimodal images and advancements in hardware technologies that support interface devices; however, there was considerable work remaining to realize improvements in the general comfort associated with donning these devices. In revisiting this agenda, it is evident that substantial further progress has been made from a technological perspective over the past decade, yet advances in the areas of psychological considerations and evaluation continue to lag behind (see Table 1.1).

1.2 TECHNOLOGY

The technology used to generate VE systems has rapidly matured over the past decade, driven primarily by the gaming industry, which desires greater realism, more natural and intuitive interaction, and enhanced usability. According to Moore's law, there has been multiple doublings of computer power in this time span and significant advances in massively parallel graphics processing brought about by the widespread adoption of graphical processing units (GPUs) (Mims, 2010) and field-programmable gate arrays (FPGAs) (Inta, Bowman, & Scott, 2012). In fact, GPUs and FPGAs have exceeded traditional Moore's law trends; whereas central processing unit (CPU) have doubled in speed every 18 months, GPUs have increased by a factor of 5 every 18 months (Geer, 2005; Luebke & Humphreys, 2007) and FPGAs promise to be faster than GPUs for some operations (Thomas, Howes, & Luk, 2009), while not for others (Cope, Cheung, Luk, & Witt, 2005). The increased speed, programmability, and parallelism brought about by these specialized circuits allow for high-performance, visually and aurally rich, interactive 3D virtual experiences. Beyond processing capability, technology advances have come in many areas, including human–machine interfaces, the hardware and software used to generate the VE, electromechanical systems used in telerobotics, as well as the communication networks that can be used to transform VE systems into shared virtual worlds.

1.2.1 Human–Machine Interface

Human–machine interfaces consist of the multimodal display devices, including visual, auditory, haptic, and olfactory displays used to present information to VE users, and multimodal user input devices used to control movement throughout a virtual world, including the mouse, touch screen, haptic gloves, cameras (to capture gestures, eye gaze, facial expression, etc.), and microphones (to capture voice; see Popescu, Burdea, & Trefftz, 2014, Chapter 17). Advances in this area have been substantial.

A virtual world aims to be as information rich to our eyes as the real world in order to create visual immersion. The state of the art in visual immersion has been driven in large part by the state of the art in real-time computer graphics, which has progressed in terms of detail from flat polygons to smooth shading to texture mapping and on to programmable shaders, the latter of which provide

TABLE 1.1
Status of Durlach and Mavor's (1995) Recommendations for Advancing VE Technology

Area	Recommendation	Status
Technology: human–machine interface	Address issues of information loss due to visual display technology shortcomings (e.g., poor resolution, limited field of view, deficiencies in tracker technology, bulkiness, ruggedness)	S
	Improvements in spatialization of sounds, especially sounds to the front of a listener and outside of the *sweet spot* surrounding a listener's head	S (see Chapter 4)
	Improvements in sound synthesis for environmental sounds	S
	Improvements in real-time sound generation	M
	Better understanding of scene analysis (e.g., temporal sequencing) in the auditory system	M
	Improvements in tactile displays that convey information through the skin	M (see Chapter 5)
	Better understanding of the mechanical properties of skin tissues that come in contact with haptic devices, limits on human kinesthetic sensing and control, and stimulus cues involved in the sensing of contact and object features	M
	Improvements in locomotion devices beyond treadmills and exercise machines	M (see Chapter 10)
	Address fit issues associated with body-based linkage tracking devices; workspace limitations associated with ground-based linkage tracking devices; accuracy, range, latency, and interference issues associated with magnetic trackers; and sensor size and cost associated with inertial trackers	M (see Chapter 7)
	Improvements in sensory, actuator, and transmission technologies for sensing object proximity, object surface properties, and applying force	S (see Chapter 5)
	Improvements in the vocabulary size, speaker independence, speech continuity, interference handling, and quality of speech production for speech communication interfaces	S
	Improvements in olfactory stimulation devices	L (see Chapter 6)
	Improvements in physiological interfaces (e.g., direct stimulation and sensing of neural systems)	M
	Address ergonomic issues associated with interaction devices (e.g., miniaturization, weight, cost, power consumption, and integration methods)	M
	Better understanding of perceptual effects of misregistration of visual images in augmented reality	M
	Better understanding of how multimodal displays influence human performance on diverse types of tasks	M (see Chapter 29)
Technology: computer generation of virtual environments	Improvements in techniques to minimize the load (i.e., polygon flow) on graphics processors	S (see Chapter 11)
	Improvements in data access speeds	S
	Development of operating systems that ensure high-priority processes (e.g., user tracking) receive priority at regular intervals and provide time-critical computing and rendering with graceful degradation	M
	Improvements in rendering photorealistic time-varying visual scenes at high frame rates (i.e., resolving the trade-off between realistic images and realistic interactivity)	S
	Development of navigation aids to prevent users from becoming lost	M (see Chapter 19)
	Improvements in the ability to develop psychological and physical models that *drive* autonomous agents	M (see Chapter 14)
	Improved means of mapping how user control actions update the visual scene	M (see Chapter 12)
	Improvements in active mapping techniques (e.g., scanning-laser range finders, light stripes)	S (see Chapter 7)

(*continued*)

TABLE 1.1 (continued)
Status of Durlach and Mavor's (1995) Recommendations for Advancing VE Technology

Area	Recommendation	Status
Technology: telerobotics	Improvements in the ability to create and maintain accurate registration between the real and virtual worlds in augmented reality applications	S
	Development of display and control systems that support distributed telerobotics	M (see Chapter 44)
	Improvements in supervisory control and predictive modeling for addressing transport delay issues	M (see Chapter 44)
Technology: networks	Development of network standards that support large-scale distributed VEs	M
	Development of an open VE network	M
	Improvements in the ability to embed hypermedia nodes into VE systems	S
	Development of wide area and local area networks with the capability (e.g., increased bandwidth, speed, reliability, reduced cost) to support the high-performance demands of multimodal VE applications	S
	Development of VE-specific application-level network protocols	M
Psychological consideration	Better understanding of sensorimotor resolution, perceptual illusions, human-information-processing transfer rates, and manual tracking ability	M
	Better understanding of the optimal form of multimodal information presentation for diverse types of tasks	M (see Chapter 13)
	Better understanding of the effect of fixed sensory transformations and distortions on human performance	M
	Better understanding of how VE drives alterations and adaptation in sensorimotor loops and how these processes are affected by magnitude of exposure	M (see Chapters 31 through 33)
	Better understanding of the cognitive and social side effects of VE interaction	M (see Chapters 16, 27)
Evaluation	Establish set of VE testing and evaluation standards	M (see Chapter 28)
	Determine how VE hardware and software can be developed in a cost-effective manner, taking into consideration engineering reliability and efficiency, as well as human perceptual and cognitive features	M (see Chapter 20)
	Identify capabilities and limitations of humans to undergo VE exposure	M (see Chapters 23, 24, 26, 31 through 33)
	Examine medical and psychological side effects of VE exposure, taking into consideration effects on human visual, auditory, and haptic systems, as well as motion sickness and physiological/psychological aftereffects	M (see Chapters 23, 24, 26, 31 through 33)
	Determine if novel aspects of human–VE interaction require new evaluation tools	M (see Chapter 28)
	Conduct studies that can lead to generalizations concerning relationships between types of tasks, task presentation modes, and human performance	L (see Chapter 29)
	Determine areas in which VE applications can lead to significant gains in experience or performance	M (see Chapters 36 through 50)

Note: L, limited to no advancement; M, modest advancement; S, substantial advancement.

unprecedented flexibility (Dionisio, Burns, & Gilbert, 2013). Beyond improved processing power derived through GPUs, improvements in visual displays have also been driven by advances in display hardware, driven today primarily by the gaming industry. In terms of visual displays, the head-mounted display (HMD), which in the past has been looked at as a potential showstopper for virtual worlds, has seen some level of resolution in recent years with regard to its technological hurdles (see Badcock, Palmisano, & May, 2014; Chapter 3). For example, a newcomer to the field targeted for the gaming industry, the Oculus Rift virtual reality headset, has a wide field of view (110°), very little visible optical housing, and low-latency head tracking (due to Oculus' 1000 Hz adjacent reality tracker, which uses a combination of three-axis gyros, accelerometers, and magnetometers), with issues still remaining in terms of resolution (though they are prioritizing improvements in this area; Shilov, 2013), motion sickness, bulkiness, and untested ruggedness (Bruce, 2013; Hester, 2013). To foster widespread adoption, Oculus is aiming for an affordable consumer price point (under $500; Newman, 2013), and with Facebook's recent acquisition of Oculus VR (Zuckerberg, 2014), visually immersive applications are positioned to expand widely. Other HMD options are available, including Carl Zeiss Cinemizer, Sensics zSight, Silicon Micro Display ST1080, and Sony HMZ-T1/HMZ-T2 (Road to VR, 2013). While gains have been made, and low-to-mid cost options are available, remaining issues with HMDs (Boger, 2013) have likely stymied user acceptance. Perhaps for this reason, wearable displays, which may be more opportune for mass adoption than persistently cumbersome headsets, are making their way into virtual reality applications, including Google's Glass (Rivington, 2013), castAR (Lee, 2014), metaglasses (Ramirez, 2013), Atheer Labs' 3D mobile computing technologies (Tweney, 2013), Vuzix's M100 smart glasses (Bohn, 2013). Lumus' optical engine (OE) displays (Wollman, 2012), and Innovega's iOptik contact lens display (Robertson, 2013). Yet, whether it is head-mounted or active goggle displays, none of these solutions create a visual display that arouses the senses in the same manner as one's natural environment. Toward that end, there are attempts to build virtual data displays directly into the physical environment through 3D mapping techniques that are still in their infancy (Jagadeesh, 2013; Lawler, 2013). These nascent technologies, which recently saw a breakthrough in terms of an optical chip based on waveguide-based platform technology (Smalley, Smithwick, Bove, Barabas, & Jolly, 2013), may someday bring holograms in motion (i.e., 3D holographic video displays that are generated by a computer on the fly) and other glass-free 3D displays (Fattal et al., 2013).

A virtual world is not created by visuals alone. The ability to produce ambient audio cues also influences immersion, providing important positional and spatial cues that contribute significantly to one's sense of placement within a VE. There have been gains in the area of virtual auditory displays (see Vorlander & Shinn-Cunningham, 2014; Chapter 4). While early spatialized audio solutions (Blauert, 1997) were expensive to implement, it is currently feasible to include spatialized audio in most VE systems. (For an excellent source on spatialized audio, see Kapralos, Jenkin, and Milios [2008].) On the hardware side, stereo speakers, surround-sound speakers, speaker arrays, and headphone-based systems, along with 16-bit stereo soundcards, can deliver near real-life binaural sound (Chang & Jacobsen, 2013; Fazi & Nelson, 2010; Poletti, Fazi, & Nelson, 2011; Schobben & Aarts, 2005; Seo, Yoo, Kyeongok Kang, & Fazi, 2010). From the software side, many options are available, including Sound Lab (SLAB3D), a software-based real-time virtual acoustic environment (VAE) rendering system (Miller, 2012); SPAT, a real-time modular spatial sound processing software system and an associated semantic and syntactic specification for storing and transmitting spatial audio scene descriptions (Peters, Lossius, & Schacher, 2013; Wozniewski, Settel, Quessy, Matthews, & Courchesne, 2012); SoundScape Renderer (SSR), a tool for real-time spatial audio reproduction (Geier & Spors, 2012); Scatter, a dictionary-based method based on sound diffusion and particle-oriented approaches, which provides an alternative to time–frequency signal representations (e.g., short-term Fourier and wavelet analyses approaches; McLeran, Roads, Sturm, & Shynk, 2008); and Auro-3D, a 3D audio technology suite that provides tools to create sound in three distinct layers—surround, height and overhead (Van Baelen, Bert, Claypool, & Sinnaeve, 2011), and stratified approaches (Peters et al., 2009). These software solutions support the design of complex spatial audio scenarios through control of a variety of signal processing functions, including sound source allocation, specification of reflections, frame-accurate callback configuration, sound output destination selection, and specification of acoustic scene and renderer parameters, as well as allowing for normal head-related transfer function (HRTF) processing (see Vorlander & Shinn-Cunningham, 2014; Chapter 4). Recently, parametric methods have been developed to capture perceptually relevant information from HRTFs, which allow for binaural rendering to be performed at lower complexity compared to conventional HRTF convolution (Breebaart, Nater, & Kohlrausch, 2009; Dal Bó Silva & Götz, 2013). The use of parametric spatial processing provides a more convincing spatial reproduction for conventional stereo signals, and the combined process of spatial decoding and HRTF parameterization reduces processing requirements and increases perceived quality. Given the typically large computational requirements associated with spatialized audio, it is no surprise that the general-purpose GPU is being used to support spatial sound generation and audio processing compiling in VEs (Hamidi & Kapralos, 2009). These approaches can support efficient implementation of complex and computationally costly software-based spatial sound algorithms, although slow data transfer between GPU and CPU remains a bottleneck even with the use of a Peripheral Component Interconnect Express (PCIe) bus (Inta et al., 2012).

Haptic display technology aims to allow those who traverse a virtual world to manipulate and interact with the virtual objects they encounter (Dindar, Tekalp, & Basdogan, 2013, Chapter 5) and some are using haptic displays to communicate via tactile languages (Fuchs, Johnston, Hale, & Axelsson, 2008). The realization of such naturalistic physical interaction is still largely an area of research found in laboratories. Haptic research and development to date has generally been application specific, with many applications focused on the field of medicine (Demain, Metcalf, Merrett, Zheng, & Cunningham, 2013; Kawasaki, Endo, Mouri, Ishigure, & Daniulaitis, 2013). Some work aims to provide means of integrating haptics into existing platforms, such as *Second Life* (de Pascale, Mulatto, & Prattichizzo, 2008). Methods for capturing, modeling, and reproducing touch-related object properties have also been an area of research. To create virtual interaction, haptic interfaces can be used to measure the motion of the human hand, map these movements onto the VE, and trigger haptic device actuators to provide users with appropriate force and vibration feedback, thereby converting virtual contacts into physical interactions. Recently, this mapping has moved from the use of surface models that require extensive and subjective hand tuning via simple parametric relationships to *haptography*, which bases tactile models on haptic data recorded via a handheld stylus that is used to capture haptic properties of items in the real world (Kuchenbecker, 2008). From these data, a *haptograph* is produced, which is a haptic impression of an object or surface patch, including various haptic properties (e.g., shape, stiffness, friction, and texture). Using the haptograph, the feel of an object or surface can then be recreated via a haptic interface. Haptography aims to provide a generalizable approach to designing authentic virtual touch—much as a camera lens provides a means of reproducing the visual world—through the specification of haptic sensations (i.e., determining how to mimic the human's afferent nervous systems), understanding of haptic distillation (i.e., developing algorithms that can automatically reconstruct these haptic impressions), and enhanced haptic rendering (i.e., developing haptic devices that increase the bandwidth of current amplifiers and mechanical linkages). In terms of the latter, considerable research has focused on the design of haptic devices (Bullion & Gurocak 2009; Folgheraiter, Gini, & Vercesi, 2008; Withana et al., 2010). This has evolved into various device options, including electrical tactile systems, vibromechanical systems (both electromechanical [e.g., rotary inertial, linear actuators] and pneumatic tactors), static low-frequency tactors (e.g., pin-based tactile displays, hydraulic), piezoelectric-based devices, and burgeoning approaches such as electroactive polymers and microelectromechanical systems (McGrath et al., 2008). Initially, linear actuators were the most commonly used; however, rotary inertial tactors have become more popular in recent years. Each solution has strengths and limitations and thus the best option is application dependent. Before such technology is embraced by the masses, advances are needed in many areas including miniaturization, weight, cost, power consumption, and integration methods (i.e., means of incorporating tactors into clothing, seats, harnesses, etc.).

Until recently, most virtual worlds did not include stimulation of senses beyond the big three—visual, auditory, and haptic. Advances in olfactory displays and interfaces are making the inclusion of virtual odors a possibility (Nakamoto, 2013; also see Jones, Dechmerowski, Oden, Lugo, Wang-Costello, & Pike, 2014; Chapter 6). Currently, similar to haptic displays, olfactory display techniques have not moved much beyond the research laboratory. There are various scent generation methods and scent delivery methods being explored. In terms of scent generation methods, advances are needed in the areas of vaporization/atomization techniques, scent switching techniques, and formulation techniques (Yanagida & Tomono, 2013). There are various scent delivery methods being explored, including odor releasing vents, aroma chips using functional polymer gels, inkjet printer trigger mechanisms, and projection-based systems (Kim, 2013; Matsukura, Yoneda, & Ishida, 2012; Nakaizumi, Noma, Hosaka, & Yanagida, 2006; Sugimoto & Okada, 2013). Perhaps one day, not too far away, it will be possible to travel through a virtual world with tantalizing smells wafting toward us from various virtual objects.

Since the initial edition of this handbook, touch-based input device technology has expanded in the commercial marketplace beyond conventional control interfaces (e.g., keyboard, mouse, game controller) to realize more natural interaction techniques (e.g., via touch screens, haptic devices, motion tracking, gestures) (Turk, 2014, Chapter 9). The widespread use and adoption of the Nintendo Wii®, released in 2006, and the iPhone®, released in 2007, transformed user expectations regarding acceleration and touch interfaces. Subsequently, gesture (e.g., Microsoft Kinect®, Leap Motion®) and integrated gesture and natural language interfaces (e.g., Creative Senz3D™ Interactive Gesture Camera) have provided *touchless* interactions that are "disrupt[ing] user experience, from the heuristics that guide us, to our design patterns and deliverables" (Pagan, 2012, p. 1). The primary sensing methods for capturing gestural interaction include movement based (e.g., touch interfaces, glove based, and acceleration based) and vision based (e.g., camera systems) (Dardas & Alhaj, 2011; Rautaray & Agrawal, 2012). The gaming industry and interaction researchers have developed gestural libraries for specific needs (e.g., in operating rooms; Ruppert, Reis, Amorim, de Moraes, and Da Silva [2012]) and gesture ontologies for individual users within VEs (Lanier et al., 2013) and multiuser applications (Roman, Lazarov, & Majumder, 2009). There remains a need for gesture-based interface standards, which minimize instances where the same gesture is used to represent differing meanings/actions across platforms or applications, which should, in turn, facilitate learning while minimizing user frustration and errors.

Advances in motion tracking are needed to extend gestural interaction to full-body motion, which would allow virtual images to be calibrated to the head and/or body position of the individual traversing a virtual world. There are several possible approaches to motion tracking including optical, mechanic, magnetic, and inertial. The most popular approach seems to be *outside-in* optical tracking systems. Advances in tracking technology have been realized in terms of 9DOF microelectronic mechanical system (MEMS) sensors, inertial measurement unit (IMU)-enabled GPS devices, emitter tower constellations, and independent acoustic source positioning (Park, 2013; Sun, Ma, Han, Ross, & Wee, 2013; Zou, 2013). The future of tracking technology continues to trend toward hybrid tracking systems, with a hybrid optical–inertial approach recently developed that uses the inside-out concept of the InterSense VisTracker technology (acquired by Thales Visionix), coupled with a tiny high-performance NavChip IMU (Atac & Foxlin, 2013) and several others in research labs (e.g., magnetic inertial, optical magnetic, radio frequency inertial). In addition, ultrawideband radio technology holds promise for an improved method of omnidirectional point-to-point ranging (Venkatesh & Buehrer, 2007).

If those who traverse and interact in virtual worlds are to become truly immersed, they will need to have the ability to seamlessly converse with the virtual agents they encounter. The quality of speech recognition (natural language processing) and synthesis systems is thus of paramount importance. Fluent speech-to-speech applications, fueled by real-time, speaker-independent, automatic speech recognition software and systems, are coming closer to reality (Aggarwal & Dave, 2012; Picheny et al., 2011; Seide, Li, & Yu, 2011). The accuracy of such systems has been substantially improved through recent advances in graphical model–based machine-learning techniques, with computationally tractable training algorithms, and advanced neural-network modeling techniques (e.g., context-dependent deep-neural-network hidden Markov models). Once issues associated with acoustic and language modeling algorithms are fully resolved, the possibility to develop speech recognition systems that can read voice intonation and integrate with body language and facial expression recognition systems so that emotion and intent can be better understood will likely become the holy grail.

Taken together, these display and user input technological advancements, along with those poised for the near future, provide the infrastructure on which to build complex, immersive multimodal VE applications.

1.2.2 Computer Generation of Virtual Environments

So what exactly is a virtual world? Gilbert (2011) identified five essential characteristics of contemporary state-of-the-art virtual worlds, including the following:

1. A 3D graphical interface and integrated audio (not text based)
2. Immersion derived through spatial, environmental, and multisensory realism that is capable of producing a sensation of presence
3. User-generated activities and goals (not prescripted) with the ability to create content to personalize the VE experience
4. A persistent world that continues to exist even after a user exits it
5. Simultaneous, massively multiuser remote and distributed interactivity

Thus, several elements are involved when generating VEs, including graphics and audio generators, software for effective and flexible content generation, and networks that support online environments. In terms of graphics and audio generators, computer generation of VEs requires very large physical memories, high-speed processors, high-bandwidth mass storage capacity, and high-speed interface ports for input/output devices (Durlach & Mavor, 1995). Remarkable advances in hardware technologies have been realized in the past decade that support generation of simultaneous, massively multiuser virtual worlds. While Moore's law may be nearing its end (theoretical physicist Michio Kaku posits that Moore's law has about 10 years of life remaining; Paul, 2013), the semiconductor industry is working feverishly to solve this problem by introducing trigate (3D) transistor technology, heterogeneous system architectures (HSAs) that use parallel computing schemes (e.g., accelerated processing units [APUs] that couple CPUs with GPUs, FPGAs, or other such specialized processing systems), and a possible future of other-than-silicon-based transistors (e.g., gallium arsenide processors, molecular transistors, quantum computing) sometime between 2018 and 2026 (Chacos, 2013; ITRS, 2011). When using parallel computing schemes, GPU path tracing can be accelerated by an order of magnitude compared to the CPU, which in turn supports near real-time frame rates that enable rendering photorealistic time-varying visual scenes (Rahikainen, 2013). In addition, the first dedicated ray tracing hardware—*Caustic Series2 ray tracing acceleration boards*—has been released, although its current use is not intended for real-time execution but rather for accelerating lighting, *look development*, and design visualization workflows (Imagination, 2013). One caveat: While such advances needed to support real-time generation of virtual worlds will keep making substantial gains, power and resulting heat may become limiting factors. As the NRC has noted, "even as multicore hardware systems are tailored to support software that can exploit multiple computation units, thermal constraints will continue to be a primary concern" (Fuller & Millett, 2011, p. viii).

From a software perspective, over the past decade, it has become possible to rapidly build and render complex virtual worlds. Virtual reality software toolkits (e.g., Vizard, Goblin XNA, Demotride); server platforms (e.g., Open Cobalt, Open Wonderland, OpenSimulator, Solipsis); standard application program interfaces (APIs) (e.g., OpenGL, Direct-3D); cross platform scene graph-based 3D APIs (e.g., Java-3D, H3DAPI, Ardor3D, jMonkeyEngine, Espresso3D, Jreality); 3D modeling languages, toolkits, content generators, and photorealistic rendering tools (e.g., AC3D Modeler, Autodesk 3ds Max, Cobalt, CyberX3D, Java 3D VRML Loader, LightWave 3D, Maya, MMDAgent, Modo, Photosynth, Presagis Creator, Raster3D, Remo 3D, SimVRML, VRML, X3D ToolKit); and real-time image processing libraries (e.g., OpenCV, VXL, IVT) are all accelerating the development process. Using these tools, commercial application developers can build a range of VEs, from the most basic mazes to complex medical simulators and from low-end single-user PC platform applications to massively online collaborative applications supported by client–server environments. Further, today's VE software toolkits support integration with most VE hardware (e.g., HMDs, goggles, 3D projection systems, motion trackers, 3D sound systems, haptic interfaces, datagloves), readily import 3D models and sounds, and provide the ability to build VE applications as executables, and with scripting languages (e.g., Python, OpenSim, Linden, GAML), it is now possible for even nonprogrammers to develop virtual worlds. One aspect of VE content creation that is lagging behind is advances in agent artificial intelligence (AI), especially with regard to techniques to develop emotionally responsive agents and other nonplayer entities (NPEs) (Slater, Moreton, Buckley, & Bridges, 2008). There are tools available to develop *visually* lifelike human avatars (e.g., *CoJACK*, *DI-Guy*; see Feng, Shapiro, Lhommet, & Marsella, 2014; Chapter 14); however, their expressive behavior comes nowhere close to their photorealistic visual fidelity. There have been considerable advances with regard to integration of telerobotic techniques (see Kheddar, Chellali, & Coiffet, 2014; Chapter 43) into autonomous agent design, which is leading to a new breed of social interaction (e.g., virtual humanoid *U-Tsu-Shi-O-Mi*, *Jeeves*, and *MiRA*) that is supported by a combination of robotic and virtual components (Holz, Dragone, & O'Hare, 2009). Interactions with the Kinect®, Wii, and other more naturalistic interfaces are further breaking down the barriers typically found in third person representations of human interactions. The next step may be integrating electroencephalography (EEG) and other neurophysiological measures of human emotional state and rendering these onto virtual agents (Kokini et al., 2012).

When moving about a VE, research has shown that travelers plan ahead, using a spatial map of the environment (Tyson, 2013). Thus, to enable user-generated wayfinding within virtual worlds, navigational techniques should foster the development of spatial maps. Several such techniques have been investigated, including maps, landmarks, trails, and direction finding (see Darken & Peterson, 2014; Chapter 19). Many of these techniques have been perfected by the game industry.

Some informal guidelines to support traversing virtual worlds have evolved. For closed VEs (e.g., buildings), tools that demonstrate the surrounding area (maps, exocentric 3D views) are recommended if training or exposure time is short, while internal landmarks (i.e., along a route) are recommended for longer exposure durations (Stanney, Chen, & Wedell, 2000). For semiopen (e.g., urban areas) and open environments (e.g., sea, sky), demonstrating the surround is appropriate for short exposures, while the use of external landmarks (i.e., outside a route) is recommended for long exposure times. For navigation to far away virtual places, image plane interaction, scaled-world grabbing, steering-based multiscale navigation, target-based multiscale navigation, and World in Miniature (WIM) techniques may prove effective (Kopper, Ni, Bowman, & Pinho, 2006). In addition, Microsoft Kinect–based techniques, such as footpad and multitouch pad controllers, offer an alternative to purely virtual navigation aids (Dam, Braz, & Raposo, 2013). Brain-controlled navigational interfaces provide another alternative to virtual aids, where an individual navigates a virtual world using only their cerebral activity (Lécuyer et al., 2008; Vourvopoulos, Liarokapis, & Petridis, 2012). Any such VE navigational techniques should carefully consider the fact that spatial orientation and navigation in natural environments rely heavily on locomotion and associated activation of motor, vestibular, and proprioceptive systems; thus, the impact of the absence of these motion-based cues in VEs is important to consider when evaluating the efficacy of navigational aids (Taube, Valerio, & Yoder, 2013). More work is needed in the area of navigational aiding because becoming lost or disoriented in a virtual world has been found to be one of the most common usability issues experienced (Sawyerr, Brown, & Hobbs, 2013).

The NRC report (Durlach & Mavor, 1995) indicated the need for a real-time operating system (RTOS) for VEs; however, while it used to be the case that supporting VEs required an RTOS, this no longer holds true. Due to the aforementioned multiple doublings in processing power, the corresponding need for an RTOS has been obviated by the utilization of a common OS (e.g., Windows, Linux), which can be used as a proxy for a realistic VE. This increase in capability of the computer OS is in part due to the increase in popularity of computer gaming, which requires significant processing power in both the processor (to handle relevant computations) and the GPU (to handle rendering of complex VEs). The combination of these technologies has led to the popularity of massively multiplayer online role-playing games (MMORPGs), such as *World of Warcraft*, *Rift*, and *Lord of the Rings* online. These MMORPGs are, for all intents and purposes, the latest incarnation of a VE in that they share the concept of a shared virtual world, a representation of the player as an avatar, and ways to mitigate latency and jitter that were commonplace in older OSs. These have become the de facto standard in deploying this type of software.

1.2.3 Telerobotics

Beyond the advantages to autonomous agent design discussed earlier, there are many areas (e.g., sensing, navigation, object manipulation) in which VE technology can prosper from the application of robotic techniques. Yet, if these techniques are to be adopted, issues of stability and communication time delay (i.e., transport delay), link flexibility, and real-time control architecture design (e.g., object recognition and pose estimation, fusion of vision, tactile, and force control for manipulation) must be resolved (Ambrose et al., 2012; Atashzar, Shahbazi, Talebi, & Patel, 2012). Chapter 44 discusses a number of techniques for addressing these issues.

1.2.4 Networks

The NRC report (Durlach & Mavor, 1995) suggested that with improvements in communications networks, VEs would become shared experiences in which individuals, objects, processes, and autonomous agents from diverse locations interactively collaborate. This has occurred with a number of virtual worlds currently actively populated with online communities (e.g., *Call of Duty*, *Active World*, *Second Life*, *SimCity*, *World of Warcraft*). As these virtual communities grow and online demand increases in general (e.g., crowdsourcing), shortcomings with regard to performance, reliability, scalability, and security are being explored through Future Internet research efforts, which are ongoing around the world (e.g., the United States, Global Environment for Network Innovations [GENI] and Future Internet Network Design [FIND]; Korea, Future of the Internet for Korea [u-IT839]; European Union, FP6 and FP7, Network of Excellence [Euro-NGI, Euro-FGI], and EIFFEL; Japan, Collaborative Overlay Research Environment [CORE]; and Germany, IKT 2020). These efforts are working toward advances in networks (e.g., application-centric multinetwork service and edge-based intelligence [i.e., boundary between providers and users is disappearing; Peer-to-Peer (P2P) content delivery networks]) and service evolution (e.g., overlay and self-organizing networks, changes in user traffic behavior, functional versus stochastic scalability [e.g., Voice over Internet Protocol (VoIP) signaling platform with an overlay network]). Significant challenges remain with regard to transitioning the focus from quality of service or quality of experience and charting the evolutionary path to the Future Internet (e.g., PlanetLab test bed support of GENI) (Tran-Gia, 2007). In addition, commercial successes such as Google Fiber, which aims to provide 1 Gbps networking to residential clients, mean that networked VE applications will become more ubiquitous as time goes on.

1.3 PSYCHOLOGICAL CONSIDERATION

There are a number of psychological considerations associated with the design and use of VE systems. Some of these considerations focus on techniques and concerns that can be used to augment or enhance VE interaction and transfer of training (e.g., perceptual illusions, design based on human-information-processing transfer rates), while others focus on adverse effects due to VE exposure. In terms of the former, we know that perceptual illusions exist, such as auditory–visual cross-modal perception phenomena, yet little is known about how to leverage these phenomena to reduce development costs while enhancing one's experience in a VE. Perhaps, one exception is vection (i.e., the illusion of self-movement), which is known to be related to a number of display factors (see Table 18.1 of Hettinger, 2014; Chapter 18). By manipulating these display factors, designers can provide VE users with a compelling illusion of self-motion throughout a virtual world, thereby enhancing their sense of presence (see Chertoff & Schatz, 2014; Chapter 34) often with the untoward effect of motion sickness (see Keshavarz, Heckt, & Lawson, 2014; Lawson, 2014a, 2014b; Chapters 23, 24, and 26). Other such illusions exist (e.g., visual dominance) and could likewise be leveraged. While current knowledge of how such perceptual illusions occur is limited, it may be sufficient to know that they do occur in order to leverage them to enhance VE system design and reduce development costs. Substantially more research is needed in this area to identify perceptual and cognitive design principles (see Munro, Carroll, Sheldon, & Patrey, 2014; Chapter 16) that can be used to trigger and capitalize on these illusory phenomena.

Another psychological area in need of research is that of transfer of training (see Champney, Carroll, Surpris, & Cohn, 2014; Chapter 30). Stanney, Mourant, and Kennedy (1998, p. 330) suggest that

> To justify the use of VE technology for a given task, when compared to alternative approaches, the use of a VE should improve task performance when transferred to the real-world task because the VE system capitalizes on a fundamental and distinctively human sensory, perceptual, information processing, or cognitive capability.

But what leads to such transfer? VEs provide the ability to reconstruct similar conditions to those in the operational world, which would otherwise be too risky, costly, or cumbersome to reproduce, which imparts them with high face validity. While face validity is important, there are still limited data-grounded best practices that can be used to direct the design of VE training solutions such that they optimize skill acquisition and retention (Burke & Hutchins, 2008). Despite this lack of knowledge, transfer of training from VEs to real-world tasks has been demonstrated across a range of applications from simple sensorimotor tasks (Kenyon & Afenya, 1995 [see refuting evidence in Kozak, Hancock, Arthur, & Chrysler, 1993]; Rose et al., 1998; Rose, Brooks, & Attree, 2002) and procedural tasks to ones that are more complex, both procedurally (Brooks, Rose, Attree, & Elliot-Square, 2002) as well as cognitively and spatially (Foreman, Stanton, Wilson, & Duffy, 2003). While these examples demonstrate the potential of achieving training transfer with VE systems, there is a need for better understanding of the types of tasks or activities for which the unique characteristics of VEs (i.e., egocentric perspective, stereoscopic 3D visualization, real-time interactivity, immersion, and multisensory feedback) can be leveraged to provide significant gains in human performance, knowledge, or experience.

In contrast to the limited knowledge concerning perceptual and cognitive design principles that augment or enhance VE interaction, more is known about identifying and controlling the adverse effects of VE exposure. Adverse effects are of particular concern because they can persist for some time after exposure, potentially predisposing those exposed to harm. These effects are both physiological (see DiZio & Lackner, 2014; Keshavarz, Heckt, & Lawson, 2014; Lawson, 2014a, 2014b; Stanney, Kennedy, & Hale, 2014; Wann & Mon-Williams, 2014; Chapters 23, 24, 26, 31 through 33) and psychological (see Calvert, 2014; Chapter 27), with considerable effort currently focused on the former and less on the latter. With regard to the latter, some are concerned that exposure to VEs that portray violent content, as is often found in entertainment venues, could lead to aggressive, antisocial, or criminal behavior (see Calvert, 2014; Chapter 27). Thus, a proactive approach is needed that weighs the physiological and psychological risks and potential consequences associated with VE exposure against the benefits. Waiting for the onset of harmful consequences should not be tolerated.

Taken together, the research into psychological considerations of VE exposure indicates that more research is needed to derive perceptual and cognitive design strategies that enhance VE interaction and that there are risks associated with VE exposure. However, usage protocols have been developed that, if successfully adopted, can assist in minimizing these risks (see Stanney, Kennedy, & Hale, 2014; Chapter 31). Thus, VE technology is not something to be eschewed as it has many advantages for enticing and didactic experiences; it is rather something to leverage wholly yet vigilantly, taking care to address the associated risks.

1.4 EVALUATION

Most VE user interfaces are fundamentally different from traditional graphical user interfaces, requiring that their design address unique input/output (I/O) devices, perspectives, and physiological interactions (see McMahan, Lopper, & Bowman, 2014; Chapter 12). Thus, when developers and usability practitioners attempt to apply traditional usability engineering methods to the evaluation of VE systems, they find few if any that are particularly well suited to these environments (see Gabbard, 2014; Chapter 28). There is a need to address key characteristics unique to VEs (e.g., perceived presence and real-world fidelity, multidimensional interactivity, immersion, spatial navigation, orientation), for which existing usability methods fall short in their ability to assess VE systems (see Table 1.2). Several have sought to fill this gap.

Stuart (1996) provided basic methods for evaluating general usability components of VEs. Salzman, Dede, and Loftin (1995) developed formative usability evaluation methods for assessing virtual worlds. Bowman, Koller, and Hughes (1998) developed usability techniques specific to the evaluation of various VE travel techniques. Gabbard and Hix (2000; see Gabbard, 2014; Chapter 28) developed a taxonomy of VE usability characteristics that can serve as a foundation for identifying usability criteria that existing evaluation techniques fail to fully characterize. Stanney, Mollaghasemi, and Reeves (2000) used this taxonomy as the foundation on which to develop an automated system, the Multicriteria Assessment of Usability for Virtual Environments (*MAUVE*), which organizes VE usability characteristics into 2 primary usability attributes (VE system usability and VE user considerations), 4 secondary attributes (interaction, multimodal system output, engagement, and side effects), and 11 tertiary attributes (navigation, user movement, object selection and manipulation, visual output, auditory output, haptic output, presence, immersion, comfort, sickness, and aftereffects). Similar to the manner in which traditional heuristic evaluations are conducted, *MAUVE* can be used at various stages in the usability engineering life cycle, from initial storyboard design to final evaluation and testing. It can also be used to compare system design alternatives. The results of a *MAUVE* evaluation not only identify a system's problematic usability components and techniques but also indicate why they are problematic. Such results may be used to remedy critical usability problems as well as to enhance the design for usability of subsequent system development efforts. Sawyerr et al. (2013) developed a hybrid evaluation method that combines a three-cycle (task action, navigation, and system initiative) cognitive walkthrough method and VE-specific usability

TABLE 1.2
Limitations of Traditional Usability Methods for Assessing Virtual Environments

Traditional measurement techniques only capture point-and-click interactions, which are not representative of the multidimensional object selection and manipulation characteristics of 3D space.
Quality of multimodal system output (e.g., visual, auditory, haptic) is not comprehensively addressed by traditional evaluation techniques.
Means of assessing sense of presence and aftereffects have not been incorporated into traditional usability methods.
Traditional performance measurements (i.e., time and accuracy) do not comprehensively characterize VE system interaction (e.g., spatial navigation, orientation).
Traditional single-user task-based assessment methods do not consider VE system characteristics in which two or more users interact in the same environment.

heuristics organized into three categories (i.e., design and esthetics, control and navigation, and errors and help) (Munoz, Barcelos, & Chalegre, 2011; Rusu et al., 2011).

Beyond usability, the cost-effectiveness of VE systems should also be evaluated (see Gross, 2014; Chapter 20), as well as the potential for any product liability concerns (see Kennedy, Kennedy, Kennedy, Wasula, & Bartlett, 2014; Chapter 21). With these aspects considered, developers can evaluate if VE technology offers financial advantages as well as acceptable risks over current practices or technologies. This is an essential determination if VE technology is to thrive both commercially and in research domains, which is looking possible as applications have grown in both areas (see Chapters 36 through 50).

1.5 CONCLUSION

With this second coming of virtual reality, we are more hardware and software capable than ever before. The biggest difference this time, though, is that it is the game industry that is getting us the better graphics hardware and low-cost HMDs, as opposed to the mid-1990s Department of Defense (DoD) VR agenda. The 1997 NRC study entitled *Modeling and Simulation—Linking Entertainment and Defense* (Zyda & Sheehan, 1997) predicted this trend. The largest networked VEs today are coming straight out of the game industry—the *Call of Duty* and *World of Warcraft* series of games have networked infrastructures better architected than anything ever conceived or built by the US DoD; DoD may thus be ceding their leadership in the hardware and software base underlying this second coming of VR. These games definitely move toward the ultimate form of VEs, where users are immersed in an alternate *reality* that stimulates at least the visual and aural senses and captures them effectively enough to garner hours and hours of their attention. Imagine when the same powers of immersion are brought to the classroom! As olfaction and gesturing are added, the future is destined to bring a *Matrix-like* experience to the virtual world. Thus, while the VEs to date have yet to meet with expectations, we anticipate the second go around will fulfill many of the early fantasies.

Some of the things we will be looking for as this new VR push goes forward are perhaps solving some of the harder issues with respect to populating our VR worlds—how do we create AI characters or other NPEs that both perceive and display cognition and emotions? We imagine not just a visual display, say from an Oculus Rift HMD, but also a low-cost, hybrid EEG device that reads our emotional state and transmits that to the core virtual world such that our emotionally cognoscent AI characters can appropriately interact with us more subtlety than is currently achieved through the firing of a virtual weapon. The real question then is: who will fund this careful and hard melding of cognition and emotion such that we can achieve these great, next-generation VEs? Our guess is that the entertainment industry will recognize the value in building AI characters that perceive and display emotions as they turn toward generating interactive VR stories imbued with the emotional subtleties currently only found in the cinema.

So, there are lots of great VR research works ahead with this game industry–driven second coming, and it is going to move us toward Descartes' vision—virtual sensory experiences that are so deceiving we take them as truths from which we may have little desire to withdraw.

REFERENCES

Aggarwal, R. K., & Dave, M. (2012). Recent trends in speech recognition systems. In U. S. Tiwary & T. J. Siddiqui (Eds.), *Speech, image, and language processing for human computer interaction: Multi-modal advancements* (pp. 101–127). Hershey, PA: IGI Global.

Ambrose, R., Wilcox, B., Reed, B., Matthies, L., Lavery, D., & Korsmeyer, D. (2012). *Robotics, tele-robotics and autonomous systems roadmap: Technology area 04* [Online]. Available: http://www.nasa.gov/pdf/501622main_TA04-ID_rev6b_NRC_wTASR.pdf.

Atac, R., & Foxlin, E. (2013, May 16). Scorpion hybrid optical-based inertial tracker (HObIT). *Proceedings of SPIE, Head- and Helmet-Mounted Displays XVIII: Design and Applications* (Vol. 8735), 873502, Baltimore, MD. doi:10.1117/12.2012194.

Atashzar, S. F., Shahbazi, M., Talebi, H. A., & Patel, R. V. (2012, October 7–12). Control of time-delayed telerobotic systems with flexible-link slave manipulators. In *International Conference on Intelligent Robots and Systems* (*IEEE/RSJ*) (pp. 3035–3039), Vilamoura, Algarve, Portugal.

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: Design, implementation, and applications* (2nd ed., pp. 39–86). Boca Raton, FL: Taylor & Francis Group, Inc.

Blauert, J. (1997). *Spatial hearing* (Rev. ed.). Cambridge, MA: MIT Press.

Boger, Y. (2013). *Why HMDs sometimes fail (and what to do about it): The 2013 survey on HMD reliability*. Columbia, MD: Sensics, Inc. [Online]. Available: http://sensics.com/wp-content/uploads/2012/03/why-do-hmds-fail-report.pdf.

Bohn, D. (2013, January 6). Vuzix M100 smart glasses: Hands on with the contender trying to beat Google Glass to market. *The Verge* [Online]. Available: http://www.theverge.com/2013/1/6/3843760/vuzix-m100-smart-glasses-hands-on-beating-project-glass.

Bowman, D., Koller, D., & Hughes, L. (1998). A methodology for the evaluation of travel techniques for immersive virtual environments. *Virtual Reality: Journal of the Virtual Reality Society*, *3*, 120–131.

Breebaart, J., Nater, F., & Kohlrausch, A. (2009). Parametric binaural synthesis: Background, applications and standards. In M. Boone (Ed.), *Proceedings of the NAG-DAGA International Conference on Acoustics 2009* (pp. 172–175). Rotterdam, the Netherlands: Curran Associates, Inc.

Brooks, B. M., Rose, F. D., Attree, E. A., & Elliot-Square, A. (2002). An evaluation of the efficacy of training people with learning disabilities in a virtual environment. *Disability and Rehabilitation*, *24*(11–12), 622–626.

Bruce, J. (2013, May 16). Oculus Rift development kit review and giveaway. *MakeUseOf* [Online]. Available: http://www.makeuseof.com/tag/oculus-rift-development-kit-review-and-giveaway/.

Bullion, C., & Gurocak, H. (2009). Haptic glove with MR brakes for distributed finger force feedback. *Presence: Teleoperators and Virtual Environments*, *18*, 421–433.

Burke, L. A., & Hutchins, H. M. (2008). A study of best practices in training transfer and proposed model of transfer. *Human Resource Development Quarterly*, *19*(2), 107–128.

Calvert, S. L. (2014). Social impact of virtual environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 699–718). Boca Raton, FL: Taylor & Francis Group, Inc.

Chacos, B. (2013, April 11). Breaking Moore's Law: How chipmakers are pushing PCs to blistering new levels. *PC World* [Online]. Available: http://www.pcworld.com/article/2033671/breaking-moores-law-how-chipmakers-are-pushing-pcs-to-blistering-new-levels.html.

Champney, R. K., Carroll, M., Surpris, G., & Cohn, J. V. (2014). Conducting training transfer studies in virtual environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 781–796). Boca Raton, FL: Taylor & Francis Group, Inc.

Chang, J.-H., & Jacobsen, F. (2013). Experimental validation of sound field control with a circular double-layer array of loudspeakers. *Journal of Acoustical Society of America*, *133*(4), 2046–2054.

Chertoff, D., & Schatz, S. (2014). Beyond presence. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 855–870). Boca Raton, FL: Taylor & Francis Group, Inc.

Cope, B., Cheung, P. Y. K., Luk, W., & Witt, S. (2005). Have GPUs made FPGAs redundant in the field of video processing? In G. J. Brebner, S. Chakraborty, & W.-F. Wong (Eds.), *Proceedings of the IEEE International Conference on Field Programmable Technology* (Vol. 1, pp. 111–118). FPT 2005, December 11–14, 2005, Singapore. IEEE 2005 ISBN 0-7803-9407-0.

Dal Bó Silva, B., & Götz, M. (2013). A structural parametric binaural 3D sound implementation using open hardware. *IFIP Advances in Information and Communication Technology*, *403*, 306–317.

Dam, P., Braz, P., & Raposo, A. B. (2013, July 21–26). A study of navigation and selection techniques in virtual environments using Microsoft Kinect®. In *Proceedings of the 15th International Conference on Human-Computer Interaction – HCI International 2013*, Las Vegas, NV.

Dardas, N. H., & Alhaj, M. (2011). Hand gesture interaction with a 3D virtual environment. *The Research Bulletin of Jordan ACM*, *2*(3), 186–194. Available: http://ijj.acm.org/volumes/volume2/no3/ijjvol2no3p9.pdf, viewed October 4, 2013.

Darken, R. P., & Peterson, B. (2014). Spatial orientation, wayfinding, and representation. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 467–492). Boca Raton, FL: Taylor & Francis Group, Inc.

Demain, S., Metcalf, C. D., Merrett, G. V., Zheng, D., & Cunningham, S. (2013). A narrative review on haptic devices: Relating the physiology and psychophysical properties of the hand to devices for rehabilitation in central nervous system disorders. *Disability and Rehabilitation: Assistive Technology*, *8*(3), 181–189.

de Pascale, M., Mulatto, S., & Prattichizzo, D. (2008). Bringing haptics to Second Life. In R. Whitaker & B. Liang (Eds.), *Proceedings of the Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems* (*HAS'08*) (pp. 6:1–6:6). Brussels, Belgium: Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering (ICST).

Dindar, N., Tekalp, A. M., & Basdogan, C. (2013). Dynamic haptic interaction with video. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 115–130). Boca Raton, FL: Taylor & Francis Group, Inc.

Dionisio, J. D. N., Burns, W. G., & Gilbert, R. (2013). 3D virtual worlds and the metaverse: Current status and future possibilities. *ACM Computing Surveys*, *45*(3), Article 34. doi:http://dx.doi.org/10.1145/2480741.2480751.

DiZio, P., Lackner, R., & Champney, R. K. (2014). Proprioceptive adaptation and aftereffects. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 833–854). Boca Raton, FL: Taylor & Francis Group, Inc.

Durlach, B. N. I., & Mavor, A. S. (1995). *Virtual reality: Scientific and technological challenges*. Washington, DC: National Academy Press.

Eisenstein, S. (1987). *Non-indifferent nature: Film and the structure of things* (Trans. Herbert Marshall). Cambridge, MA: Cambridge University Press.

Fattal, D., Peng, Z., Tran, T., Vo, S., Fiorentino, M., Brug, J., & Beausoleil, R. G. (2013). A multi-directional backlight for a wide-angle, glasses-free three-dimensional display. *Nature*, *495*(7441), 348–351.

Fazi, F. M., & Nelson, P. A. (2010, May 26–28). Sound field reproduction with a loudspeaker array. In *Proceedings of the 37th Convegno Nazionale Associazione Italiana Acustica* (p. 4), Siracusa, Italy.

Feng, A., Shapiro, A., Lhommet, M., & Marsella, S. (2014). Embodied autonomous agents. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 335–350). Boca Raton, FL: Taylor & Francis Group, Inc.

Folgheraiter, M., Gini, G., & Vercesi, D. (2008). A multi-modal haptic interface for virtual reality and robotics. *Journal of Intelligent Robotics Systems*, *52*, 465–488.

Foreman, N., Stanton, D., Wilson, P., & Duffy, H. (2003). Spatial knowledge of a real school environment acquired from virtual or physical models by able-bodied children and children with physical disabilities. *Journal of Experimental Psychology: Applied*, *9*(2), 67–74.

Fuchs, S., Johnston, M., Hale, K. S., & Axelsson, P. (2008). Results from pilot testing a system for tactile reception of advanced patterns (STRAP). *Proceedings of the HFES 52nd Annual Meeting*, New York, September 2008, pp. 1302–1306.

Fuller, S. H., & Millet, L. I. (2011). *The future of computing performance: Game over or next level?* Washington, DC: The National Academies Press.

Gabbard, J. L., & Hix, D. (2000, July 31). *A taxonomy of usability characteristics in virtual environments* [Online]. Available: http://csgrad.cs.vt.edu/~jgabbard/ve/taxonomy/.

Gabbard, J. L. (2014). Usability engineering of virtual environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 721–748). Boca Raton, FL: Taylor & Francis Group, Inc.

Geer, D. (2005). Taking the graphics processor beyond graphics. *IEEE Computer*, *38*(9), 14–16.

Geier, M., & Spors, S. (2012, November 22–25). Spatial audio with the SoundScape Renderer. In *Proceedings of the 27th Tonmeistertagung – VDT International Convention*, Cologne, Germany.

Gilbert, R. L. (2011). The P.R.O.S.E. Project: A program of in-world behavioral research on the Metaverse. *Journal of Virtual Worlds Research, 4*(1), 3–18.

Gross, D. (2014). Technology management and user acceptance of virtual environment technology. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 493–504). Boca Raton, FL: Taylor & Francis Group, Inc.

Hamidi, F., & Kapralos, B. (2009). A review of spatial sound for virtual environments and games with graphics processing units. *The Open Virtual Reality Journal*, *1*, 8–17.

Hester, L. (2013, June 17). The Oculus Rift will make you a virtual believer. *Complex Gaming* [Online]. Available: http://www.complex.com/video-games/2013/06/the-oculus-rift-will-make-you-a-virtual-believer.

Hettinger, L. J., Schmidt, T., Jones, D. L., & Keshavarz, B. (2014). Illusory self-motion in virtual environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 467–492). Boca Raton, FL: Taylor & Francis Group, Inc.

Holz, T., Dragone, M., & O'Hare, G. M. P. (2009). Where robots and virtual agents meet: A survey of social interaction across Milgram's Reality-Virtuality Continuum. *International Journal of Social Robotics*, *1*(1), 83–93.

Imagination. (2013). *Imagination, Caustic* [Online]. Available: https://www.caustic.com/.

Inta, R., Bowman, D. J., & Scott, S. M. (2012). The "Chimera": An off-the-shelf CPU/GPGPU/FPGA hybrid computing platform. *International Journal of Reconfigurable Computing*, *2012*, Article ID 241439, doi:10.1155/2012/241439.

ITRS. (2011). *International technology roadmap for semiconductors (2011 Edition): Emerging research materials* [Online]. Available: http://www.itrs.net/Links/2011ITRS/2011Chapters/2011ERM.pdf.

Jagadeesh, K. (2013, June 13). Virtual race game in the physical setting of your home allows for books and cereal bowls to become obstacles. *PSFK* [Online]. Available: http://www.psfk.com/2013/06/tabletop-augmented-reality-game.html.

Jones, D. L., Dechmerowski, S., Oden, R., Lugo, V., Wang-Costello, J., & Pike, W. (2014). Olfactory interfaces. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual* environments*: Design, implementation, and applications* (2nd ed., pp. 131–162). Boca Raton, FL: Taylor & Francis Group, Inc.

Kapralos, B., Jenkin, M., & Milios, E. (2008). Virtual audio systems. *Presence: Teleoperators and Virtual Environments*, *17*(6), 527–549.

Kawasaki, H., Endo, T., Mouri, T., Ishigure, Y., & Daniulaitis, V. (2013). HIRO: Multi-fingered haptic interface robot and its medical application systems. In I. Galiana & M. Ferre (Eds.), *Multi-finger haptic interaction*, Springer Series on Touch and Haptic Systems (Chapter 5, pp. 85–107). London, U.K.: Springer.

Kennedy, R. S., Kennedy, R. C., Kennedy, K. E., Wasula, C., & Bartlett, K. M. (2014). Virtual environments and product liability. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 505–518). Boca Raton, FL: Taylor & Francis Group, Inc.

Kenyon, R., & Afenya, M. (1995). Training in virtual and real environments. *Annals of Biomedical Engineering*, *23*(4), 445–455.

Keshavarz, B., Heckt, H., & Lawson, B. D. (2014). Visually induced motion sickness. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 647–698). Boca Raton, FL: Taylor & Francis Group, Inc.

Kheddar, A., Chellali, R., & Coiffet, P. (2014). Virtual environment–Assisted teleoperation. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 1107–1142). Boca Raton, FL: Taylor & Francis Group, Inc.

Kim, D. W. (2013). Aroma chip using a functional polymer gel. In T. Nakamoto (Ed.), *Human olfactory displays and interfaces: Odor sensing and presentation* (pp. 384–400). Hershey, PA: IGI Global.

Kokini, C., Carroll, M., Ramirez-Padron, R., Hale, K., Sottilare, R., & Goldberg, B. (2012). Quantification of trainee affective and cognitive state in real time. In *Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC 2012)*, Orlando, FL. Paper # 12064.

Kopper, R., Ni, T., Bowman, D., & Pinho, M. (2006, March 25–29). Design and evaluation of navigation techniques for multiscale virtual environments. In *Proceedings of IEEE Virtual Reality Conference* (*VR 2006*), pp. 175–182, Alexandria, VA, March 25–29, 2009. IEEE Computer Society 2006 ISBN 1-4244-0224-7.

Kozak, J. J., Hancock, P. A., Arthur, E. J., & Chrysler, S. T. (1993). Transfer of training from virtual reality. *Ergonomics*, *36*(7), 777–784.

Kuchenbecker, K. J. (2008). Haptography: Capturing the feel of real objects to enable authentic haptic rendering. In R. Whitaker & B. Liang (Eds.), *Proceedings of the Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems* (*HAS'08*) (pp. 3:1–3:3). Brussels, Belgium: Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering (ICST).

Lanier, J. (2013). *Who owns the future?* New York, NY: Simon & Schuster.

Lanier, R., Bond, A., Johnston, M., LaViola, J., Hale, K., & Nguyen, N. (2013). *Ecological Gesturing Ontology (EGO)*. Final report delivered to AFRL under contract FS8650-13-M-6389.

Lawler, R. (2013, May 9). With $3.2M in funding, 'real-world operating system' startup Dekko refocuses to build its own augmented reality apps. *TechCrunch* [Online]. Available: http://techcrunch.com/2013/05/09/dekko-real-world-os/.

Lawson, B. D. (2014a). Motion sickness symptomatology and origins. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 531–600). Boca Raton, FL: Taylor & Francis Group, Inc.

Lawson, B. D. (2014b). Motion sickness scaling. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 601–626). Boca Raton, FL: Taylor & Francis Group, Inc.

Lécuyer, A., Lotte, F., Reilly, R. B., Leeb, R., Hirose, M., & Slater, M. (2008). Brain-computer interfaces, virtual reality, and videogames. *Computer*, *41*(10), 66–72.

Lee, N. (2014). castAR's vision of immersive gaming gets closer to final production engadget. http://www.engadget.com/2014/03/20/castar-update-gdc/.

Luebke, D., & Humphreys, G. (2007). How GPUs work. *IEEE Computer*, *40(*2), 96–100.

Matsukura, H., Yoneda, T., & Ishida, H. (2012, March 4–8). Smelling screen: Technique to present a virtual odor source at an arbitrary position on a screen. In *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE* (pp. 127–128), Orlando, FL. doi:10.1109/VR.2012.6180915.

McGrath, B., McKinley, A., Duistermaat, M., Carlander, O., Brill, C., Zets, G., & van Erp, J. B. F. (2008). Chapter 4 – Tactile actuator technology. In J. B. F. van Erp & B. P. Self (Eds.), *Tactile displays for orientation, navigation and communication in air, sea and land environments* (pp. 4-1–4-12) (Tech. Rep. No. RTO-TR-HFM-122). Neuilly sur Seine, France: The Research and Technology Organisation (RTO) of the North Atlantic Treaty Organization (NATO).

McLeran, A., Roads, C., Sturm, B. L., & Shynk, J. J. (2008, July 31–August 3). Granular sound spatialization using dictionary-based methods. In *Proceedings of the 5th Sound and Music Computing Conference*, Berlin, Germany.

McMahan, R. P., Kopper, R., & Bowman, D. A. (2014). Principles for designing effective 3D interaction techniques. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 285–312). Boca Raton, FL: Taylor & Francis Group, Inc.

Miller, J. D. (2012). *SLAB3D User Manual* (v6.7.0) [Online]. Available: http://home.earthlink.net/~slab3d/slab3d_user_manual.pdf.

Mims, C. (2010, October 22). Whatever happened to virtual reality? *MIT Technology Review* [Online]. Available: http://www.technologyreview.com/view/421293/whatever-happened-to-virtual-reality/.

Munoz, R., Barcelos, T., & Chalegre, V. (2011, November). Defining and validating virtual worlds usability heuristics. In *Proceedings of the 30th International Conference of the Chilean Computer Science Society (SCCC)* (pp. 171–178), Curico, Chile.

Munro, A., Patrey, J., Sheldon, E. S., & Carroll, M. (2014). Cognitive aspects of virtual environment design. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 391–410). Boca Raton, FL: Taylor & Francis Group, Inc.

Nakaizumi, F., Noma, H., Hosaka, K., & Yanagida, Y. (2006, March 25–29). SpotScents: A novel method of natural scent delivery using multiple scent projectors. In *Virtual Reality Conference 2006* (pp. 207–214), Alexandria, VA. doi: 10.1109/VR.2006.122.

Nakamoto, T. (Ed.). (2013). *Human olfactory displays and interfaces: Odor sensing and presentation.* Hershey, PA: IGI Global.

National Academy of Engineering. (2008). *Leading engineers and scientists identify advances that could improve quality of life around the world: 21st century's grand engineering challenges unveiled.* Washington, DC: Author. [online]. Available: http://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID = 02152008.

Newman, J. (2013). The Oculus Rift commercializes virtual reality. *Desktop Engineering* [Online]. Available: http://www.engineeringontheedge.com/2013/06/the-oculus-rift-commercializes-virtual-reality/.

Pagan, B. (2012, May 7). New design practices for touch-free interactions. *UX Magazine*, Article No. 824. Available: http://uxmag.com/articles/new-design-practices-for-touch-free-interactions, viewed 2013, October 4.

Park, K.-H. (2013). A ubiquitous motion tracking system using sensors in a personal health device. *International Journal of Distributed Sensor Networks*, Article ID 298209. Available: http://dx.doi.org/10.1155/2013/298209.

Paul, I. (2013). The end of Moore's Law is on the horizon, says AMD. *PC World* [Online]. Available: http://www.pcworld.com/article/2032913/the-end-of-moores-law-is-on-the-horizon-says-amd.html.

Peters, N., Lossius, T., & Schacher, J. C. (2013). The spatial sound description interchange format: Principles, specification, and examples. *Computer Music Journal*, *37*(1), 11–22.

Peters, N., Lossius, T., Schacher, J., Baltazar, P., Bascou, C., & Place, T. (2009, July 23–25). A stratified approach for sound spatialization. In *Proceedings of the SMC 2009 – 6th Sound and Music Computing Conference* (pp. 219–224), Porto, Portugal.

Picheny, M., Nahamoo, D., Goel, V., Kingsbury, B., Ramabhadran, B., Rennie, S. J., & Saon, G. (2011). Trends and advances in speech recognition. *IBM Journal of Research and Development*, *55*(5), 2:1–2:18. doi:10.1147/JRD.2011.2163277.

Poletti, M. A., Fazi, F. M., & Nelson, P. A. (2011). Sound reproduction systems using variable-directivity loudspeakers. *Journal of the Acoustical Society of America*, *129*(3), 1429–1438.

Popescu, G. V., Trefftz, H., & Burdea, G. C. (2014). Multimodal interaction modeling. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 411–434). Boca Raton, FL: Taylor & Francis Group, Inc.

Rahikainen, V. (2013). *On real-time ray tracing* (MSc thesis). University of Tampere School of Information Sciences, Computer Science, Finland [Online]. Available: http://tutkielmat.uta.fi/pdf/gradu06763.pdf.

Ramirez, E. (2013, June 12). Augmented reality, seen through Meta Glasses. *The Wall Street Journal, WSJ Blogs, Digits* [Online]. Available: http://blogs.wsj.com/digits/2013/06/12/augmented-reality-seen-through-meta-glasses/.

Rautaray, S. S., & Agrawal, A. (2012). Real time hand gesture recognition system for dynamic applications. *International Journal of Ubiquitous Computing*, *3*(1), 21–31. Available: http://airccse.org/journal/iju/papers/3112iju03.pdf, viewed 2013, October 4.

Rivington, J. (2013). Google Glass: What you need to know: Are Google's glasses more than just a gimmick? *techRadar* [Online]. Available: http://www.techradar.com/us/news/video/google-glass-what-you-need-to-know-1078114.

Road to VR. (2013). *Head mounted display (HMD)/VR headset comparison chart* [Online]. Available: http://www.roadtovr.com/head-mounted-display-hmd-vr-headset-comparison.

Robertson, A. (2013, January 10). Innovega combines glasses and contact lenses for an unusual take on augmented reality. *The Verge* [Online]. Available: http://www.theverge.com/2013/1/10/3863550/innovega-augmented-reality-glasses-contacts-hands-on.

Roman, P., Lazarov, M., & Majumder, A. (2009). A scalable distributed paradigm for multi-user interaction with tiled rear project display walls. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1623–1632.

Rose, F. D., Attree, E. A., Brooks, B. M., Parslow, D. M., Penn, P. R., & Ambihaipahan, N. (1998). Transfer of training from virtual to real environments. In P. Sharkey, F. D. Rose, & J-I Lindstrom (Eds.), *Proceedings of the International Conference Series on Disability, Virtual Reality, & Associated Technologies* (pp. 69–75). Berkshire, U.K.: University of Reading.

Rose, F. D., Brooks, B. M., & Attree, E. A. (2002). An exploratory investigation into the usability and usefulness of training people with learning disabilities in a virtual environment. *Disability and Rehabilit*ation, *24*(11–12), 627–633.

Ruppert, G. C., Reis, L. O., Amorim, P. H., de Moraes, T. F., & Da Silva, J. V. (2012). Touchless gesture user interface for interactive image visualization in urology surgery. *World Journal of Urology*, *30*(5), 687–691.

Rusu, C., Munoz, R., Roncagliolo, S., Rudloff, S., Rusu, V., & Figueroa, A. (2011, August). Usability heuristics for virtual worlds. In *Proceedings of the 3rd International Conference Advances in Future Internet (AFIN)* (pp. 16–19), French Riviera, Nice/Saint Laurent du Var, France.

Salzman, M., Dede, C., & Loftin, R. (1995, October). Usability and learning in education virtual realities. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 486–490), San Diego, CA.

Sawyerr, W., Brown, E., & Hobbs, M. (2013). Using a hybrid method to evaluate the usability of a 3D virtual world user interface. *International Journal of Information Technology & Computer Science*, *8*(2), ISSN No: 2091-1610 [Online]. Available: http://www.ijitcs.com/volume%208_No_2/William.pdf.

Schobben, D. W. E., & Aarts, R. M. (2005). Personalized multi-channel headphone sound reproduction based on active noise cancellation. *Acta Acustica*, *91*(3), 440–450.

Seide, F., Li, G., & Yu, D. (2011, August 28–31). Conversational speech transcription using context-dependent deep neural networks. In *Interspeech 2011* (pp. 437–440), Florence, Italy.

Seo, J., Yoo, J. H., Kyeongok Kang, K., & Fazi, F. M. (2010, May 22–25). 21-Channel surround system based on physical reconstruction of a three-dimensional target sound field. In *Proceedings of the 128th Convention of the Audio Engineering Society* (Paper 7973, p. 11), London, U.K.

Shilov, A. (2013, June 13). Oculus VR demonstrates full-HD Oculus Rift virtual reality helmet. *Xbit Laboratories* [Online]. Available: http://www.xbitlabs.com/news/multimedia/display/20130613230543_Oculus_VR_Demonstrates_Full_HD_Oculus_Rift_Virtual_Reality_Helmet.html.

Slater, M., Moreton, R., Buckley, K., & Bridges, A. (2008). A review of agent emotion architectures. *Journal for Computer Game Culture*, *2*(2), 203–214.

Smalley, D. E., Smithwick, Q. Y. J., Bove, V. M., Barabas, J., & Jolly, S. (2013). Anisotropic leaky-mode modulator for holographic video displays. *Nature*, *498*(7454), 313–317.

Stanney, K. M., Chen, J., & Wedell, B. (2000). *Navigational metaphor design* (Final Rep. Contract No. N61339-99-C-0098). Orlando, FL: Naval Air Warfare Center Training Systems Division.

Stanney, K. M., Kennedy, R. S., & Hale, K. S. (2014). Virtual environment usage protocols. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 796–808). Boca Raton, FL: Taylor & Francis Group, Inc.

Stanney, K. M., Mollaghasemi, M., & Reeves, L. (2000). *Development of MAUVE, the multi-criteria assessment of usability for virtual environments system* (Final Rep. Contract No. N61339-99-C-0098). Orlando, FL: Naval Air Warfare Center Training Systems Division.

Stanney, K. M., Mourant, R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments*, *7*(4), 327–351.

Stone, R. J. (2012). *Human factors guidance for designers of interactive 3D and games-based training systems*. Birmingham, U.K.: University of Birmingham.

Stuart, R. (1996). *The design of virtual environments*. New York: McGraw-Hill.

Sugimoto, S., & Okada, K. (2013). Olfactory display based on ink jet printer mechanism and its presentation techniques (pp. 401–414). In T. Nakamoto (Ed.), *Human olfactory displays and interfaces: Odor sensing and presentation.* Hershey, PA: IGI Global.

Sun, Y., Ma, T., Han, C. Y., Ross, J., & Wee, W. (2013). A study of tracking system using multiple emitter towers. *Advanced Materials Research*, *694–697*, 927–935.

Taube, J. S., Valerio, S., & Yoder, R. M. (2013). Is navigation in virtual reality with fMRI really navigation? *Journal of Cognitive Neuroscience*, *25*(7), 1008–1019. doi:10.1162/jocn_a_00386.

Thomas, D. B., Howes, L., & Luk, W. (2009, February 22–24). A comparison of CPUs, GPUs, FPGAs, and massively parallel processor arrays for random number generation. In *Proceedings of the ACM/SIGDA International Symposium on Field Programmable Gate Arrays – FPGA'09* (pp. 63–72), Monterey, CA.

Tran-Gia, P. (2007, May 21–23). Trends towards next generation internet. In *Proceedings of NGI 2007: 3rd EURO-NGI Conference on Next Generation Internet Networks*, Trondheim, Norway. doi:10.1109/NGI.2007.371188.

Turk, M. (2014). Gesture recognition. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments* (2nd ed.). Boca Raton, FL: Taylor & Francis Group, Inc.

Tweney, D. (2013, May 30). Atheer Labs demonstrates a 3D, virtual-reality headset technology. *VentureBeat* [Online]. Available: http://venturebeat.com/2013/05/30/atheer-labs-demonstrates-a-3d-virtual-reality-headset-technology.

Tyson, A. (2013, March). What do we know about spatial navigation, and what else could model-based fMRI tell us? *The Einstein Journal of Biology and Medicine*, Online First: E1–E11. http://www.einstein.yu.edu/uploadedFiles/Pulications/EJBM/2013% 20Tyson%20EJBM%20Online%20First.pdf.

Van Baelen, W., Bert, T., Claypool, B., & Sinnaeve, T. (2011, December). A new dimension in cinema sound: Auro-3D. *Cinema Technology*, *24*(4), 26–31.

Venkatesh, S., & Buehrer, R. M. (2007). Non-line-of-sight identification in ultra-wideband systems based on received signal statistics. *IET Microwaves, Antennas and Propagation*, *1*(6), 1120–1130.

Vorländer, M. & Shinn-Cunningham, B. (2014). Virtual auditory displays. In K. S. Hale & K. M. Stanney, (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 87–114). Boca Raton, FL: Taylor & Francis Group, Inc.

Vourvopoulos, A., Liarokapis, F., & Petridis, P. (2012, September 2–5). Brain-controlled serious games for cultural heritage. In *Proceedings of the 18th International Conference on Virtual Systems and Multimedia, Virtual Systems in the Information Society* (*VSMM-2012*) (pp. 291–298), IEEE Computer Society, Milan, Italy.

Wann, J. P., White, A. D., Wilkie, R. M., Culmer, P. R., J. Peter A. Lodge, & Mon-Williams, M. (2014). Measurement of visual aftereffects following virtual environment exposure. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 809–832). Boca Raton, FL: Taylor & Francis Group, Inc.

Withana, A., Kondo, M., Makino, Y., Kakehi, G., Sugimoto, M., & Inami, M. (2010). ImpAct: Immersive haptic stylus to enable direct touch and manipulation for surface computing. *Computer Entertainment*, *8*(9):1–9:16.

Wollman, D. (2012, February 3). Lumus' OE-31 optical engine turns motorcycle helmets, other eyewear into wearable displays. *Engadget* [Online]. Available: http://www.engadget.com/2012/02/23/lumus-oe-31-optical-engine-turns-motorcycle-helmets-other-eyew/.

Wozniewski, M., Settel, Z., Quessy, A., Matthews, T., & Courchesne, L. (2012, September 9–15). SpatOSC: Providing abstraction for the authoring of interactive spatial audio experiences. *International Computer Music Conference (ICMC 2012)*, Ljubljana, Slovenia.

Yanagida, Y., & Tomono, A. (2013). Basics for olfactory display. In T. Nakamoto (Ed.), *Human olfactory displays and interfaces: Odor sensing and presentation* (pp. 60–85). Hershey, PA: IGI Global.

Zou, W. (2013). A distributed system for independent acoustic source positioning using magnitude ratios. In Y. Yang & M. Ma (Eds.), *Proceedings of the 2nd International Conference on Green Communications and Networks 2012 (GCN 2012)* (Vol. 2, Lecture Notes in Electrical Engineering, Vol. 224, pp. 109–115). Berlin, Germany: Springer.

Zuckerberg, M. (2014, March 25). https://www.facebook.com/zuck/posts/10101319050523971.

Zyda, M., & Sheehan, J. (Eds.). (1997). *Modeling and simulation: Linking entertainment & defense*. Washington, DC: National Academy Press.

2 Virtual Environments Standards and Terminology

Richard A. Blade and Mary Lou Padgett

CONTENTS

2.1 INTRODUCTION

In just the past decade, virtual environment (VE) applications have emerged in entertainment, training, education, and other areas (see Chapters 37 through 49, this book). In that time, extensive research in VE technology has also been conducted. However, the terminology used to characterize this technology is still evolving. In fact, Durlach and Mavor (1995, p. 2) indicate that "inadequate terminology [is] being used" to describe VE technology and its applications. It is thus important to describe the key terms that are used in this handbook. The objective is not to resolve differences between disparate uses (in fact, often multiple, even conflicting definitions are presented) but rather to provide a coherent set of commonly used terms. While it is customary to present a glossary at the end of handbooks such as this one, this work starts out with a glossary so that readers may develop a common understanding of the terms used throughout the handbook. Paradoxically, the one term that remains particularly elusive is *virtual environment*. Many authors, especially those among the application chapters (see Chapters 36 through 50, this book), have catered the definition of VE to fit the forms of the technology that best suit their needs. Perhaps this definitional multiplicity demonstrates the versatile nature of VE technology and its wide array of potential uses.

While the definitions in this chapter have been presented in a relatively informal manner, the profession continues to work toward a set of standard definitions through the VR Terminology Project. This involves a multiphase effort on the part of the IEEE Computer Society under the auspices of the IEEE Standards Association, of which the first phase is the establishment of a working group on virtual reality (VR) terminology (VI-1392). The first author of this chapter chaired the VI-1392 working group in the very early years of VR and in this chapter attempts to present the most fundamental definitions with the greatest generality.

2.2 IMPORTANCE OF STANDARDS

Standards are critical for systematic and robust development of any emerging technology. *Specification standards* provide for practical descriptions of product characteristics and limitations, critical to an end user. *Interface standards* allow for interchangeability of components developed by different manufacturers, thus permitting specialization and robust competition in the marketplace.

Safety standards ensure the health and safety of product users. *Terminology standards* ensure that technical terminology is used in a consistent and rigorous manner, thus preventing confusion and ambiguity in scientific and technical reports and specifications.

The process of establishing standards is quite possibly as important as the standards ultimately produced for that process establishes a forum for open and systematic dialog between researchers, developers, manufacturers, and end users on the current and future needs and directions of an industry. It often progresses at a *glacial pace*, meaning it could take years to actually reach agreement on a standard, which ensures a systematic, objective, and rigorous examination of all aspects of an issue by all who have interest or involvement.

There are many standards-setting organizations, each having its own rules and procedures. For electrical, electronic, and computer-related standards, the two primary international organizations are the *International Electrotechnical Commission* (IEC) and the *International Organization for Standardization* (ISO). Most standards related to VEs will originate in the IEEE Computer Society, which is a part of the *Institute of Electrical and Electronic Engineers* (IEEE). *The IEEE Standards Association* (IEEE-SA), in turn, belongs to the *American National Standards Institute* (ANSI). Standards organizations can and do work together, but there are no requirement that they do so. Rather, obtaining the broadest possible acceptance of standards is the motivation for cooperation and collaboration. For example, the ISO and IEC have formed the *Joint Technical Committee Number One* (JTC1) to deal with information technology standards. They also cooperate on issues involving safety, electromagnetic radiation, and the environment.

The first step to establishing one or more VE standards (usually a group of related standards) in the IEEE Computer Society is for an individual to prepare a *project authorization request* (PAR) for submission to the *IEEE Standards Board*. If approved, that PAR is given an identifying label (e.g., vi-1392 in the case of VE terminology), and a working group is set up under that label to study the problem and make recommendations. All interested parties are invited to participate and every attempt is made to ensure a broad representation. After much informal discussion and deliberation, the working group prepares draft standards. A *balloting group* is then assembled to vote on the standards and/or modify them if necessary. In contrast to the working group, the process and membership of the balloting group become very formal to ensure a balanced and fair consideration of the issues and proposed standards. The last step in establishing a standard by the IEEE-SA is for the *IEEE Standards Board* to approve the submission of the balloting group. When the standards being proposed overlap between two or more existing groups, a *standards coordinating committee* (SCC) is involved to ensure proper coordination and collaboration. The work of the IEEE-SA is supported by its publication of the standards. Various procedures must be followed throughout the standards process to ensure that the publication rights to each standard produced belong to the IEEE.

2.3 IMPORTANCE OF OFFICIAL TERMINOLOGY

Any standard set by the IEEE or other standards development organizations contains definitions. When multiple uses of a term are found in the literature or are in common use, it should help to have a careful record of the variations, including recommendations and cautions. Consider, for example, the term *megabyte*. It may mean $10^6 = 1{,}000{,}000$ bytes, or it may mean $2^{10} = 1{,}024 \times 1{,}024 = 1{,}048{,}576$ bytes, or even $1{,}000 \times 1{,}024 = 1{,}024{,}000$ bytes. A consumer needs to be aware of these disparities. One important task of the VR Terminology Project is to provide consumers, including government contractors, reliable definitions so they can properly evaluate product descriptions and proposals. If these definitions are readily available to users and producers, reliable commerce will be enhanced.

A short glossary of terms is given in the next section. Most are technical terms that were provided by the authors of other chapters in this handbook due to their occurrence in those chapters, and many appear in more than one chapter. Although all have definitions in the professional literature, almost none of those definitions have been universally adopted by any standards organization. In preparing this chapter, much discussion took place with the authors, as well as among various authors of this handbook, to ensure a reasonable degree of clarity and consistency. However, in some cases, we found it necessary to provide multiple, sometimes even contradictory, definitions reflecting current usage.

2.4 BASIC GLOSSARY

6DOF: abbreviation for six degrees of freedom, the six being freedom of a three-dimensional object to move in the directions of three perpendicular axes (forward/backward, up/down, and left/right) and rotate about three perpendicular axes (pitch, yaw, and roll).

Accommodation: change in the focal length of the eye's lens to maintain focus on a moving close object.

Active virtual reality: virtual reality where human actions control the model of reality.

Actuator: mechanical means used to provide force or tactile feedback to a user.

Aftereffect: any effect of *VE* exposure that is observed after a participant has returned to the physical world.

Archetypes: prototypes that provide templates to guide learning, development, and the construction of the personality or psyche.

Articulation: objects composed of several parts that are separably moveable.

Artificial reality: same as virtual reality or virtual environment.

Assistive agents: artificial intelligence algorithms developed to guide users through a VR world and to coach the user on available choices within the world.

Attentional inertia: attentional mechanism in which the person's visual system gets locked into engaging, interesting experiences.

Aubert effect or phenomenon: the apparent displacement of an isolated vertical line in the direction opposite to which the observer is tilted. This happens when an observer has a large tilt, for example, 90°.

Audification: an acoustic stimulus involving direct playback of data samples. See also *Sonification*.

Augmented reality: form of virtual reality where the human interacts with a combination of the reality model and true reality, usually through the use of special eyeglasses displaying both data from the model and data from the real world. (An industrial example is Boeing workers constructing wire harnesses using special glasses to display the actual path of each wire in the harness.)

Autonomy: performance or action of the object on the rule of physics, biology, or a virtual world, but not by independent decision of a human operator.

Avatar: an interactive representation of a human in a virtual reality environment.

Back clipping plane: region at a distance beyond which objects are not shown.

Backdrop: stationary background in a virtual world; the boundary of the world that cannot be moved or broken into smaller elements.

Backface removal: elimination of those portions of a displayed object that are facing away from the viewer.

Binocular: displaying a slightly different view to each eye for the purpose of stereographic viewing.

Binocular Omni-Orientation Monitor (BOOM): 3D display device suspended from a weighted boom that can swivel freely so the viewer can use the device by bringing the device up to the eyes and viewing the 3D environment while holding it. The boom's position and orientation communicates the user's point of view to the computer.

Binocular parallax: the means whereby the eyes can judge distance by noticing how closer objects appear to move more than distant ones when the observer moves. See also *Parallax* and *Motion parallax*.

Bi-ocular: displaying the same image to each eye. Sometimes, this is done to conserve computing resources when depth perception is not critical. See also *Stereopsis*.

Biosensors: sensor devices that monitor the state of the body.

Bots: robots or intelligent agents who roam multiuser domains (MUDs) and other virtual environments.

Browser: overviews, such as indexes, lists, or animated maps, which provide a means of navigating through the physical, temporal, and conceptual elements of a virtual world.

CAVE: *cave automatic virtual environment*; virtual images projected on the walls, floor, and ceiling of a room that surrounds a viewer. Oftentimes, a principal viewer's head position is tracked to determine view direction and content, while other viewers *come along for the ride*.

Cognitive map: mental representation of an environment (also referred to as a mental map).

Communication channel: when applied to HCI, it is a pathway between the user and the simulation that allows human–computer interaction.

Computer-assisted teleoperation (CAT): bilateral control of teleoperation through computers, including computer assistance in both robot control and information feedback.

Convergence: occurs in stereoscopic viewing when the left- and right-eye images become fused into a single image.

Convolve: to filter and intertwine signals (e.g., sounds) and render them 3D. Used in VE applications to recreate sounds with directional cues.

Convolvotron: output system for controlling binaural sound production in a VR world.

Coordinates: set of data values that determine the location of a point in a space. The number of coordinates corresponds to the dimensionality of the space.

Coriolis, or cross-coupling, effect: effect resulting from certain kinds of simultaneous multi-axis rotations, especially making head movements while rotating. This illusion is characterized by a feeling of head or self-velocity in a curved path that is roughly orthogonal to the axes of both body and head rotation, which can lead to simulator sickness.

Cue conflict: theory to explain the kind of motion sickness caused when the body tries to interpret conflicting clues being received by the senses. Frequent causes are faulty calibration of eye devices or delay between the sensory inputs and output display.

Culling: removing invisible pieces of geometry and only sending potentially visible geometry to the graphics subsystem. Simple culling rejects entire objects not in the view. More complex systems take into account occlusion of some objects by others, for example, building hiding trees behind it.

Cutaneous senses: skin senses, including light touch, deep pressure, vibration, pain, and temperature.

Cybersickness: sensations of nausea, oculomotor disturbances, disorientation, and other adverse effects associated with VE, typically because the sensory data generated do not properly match reality.

Cyberspace: virtual universe of digital data.

Cyborg: robotic humanoid modeled directly from digital readings of a real human and transformed into a photo realistic, animated character produced via illusionary metamorphosis.

DataGlove: device for sensing hand gestures, which uses fiber-optic flex sensors to track hand orientation and position, as well as finger flexure.

Data sonification: assignment of sounds to digitized data that may involve filtering to give illusion of localized sound.

Data specialization: assignment of orientation (yaw, pitch) and position coordinates (x, y, z) to digital sounds assigned to data.

Deformable object technology (DOT): virtual objects, which bend and deform appropriately when touched.

Depth cueing: use of shading, texture mapping, color, interposition, or other visual characteristics to provide a cue for the distance of an object from the observer.

Desktop virtual systems: virtual experiences that are displayed on a 2D desktop computer; a person can see through the eyes of the character on the screen, but the experience is not 3D.

Dolly shot: display of a scene while moving forward or backward. See also *Pan shot* and *Track shot*.

Dynamic accuracy: system accuracy as a tracker's sensor is moved. See also *Static accuracy.*

Dynamics: rules that govern all actions and behaviors within the environment.

EBAT: *event-based approach to training;* provides the strategies, methods, and tools that are essential for an effective learning environment in a structured and measurable format for training and testing specific knowledge, skills, and abilities.

"E" effect: See *Muller effect.*

Effectors: interfacing devices used in virtual environments for input/output, tactic sensation, and tracking. Examples are gloves, head-mounted displays, headphones, and trackers.

Egocenter: sense of one's own location in a virtual environment.

Embodiment: within the body.

Ergonomics: study of human factors, that is, the interaction between the human and his or her working environment.

Exoskeletal device: flexible interaction devices worn by users (e.g., gloves and suits) or rigid-link interaction systems (i.e., jointed linkages affixed to users).

Exoskeleton: mechanically linked structure for control of feedback from an application.

Eye clearance: most accurate figure of merit used to describe the HMD positioning relative to the eye.

Eye tracking: measurement of the direction of gaze.

Eyeball in the hand: metaphor for visualized tracking where the tracker is held in the hand and is connected to the motion of the projection point of the display.

EyePhone: A VPL Research, Inc., display device consisting of two tiny television monitors (one per eye), earphones, and a sensor for tracking head position and orientation.

Fidelity: degree to which a VE or SE duplicates the appearance and feel of operational equipment (i.e., functional fidelity), sensory stimulation (i.e., physical fidelity), and psychological reactions felt in the real world (i.e., psychological fidelity) of the simulated context.

Field of view (FOV): angle in degrees of the visual field. Since a human's two eyes have overlapping 140° FOV, binocular or total FOV is roughly 180° in most people. A FOV greater than roughly 60°–90° may give rise to a greater sense of immersion.

Fish tank VE: illusion of looking *through* a computer monitor to a virtual outside world using a stereoscopic display system. When looking *out* through the stereo *window*, the observer imagines himself or herself to be in something resembling a fish tank.

Force feedback: output device that transmits pressure, force, or vibrations to provide a VE participant with the sense of resisting force, typically to weight or inertia. This is in contrast to tactile feedback, which simulates sensation to the skin.

Formal features: audiovisual production features that structure, mark, and represent media experiences.

Formative evaluation: assess, refine, and improve user interaction by iteratively placing representative users in task-based scenarios in order to identify usability problems as well as to assess a design's ability to support user exploration, learning, and task performance.

Fractal: self-similar graphical pattern generated by using the same rules at various levels of detail. That is, a graphical pattern that repeats itself on a smaller and smaller scale. Fractals can generate very realistic landscapes with great detail using very simple algorithms.

Frustum of vision: 3D field of view in which all modeled objects are visible.

Functional fidelity: degree VE mimics functional operation or relationship between objects in real world. See also *Fidelity.*

Gesture: hand motion that can be interpreted as a sign, signal, or symbol.

Gouraud shading: shading of polygons smoothly with bilinear interpolation.

GUI: abbreviation for *graphical user interface*.

Gravitoinertial force: resultant force combining gravity and virtual forces created by acceleration.

Guidelines-based evaluation: cost-effective and popular method for evaluating a user interface design; the goal is to find usability problems in a design so that they can be attended to as part of an iterative design process—involves having a small set of evaluators examine the interface to determine its compliance with recognized usability principles (the *heuristics*). See also *Heuristic evaluation*.

Gustatory: pertaining to the sense of taste.

Haptic interface: interface involving physical sensing and manipulation.

HCI: abbreviation for *human–computer interaction*.

Head-mounted display (HMD): a visual display covering the eyes, sometimes having position tracking to provide a computer with the location and orientation of the head.

Heads-up display: display device that allows users to see graphics superimposed on their view of the real world, a form of augmented reality.

Heuristic evaluation: cost-effective and popular method for evaluating a user interface design; the goal is to find usability problems in a design so that they can be attended to as part of an iterative design process—involves having a small set of evaluators examine the interface to determine its compliance with recognized usability principles (the *heuristics*). See also *Guidelines-based evaluation*.

Hidden surface: surface of a graphics object that is occluded from view by intervening objects.

Human–computer interaction (HCI): study of how people work with computers and how computers can be designed to help people effectively use them.

Image distance: perceived distance to the object (in contrast to the real object distance, if there exists a real object).

Immersion: experience of being physically within a VE experience. The term is sometimes subcategorized into external and internal immersion and sensory and perceptual immersion. See also *Presence*.

Intelligent user interface: user interface that is adaptive and has some degree of autonomy.

Interaural amplitude: differences between a person's two ears in the intensity of a sound, typically due to the location of the sound.

Interaural time: differences between a person's two ears in the phase of a sound, typically due to the location of the sound.

Inverse kinematics: specification of the motion of dynamic systems from properties of their joints and extensions.

Kinesthesis: sensations derived from muscles, tendons, and joints and stimulated by movement and tension.

Kinesthetic: all muscle, joint, and tendon senses. Excludes skins' senses, such as touch and vestibular and visual. A subset of somatosensory, which is usually applied to limb position and movement but would also include nonvestibular sensation of head movement (e.g., via neck musculature).

Kinesthetic dissonance: mismatch between feedback and its absence from touch or motion, during VE experiences.

Latency: lag between user motion and tracker system response. Delay between actual change in position and reflection by the program. Delayed response time.

Level of detail (LOD): model of a particular resolution among a series of models of the same object. Greater graphic performance can be obtained by using a lower LOD when the object occupies fewer pixels on the screen or is not in a region of significant interest.

Locomotion: means of travel restricted to self-propulsion. Motion interfaces can be subdivided into those for passive transport (inertial and noninertial displays) and those for active transport (locomotion interfaces).

Magic wand: 3D input device used for pointing and interaction. A sort of 3D mouse.

Master–slave system: teleoperator consisting of a control, called the master, and a remote executing device, called the slave.

Matching: degree to which a VE mimics reality not only in form but also in terms of function and the behaviors the VE elicits from users.

Metaball: surface defined about a point specified by a location, a radius, and an *intensity*. When two metaballs come in contact, their shapes blend together.

Metallic distortion: noise interference or degraded performance in electromagnetic trackers when used near large metallic objects.

Model: simulation of something real.

MOO: object-oriented multiuser domain, a text-based MUD. See also *MUD*.

Motion parallax: the means whereby the eyes can judge distance by noticing how closer objects appear to move more than distant ones when the observer moves. See also *Binocular parallax* and *Parallax*.

Motion platform: platform, often carrying a passenger *pod* resembling an automobile, airplane, or spacecraft that provides the sensations of motion through orientation, vibrations, and jerking and occasionally through spinning.

Move-and-wait strategy: typical teleoperation control mode when time delays occur (with direct feedback).

MUD: *multiuser domain* where users can jointly interact and play games, such as dungeons and dragons.

Muller effect: tendency to perceive an objectively vertical line as slightly tilted in the same direction as the observer when the observer is tilted by only a moderate amount. See also *"E" effect*.

Multimodal command: command issued by the user to a computer simulation using several input communication channels.

Multimodal system: when applied to HCI, a system that allows communication with computers via several modalities, such as voice, gesture, gaze, and touch.

Nanomanipulator: device that allows manipulation on a very microscopic scale.

Navigation: aggregate task of wayfinding and motion (i.e., motoric element of navigation).

Neural interface: ultimate human–computer interface, which connects directly to the human nervous system.

Nystagmus: reflexive eye movements that usually serve to keep the world steady on the retina during self-motion. These eye movements are driven by vestibulo-ocular, cervical-ocular, and optokinetic reflexes. A rapid sideways snap of the eye followed by a slow return to normal fixation or rapid oscillatory movements of the eye, as in following a moving target, in blindness, or after rotation of the body.

Occipital cortex: back of the brain receiving retinotopic projections of visual displays.

Occlusion: hiding an object or a portion of an object from sight by interposition of other objects.

Oculogravic illusion: illusion of visual target displacement that is actually caused by body acceleration. The visual component of the somatogravic illusion (e.g., when an aircraft accelerates and there is a backward rotation of the resultant force vector, in addition to feeling a *pitch-up* sensation, the pilot may sense an apparent upward movement of objects in his or her visual field).

Oculogyral illusion: illusion of visual target displacement that is actually caused by body rotation (e.g., the illusory movement of a faint light in a darkened room following rotation of the body).

Olfactory: pertaining to the sense of smell.

Otoconia: hair cell mechanoreceptors that are embedded in gelatinous membranes containing tiny crystals of calcium carbonate through which the otolith organs detect changes in the magnitude or direction of linear acceleration vectors.

Pan: angular displacement of a view along any axis of direction in a 3D world.

Pan shot: display of a scene while moving about any axis. See also *Dolly shot* and *Track shot*.

Parallax: difference in viewing angle created by having two eyes looking at the same scene from slightly different positions, thereby creating a sense of depth. Also referred to as binocular parallax. See also *Motion parallax*.

Parietal cortex: area of the brain adjacent and above the occipital cortex, which processes spatial location and direction information.

Passive virtual reality: virtual reality in which there is no control of the model by the human. That is, the model only provides information to the human senses.

Persona: public display, or mask, of the self.

Perspective: rules that determine the relative size of objects on a flat-viewing surface to give the perception of depth.

Phase lag: when output from computer-generated images lags the actual position of tracking sensors.

Phi phenomenon: illusion of continuous movement through space induced by a sequence of discontinuous events (e.g., lights flashing in sequence, as on a marquee).

Phong shading: method for calculating the brightness of a surface pixel by linearly interpolating points on a polygon and using the cosine of the viewing angle. Produces realistic shading.

Photo realism: attempt to create realistic appearing images with great detail and texture.

Physical fidelity: degree VE mimics real world in regard to the visual, auditory, and haptic cues present (more similar to sensory simulation). See also *Fidelity*.

Pitch: angular displacement about the horizontal axis perpendicular to the lateral axis. A measure of dipping forward or backward.

Portal: simulated openings that a user can pass through in a virtual space to automatically load a new world or execute a user-defined function.

Position sensor: tracking device that provides information about its location and/or orientation.

Position trigger: hot spot, sensitive spot, or button that causes a change in the computer program when touched in some way.

Presence: illusion of being part of a virtual environment. The more immersive a VE experience, the greater the sense of being part of the experience. See also *Immersion*.

Proprioceptive: internal sense of body position and movement. Includes kinesthetic and vestibular senses but excludes outwardly directed senses such as vision and hearing. This term is not a subset of somatosensory nor is it synonymous with kinesthesia.

Prosocial behavior: socially valued behaviors such as helping, sharing, cooperating, and engaging in constructive imaginary activities.

Pseudo: false.

Radiosity: diffuse illumination calculation system for graphics based on energy balancing that takes into account multiple reflections off many walls.

Ray tracing: technique for displaying a 3D object with shading and shadows by tracing light rays backward from the viewing position to the light source.

Real time: action taking place with no perceptible or significant delay after the input that initiates the action.

Real-time imaging: graphics or images synchronized with real-world time and events.

Reality engine: computer system for generating virtual objects and environments in response to user input, usually in real time.

Refresh rate: frequency with which an image is regenerated on a display surface.

Registration: correspondence between a user's actual position and orientation and that reported by position trackers.

Render: conversion of image data into pixels to be displayed on a screen.

Retinal binocular disparity (RBD): ratio of the convergence angle of an image to the convergence angle of an object.

Roll: angular displacement about the lateral axis.

Tactile display: device that provides tactile and kinesthetic sensations.

Scenes view: virtual display viewed on a large screen or through a terminal window rather than with immersive devices.

Scientific visualization: graphical representation of complex physical phenomena in order to assist scientific investigation and make inferences that are not apparent in numerical form.

Second life: popular website where people construct an avatar that represents themselves and interact socially in an online VE.

Second-person virtual systems: virtual environment experiences in which a person is represented by an on-screen avatar rather than being fully immersed in the environment.

Semantic unification: process of integration and synchronization of information from several input modalities (gesture, speech, gaze, etc.).

Semiocclusion: occlusion to one eye only.

Sensorial redundancy: presenting same or related sensory information to a user using several communication channels.

Sensorial substitution: using a different sensory channel to present information normally presented in the replaced modality.

Shared mental models: overlapping knowledge or understanding a task, team, equipment, and situation between team members.

Shared worlds: virtual environments that are shared by multiple participants.

Shutter glasses: glasses that alternately block out the left- and right-eye views in synchrony with the computer display of left- and right-eye images to provide stereoscopic images on the computer screen.

Side effect: any effect of *VE* exposure that is observed after a participant has returned to the physical world. See also *Aftereffect*.

Simulation overdose: spending too much time in virtual environments.

Simulator sickness: various disturbances, ranging in degree from a feeling of unpleasantness, disorientation, and headaches to extreme nausea, caused by various aspects of a synthetic experience. Possible factors include sensory distortions such as abnormal movement of arms and heads because of the weight of equipment, long delays or lags in feedback, and missing visual cues from convergence and accommodation.

Situational awareness: perception of elements in an environment within a volume of time and space, comprehension of their meaning, and projection of their status in the near future; *up-to-the-minute* cognizance required to operate or maintain the state of a system.

Six degrees of freedom (6DOF): ability to move in three independent directions and rotate about three independent axes passing through the center of the body. Thus, the location and orientation are specified by six coordinates.

Somatogravic illusion: illusions of false attitude occurring mainly in pitch or roll. When lacking appropriate visual cues, the balance organs and brain are unable to sort out a force vector (such as delivered by motion of an aircraft), which differs in magnitude and/or direction from the gravitational vector.

Somatosensory or somesthesia: stimuli or senses arising from the cutaneous, muscle, and joint receptors. Touch and pressure cues. Includes all body senses of the skin, muscle, joint, and internal organs (including Mittelstaedts kidney receptors) but excludes vestibular, visual, auditory, and chemical senses (taste and smell). Forms a superset of kinesthetic, tactile, and haptic stimuli or senses.

Sonification: data are used to control various parameters of a sound generator, thereby providing the listener with information about the controlling data. See also *Audification*.

Sopite syndrome: chronic fatigue, lethargy, drowsiness, nausea, etc.

Spatial navigation: self-orientation and locomotion in virtual worlds.

Spatially immersive display (SID): semisurrounding projected stereo displays.

Static accuracy: ability of a tracker to determine the coordinates of a position in space. See also *Dynamic accuracy*.

Stereopsis: binocular vision of images with different views by the two eyes to distinguish depth.

Summative evaluation: typically performed after a product or design is more or less complete. Its purpose is to statistically compare several different systems, for example, to determine which one is *better*, where better is defined in advance; involves placing representative users in task-based scenarios.

Synthetic experience (SE): experience created through a virtual environment. Some authors include passive virtual environments, such as a movie ride where there is no interaction, in the SE classification, while reserving the term *virtual environment* for active synthetic experiences, where the user interacts with the virtual world.

Tactile: sensory information arising from contact with an object, detected through nerves within the skin. Sometimes restricted to passive information, such as air currents over the skin, in contrast to haptic, which is then applied to active manual exploration. See also *Proprioceptive* and *Kinesthetic*.

Technsplanation: use of VR technology and other communication technology to explain or teach.

Tele-: operating from a remote location.

Tele-existence: virtual reality experienced from remote locations.

Telemanipulation: robotic control of distant objects.

Teleoperation: technology of robotic remote control.

Teleoperation time delay: communication delay necessary to transmit control data to a remote robot or to receive information feedback from the remote location to the operator.

Teleoperator: person doing telemanipulation.

TELEOS(TM): a tool to create Silicon Graphics computer-based real-time interactive environments with *lifelike* deformable objects.

Telepresence: remote control with adequate sensory data to give the illusion of being at that remote location.

Temporal lobe: an area of the brain in front of the occipital cortex and the parietal cortex that is the receiving site for hearing.

Terrain: geographical information and models that can be either randomly generated or based on actual data.

Texture mapping: a bitmap pattern added to an object to increase realism.

Three-dimensional graphics: the presentation of data on a 2D display surface so that it appears to represent a 3D model.

Track shot: rotating display of the same scene. See also *Dolly shot* and *Pan shot*.

Tracker (VR): a device that provides numeric coordinates to identify the current position and/or orientation of an object or user in real space.

Transparency: teleoperation performance measurement (i.e., fidelity of both information feedback and direct remote robot control).

Universe: the collection of all entities and the space they are embedded in for a VR world.

Update rate: tracker's ability to output position and orientation data.

Usability: the effectiveness, intuitiveness, and satisfaction with which specified users can achieve specified goals in particular environments, particularly interactive systems. Effectiveness is the accuracy and completeness of goals achieved in relation to resources expended. Intuitiveness is the learnability and memorability of using a system. Satisfaction is the comfort and acceptability of using a system.

Usability engineering: methodical *engineering* approach to user interface design and evaluation involving practical, systematic approaches to developing requirements, analyzing a usability problem, developing proposed solutions, and testing those solutions.

Usability evaluation: any of a variety of techniques for measuring or comparing the ease of use of a computer system, including usability inspection, user interface critiques, user testing of a wide variety of kinds, safety and stress testing, functional testing, and field testing.

User-centered design: design around the needs and goals of users and with users involved in the design process and design with usability as a primary focus.

User-centered evaluation: approach to evaluating (typically a computer's user interface) that employs representative or actual system users. The evaluation typically involves users performing representative task scenarios in order to reveal how well or poorly the interface supports a user's goals and actions.

User task analysis: process of identifying and decomposing a complete description of tasks, subtasks, and actions required to use a system as well as other resources necessary for user(s) and systems to cooperatively perform tasks.

Vection: illusion of self-motion, usually elicited by viewing a moving image, but also achievable through other sensory modalities (e.g., audition, somesthesia).

Vestibular: sensory structure of the labyrinth of the inner ear that reacts to head and gross bodily movement. The sense organs embedded in the temporal bone on each side of the head are known collectively as the labyrinthine organs or simply the labyrinths. They include the organ of hearing, or cochlea, the three semicircular canals, and the otolith organs, or utricle and saccule.

Vestibular nucleus: a brain stem structure responsible for processing real or apparent motion stimuli and implicated in the maintenance of stable posture and reflexive gaze control. On each side of the brain stem, there are four principal vestibular nuclei: the lateral, medial, superior, and inferior.

Vestibulo-ocular reflex (VOR): when head moves in any direction, the vestibular apparatus senses this movement and sends velocity information directly to the oculomotor system, which responds by driving the eyes (conjunctively) at an approximately equal rate but opposite direction to compensate for head movement and help keep the visual image stabilized on the retina.

Vibratory myesthetic illusions: illusory feeling of limb or body movement that may accompany certain vibrations of the skeletal muscle. For example, vibration of the biceps tendon with the subject's arm fixed and his eyes closed can elicit an illusion of arm extension.

Viewpoints: points from which ray tracing and geometry creation occur. The geometric eye point of the simulation.

Virtual: simulated or artificial.

Virtual environment (VE): model of reality with which a human can interact, getting information from the model by ordinary human senses such as sight, sound, and touch and/or controlling the model using ordinary human actions such as position and/or motion of body parts and voice. Usually *virtual environment* and *virtual reality* are used synonymously, but some authors reserve VE for an artificial environment that the user interacts with.

Virtual heritage: use of computer-based interactive technologies to record, preserve, or recreate artifacts and sites of historic, artistic, religious, and cultural significance and to deliver the results openly to a global audience in such a way as to provide a formative educational experience through electronic manipulations of time and space.

Virtual mechanism: passive artificial modeling structure used in a haptic controller; linked to both a master device (haptic feedback interface) and controlled robot (real or virtual) to allow intuitive and stable (since passive) control and feedback.

Virtual prototype: simulation of an intended design or product to illustrate the characteristics before actual construction. Usually used as an exploratory tool for developers or as a communications prop for persons reviewing proposed designs.

Virtual prototyping: product design that is based on VE technology; provides an alternative concept for the design–test–evaluation cycle by eliminating the fabrication of physical prototypes.

Virtual reality (VR): See *Virtual environment.*

Virtual reality modeling language (VRML): a computer language for creating online virtual reality models.

Virtual reality therapy: any form of psychotherapy using virtual reality. However, the VR is most commonly used to provide desensitization of an anxiety disorder or a phobia by providing a simulation of the situation causing the phobia.

Virtual team members (VTMs): multifunctional autonomous agents or simulated images of humans within a VE that function in the role programmed to them.

Virtual world: entire virtual environment or universe within a given simulation.

Visual suppression of the vestibulo-ocular reflex: use of visual information to suppress the normal eye-beating reflex. An example would be reading text at the same time as undergoing a turning movement while riding on a bus.

Visualization: ability to graphically represent abstract data that would normally appear as text and numbers on a computer.

Voxel: 3D generalization of a pixel, an indivisible small cube that represents a quantum of volume.

Wayfinding: the cognitive element of navigation.

World in the hand: metaphor for visualized tracking where a tracker is held in the hand and is connected to the motion of an object in a display.

Yaw: angular displacement about the vertical axis.

WEBSITES FOR ADDITIONAL INFORMATION ON STANDARDS

American National Standards Institute (ANSI): http://web.ansi.org.

British Standards Institute (BSI): http://www.bsigroup.com.

Human Engineering Design Criteria for Military Systems, Equipment and Facilities (MIL-STD-1472D): http://www.everyspec.com/MIL-STD/MIL-STD-1400-1499/MIL_STD_1472D_1209/.

IEEE Computer Society: http://www.computer.org.

IEEE Standards Association (IEEE-SA): http://standards.ieee.org/. Details of the standards creation process are found at http://standards.ieee.org/resources/index.html.

IEEE Standards Process at a Glance: http://standards.ieee.org/resources/glance.html.

Institute of Electrical and Electronics Engineers (IEEE): http://www.ieee.org.

International Electrotechnical Commission (IEC): http://www.iec.ch/.

International Organization for Standardization (ISO): http://www.iso.ch/.

Italian National Standards Board (Ente Nazionale Italiano di Unificazione, or UNI): http://www.uni.com/.

Joint Technical Committee Number One (JTC1): http://www.iso.org/iso/jtc1_home.html.

Man-Systems Integration Standards Handbook (NASA-STD-3000): http://msis.jsc.nasa.gov/.

ONLINE GLOSSARIES OF VIRTUAL ENVIRONMENTS

http://www.roadtovr.com/virtual-reality-glossary-terminology.

http://www.hitl.washington.edu/scivw/scivw-ftp/other/VR-glossary.

http://inkandvellum.com/blog/2011/04/but-can-you-use-it-in-a-sentence/.

http://www.digitalspace.com/avatars/book/appendix/glossary.htm#glossary.

FURTHER READING

Aukstakainis, S., & Blatner, D. (1992). *Silicon mirage*. Berkeley, CA: Peachpit Press.

Buie, E. (1999). HCI standards: A mixed blessing. *Interactions*, *VI*(2), 36–42.

Delaney, B. (1994). Glossary and acronyms [Special Ed.]. *CyberEdge Journal* [Online]. Available: http://www.cyberedge.com/4al.html

Durlach, N. I., & Mavor, A. S. (Eds.). (1995). *Virtual reality: Scientific and technological challenges*. Washington, DC: National Academy Press.

Hamit, F. (1993). *Virtual reality and the exploration of cyberspace*. Carmel, IN: Sams.
Hix, D., & Hartson, H. R. (1993). *Developing user interfaces: Ensuring usability through product and process*. New York, NY: John Wiley & Sons.
Kalawsky, R. S. (1993). *The science of virtual reality and virtual environments*. Wokingham, England: Addison-Wesley.
Latham, R. (1991). *The dictionary of computer graphics and virtual reality*. New York, NY: Springer-Verlag.
Manetta, C., & Blade, R. (1995). Glossary of virtual reality terminology. *The International Journal of Virtual Reality*, *1*(2), 35–48.
Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25–62). New York, NY: John Wiley & Sons.
Rada, R., & Ketchel, J. (2000). Standardizing the European information society. *Communications of the ACM*, *43*(3), 21–25.
Rheingold, H. (1991). *Virtual reality*. New York, NY: Simon & Schuster.
Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Perspectives of curriculum evaluation* [American Educational Research Association Monograph]. Chicago, IL: Rand McNally.
Sheridan, T. (1992). Defining our terms. *Presence*, *1*(2), 212–274.
Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93. Also available: http://ww.cybertherapy.info/pages/telepresence.pdf
Vince, J. (2004). *Introduction to Virtual Reality*. New York, NY: Springer-Verlag.

This page intentionally left blank

Section II

System Requirements: Hardware

This page intentionally left blank

3 Vision and Virtual Environments

David R. Badcock, Stephen Palmisano, and James G. May

CONTENTS

3.1 INTRODUCTION

This chapter is intended to provide the reader with knowledge of the pertinent aspects of human visual processing that are relevant to virtual simulation of various environments. Before considering how real-world vision is simulated, it is perhaps prudent to review the kinds of information that are usually extracted by the visual system. It would, of course, be fruitless to provide visual detail that is rarely or never available to the senses, and it may be fatal to the endeavor to omit detail that is crucial. Thus, an overview of normal visual capabilities and idiosyncrasies is provided in Section 3.2. Section 3.3 reviews some of the ways that perceptual systems provide shortcuts to simulating the visual world. The existence of these phenomena allows system developers to compensate for hardware shortcomings with user inferences. One of the most exciting advantages of virtual environment (VE) technology is that it allows a more elaborate and complex interaction between the VE and the observer. Section 3.4 reviews a number of the ways in which we interact with the world and how these mechanisms might augment and detract from virtual simulation. After the discussion of what vision entails, a discussion of techniques that use 2-D renditions of the visual world to simulate normal viewing of the 3-D world is provided in Section 3.5. The emphasis here will be to address the design requirements of VE displays and to determine if existing displays are machine or observer limited. Considering what is optimally required by the user, a review of the adequacy of existing visual displays is also provided in Section 3.6. Suggestions are made as to how existing limitations might be overcome and speculations are made concerning what new technology might allow in Section 3.7.

3.2 WHAT ARE THE LIMITS OF NORMAL VISION? (USER REQUIREMENTS)

Research on human visual capacities is an active field of endeavor. There are still many aspects of visual performance that are not well understood, but knowledge has been accumulating at an accelerating pace throughout the last century. The inquiry continues, but it is important to set forth, from time to time, a compendium of what we think we know. The formal study of vision is perhaps one of the oldest disciplines and the community of visual scientists is larger than for any other sense. Since the early psychophysicists initiated the systematic investigation of the senses, the notion has been to map the subjective sensations against the physical scales of interest and express the relationship as a *psychometric function* that defines the limits of sensitivity and the subjective scaling of the dimension. That approach will be employed here, but it is important to acknowledge that, in practice, we may utilize little information near the limits of visual abilities. It is conservative, in a sense, to employ this approach to ensure that display technologies meet these user capabilities, but we may learn to *cheat* on these limits if the task requirements allow.

A comprehensive attempt to list human performance capabilities was collated in 1986 and published as the *Engineering Data Compendium: Human Perception and Performance*

(Boff & Lincoln, 1988; see also Boff, Kaufman, & Thomas, 1986). In this chapter, we will offer more in the way of explanation of the phenomena described, but the *Engineering Data Compendium* is still an excellent starting location for determining performance limits. More recent extensive reviews on specific topics can be found in *The Visual Neurosciences* (Chalupa & Werner, 2004) and in *The Senses: A Comprehensive Reference*, volumes 1 and 2 (http://dx.doi.org/10.1016/B978-012370880-9.09003-4; Masland & Albright, 2008).

3.2.1 Luminance

The human visual system is sensitive to a broad range of ambient illumination, extending from absolute threshold (~ −6 log cd/m^2) to levels where irreversible damage to the system can result (~8 log cd/m^2). Many stimulus factors (e.g., size, wavelength, retinal location) have profound influences on the lower limits of this range (Hood & Finkelstein, 1979). The human visual system contains two types of photoreceptors (rods and cones) with significantly different sensitivities. The lowest absolute threshold is mediated by the more sensitive rods, and the cone threshold is achieved about 2.5 log units above that limit (–3.5 log cd/m^2). At about 1 log cd/m^2, the rod responses begin to saturate, and at higher luminance levels, they provide no significant information because saturation is complete. The cones mediate vision above that point. Thus, the entire range can be broken up into a scotopic region (rods only), a mesopic region (rods and cones), and a photopic region (cones only). Many of the visual abilities reviewed below are different in rod- and cone-mediated vision. They also vary, with luminance level in the ranges subserved by each photoreceptor type. These adjustments take some time and are referred to as light (and dark) adaptation. For example, an observer moving from a sunny street into a dark room may take up to 30 min to reach maximum sensitivity for detecting the presence of a target. However, the reverse process, adapting to an increase in ambient light level, is much more rapid with substantial adjustment in 10–15 s (Hayhoe, Levin, & Koshel, 1992). These adjustments are required to avoid saturation of the neural response and to maintain high sensitivity to contrast changes in a wide range of natural conditions; but it remains the case that sensitivity to luminance contrast is highest when observers are adapted to the ambient illumination level (McCann & Hall, 1980). The same is true for visual acuity (Craik, 1939).

3.2.2 Spatial Abilities

Much of human vision is concerned with discerning differences in luminance and color across space. The extent of the visual field of view (FOV), visual acuity, contrast sensitivity, and spatial position accuracy is reviewed in this section. The display factors that relate directly to providing adequate stimulation for such abilities are discussed later.

3.2.2.1 Visual Fields

The monocular visual FOV is determined by having a subject fixate a point in the center of a viewing area (preferably approximately 180° H × 150° V) and presenting targets throughout the extent of the area. The limits of the field are defined as the most peripheral points at which target detection achieves some criterion (e.g., 75% correct). These limits vary somewhat with the choice of criterion and also as a result of stimulus factors such as the luminance of the background, the luminance of the target, the size of the target, and the wavelength of the target. In normal individuals with a circular target subtending approximately 20 arcmin, the horizontal extent is approximately 160°, and the vertical extent is approximately 120° (see Figure 3.1). The shape is roughly circular but is limited by the physiognomy of the nose, brows, and cheeks. Most fields show a restriction in the area of the lower nasal quadrant. A smaller target, subtending 1 arcmin, may only be detectable within a 20° diameter area (Borish, 1970), and if the target contains spatial structure, visual fields are larger for lower-spatial-frequency targets (Koenderink, Bouman, Bueno de Mesquita, & Slappendel, 1978). The measured field size is also greatly reduced if the observer needs to name the hue of the light,

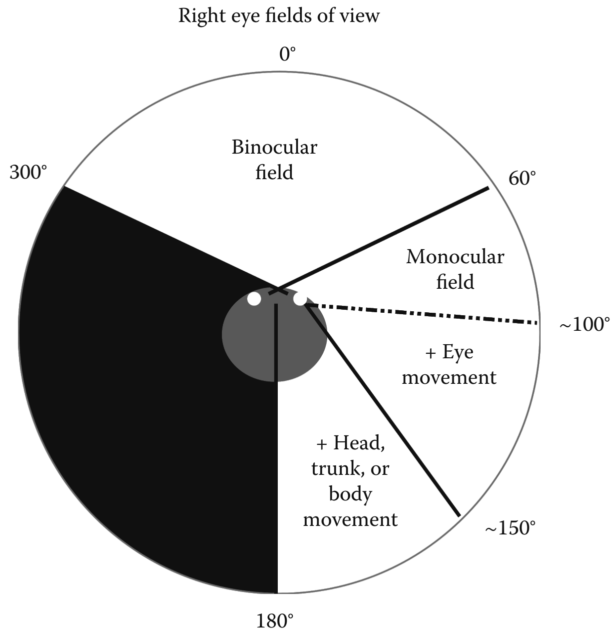

FIGURE 3.1 A depiction of the horizontal monocular and binocular visual fields of the right eye with straight ahead fixation, maximum lateral eye movements, and maximal lateral head movement. The section listing bodily movement can be viewed with maximal eye and head rotation combined, but typically submaximal eye and head movements would be combined with body rotation to view these regions. The black area is not visible by the right eye with clockwise eye and head rotations unless body rotation is included. (The authors thank Professor Paul McGraw and Dr. Andrew Astle for assistance in determining the field limits with eye and head rotations.)

rather than just detecting its presence. This effect varies with the hue to be named. Arakawa (1953) found a 105° horizontal diameter when detecting blue, but only 75° for detecting red and 55° when detecting green. Similar changes occurred vertically. The consequence of this finding suggests that tasks requiring identification of hue should be centered within the visual field.

Careful probing of the area of the visual field about 13°–18° temporal to the fixation point reveals the blind spot, a scotoma formed by the optic nerve head that contains no photoreceptors. It is roughly circular, with a diameter of about 5°. The areas of visibility that are common between the two monocular fields form the binocular visual field. It is roughly circular and is approximately 40° in diameter. Binocular viewing helps to compensate for the blind spot because the visual field location corresponding to the blind spot in one eye coincides with a functional retinal area in the other eye. However, binocular disparity cues for depth (see Section 3.2.3.2) are lost in this region.

3.2.2.2 Visual Acuity

The upper limit of the ability to resolve fine spatial detail is termed visual acuity. It has been measured with numerous techniques (Thomas, 1975), with the normal threshold for spatial resolution ranging between approximately 0.5 and 30 arcsec, depending on the task. The two detection techniques that yield the lowest values are minimum visible and vernier acuity tasks. In the first task, the width of a single line is varied, and the minimally visible width (0.5 arcmin) is determined. In the second, the lateral offset between two vertically oriented line segments is determined (1–2 arcsec). While these are both referred to as acuity tasks, they measure different abilities. The former measures contrast sensitivity (see Section 3.2.2.3) because the change in line width has a greater impact on the retinal luminance difference between the line and background than on the width of the retinal luminance change distribution. This outcome is a result of the line-spread function of the eye. Even in a good optical system, a point of light is spread while passing through the optics. In the human system, this spread covers approximately 1 arcmin. Thus, reducing the width of a small bright line will change the total amount of light passing through the optics, but the spread of light on the retina will change very little. Consequently, a very high acuity for changes can be obtained if those changes create a threshold difference in luminance contrast. Levi, Klein, and Aitsebaomo (1985) used a target composed of five parallel lines in which one of the inner lines was displaced to the left or the right. The displacement varied the width of the gaps on either side of the central line. This produced a contrast difference between the gaps on the two sides of the line, which was detectable with only a 0.5 arcsec line shift—a remarkable spatial discrimination but a normal contrast discrimination.

The vernier acuity task is more appropriately viewed as a position discrimination task. The misalignment of the bars may be detected by discriminating either a difference in the horizontal location of two vertical bars or a difference in the orientation of the overall figure (Watt, Morgan, & Ward, 1983). This task is discussed further later in this chapter.

Resolution tasks require the observer to determine the minimal separation in space of two or more luminance-defined borders. Similar results are obtained with each of two types of commonly employed patterns: gratings and letters. The gratings are high-contrast black lines on a white background with 50% duty cycle, square-wave luminance profiles. The letters are created so that the width of the lines composing the letter is one-fifth of the overall letter size. The acuity measure is calculated from the size of the line width such that the visual angle equals atan (line width/distance between observer and letter).

Since different letters vary in the orientation and position of the lines, it is common to use the Landolt rings instead. The rings have the same line width to overall size ratio but also have a gap the same size as the stroke width of the ring. The observers must resolve the gap and identify its location on the ring. Grating tasks and the Landolt ring task can also be used as discrimination tasks by varying the orientation of the stimuli. These three tasks result in similar thresholds (~30 arcsec in healthy young adult eyes). Numerous variables affect visual acuity performance. Visual acuity increases with luminance, from thresholds of about 1 arcmin at 1 cd/m^2 to 30 arcsec at 100 cd/m^2 (Shlaer, 1937). The recommended standard for illumination for visual acuity measurement is 85 cd/m^2. Under photopic conditions, visual acuity is optimal for targets presented to the center of the visual field (the fovea) and falls off rapidly (see Figure 3.2 for a demonstration). The minimum resolvable detail declines from 30 arcsec at fixation to 20 arcmin at 60° eccentricity for the Landolt ring acuity (Millodot, Johnson, Lamont, & Leibowitz, 1975). At scotopic levels,

D H I G Y R E W A M B C • D H I G Y R E W A M B C

(a)

H G R W M C • H G R W M C

(b)

O S B A W • M H P E K

(c)

FIGURE 3.2 The reader is invited to fixate the small dot in the center of the displays; while maintaining fixation of the dot, take note of how the visibility of the letters declines as more peripheral letters are considered. (a) Shows the decline in a crowded array, (b) demonstrates the larger range available when crowding is reduced, while (c) attempts to compensate for the decline in visual acuity in the periphery by increasing the size of the letters proportionally. There are individual differences in this rate of decline, but the image indicates the approximate scale of the acuity loss with target eccentricity.

when the cones are not functional, acuity is best about 4° from the fixation point. The estimates provided above assume that the targets are presented at the optimal distance. There is a limit to the ability to focus on nearby objects. Targets closer than this near point are blurred and resolution is reduced. The near point recedes with age due to hardening of the lens (presbyopia). This distance may exceed 1–2 m by age 60 and needs to be considered when providing artificial displays because resolution cannot be good if the display is closer than the near point. Optical correction will frequently be needed for older observers.

The identifiability of targets in the visual field is directly influenced by the proximity of surrounding objects. This phenomenon, known as visual crowding, was first reported by Bouma (1970) who noted that the ability to identify a target letter was greatly impaired by adjacent letters. Nandy and Tjan (2012) note that the region in which crowding occurs exhibits several characteristics. First, the size of this crowding zone grows linearly with target eccentricity. Along the axis connecting the fovea to the target, this zone is large, extending to approximately half the target eccentricity. Nandy and Tjan note that this is sometimes referred to as Bouma's law. Second, flanker effects are asymmetric; flankers that are more eccentric than the target have greater crowding effects than flankers at the same distance but on the foveal side of the target. Third, the crowding zone is markedly elongated along the radial, compared to the tangential, axis (see Levi, 2008 and Whitney & Levi, 2011 for reviews). It should also be noted that such crowding effects can impact on the processing of many different stimulus dimensions (van den Berg, Roerdink, & Cornelissen, 2007). Nandy and Tjan have recently presented a physiologically motivated, computational model showing that these critical properties could arise from eye-movement-induced alterations in natural image statistics being reflected in the strength of associations between adjacent cortical receptive fields. For practical purposes, here it is recommended that displays requiring very high acuity should avoid placing objects nearby critical targets.

3.2.2.3 Contrast Sensitivity

While visual acuity has been one of the time-honored descriptors of human spatial ability, it only provides information about the extreme upper limit of the spatial dimension to which humans are sensitive. Much of what humans see and use may not involve spatial detail near these limits. A popular approach to evaluating human sensitivity for objects of different sizes has been to specify the contrast necessary to detect them. In the simplest case, a light bar on a dark background, contrast (C) can be defined as $C = \Delta L/L$, where L is the background luminance and ΔL is luminance increment (or decrement) provided by the bar. Most visual stimuli are more complex, and thus a measure is needed that can provide a sensitivity estimate for these complex patterns. An alternate method takes into account the nature of the neural units early in the visual pathways. At the first stages of processing in the cortex, units are sensitive to short, oriented line segments. Most units give a strong response to a bar of the correct orientation that is surrounded by other bars of the opposite contrast polarity. The underlying weighting function for the cell's response to light can be approximated by a Gabor function that is produced by multiplying a cosinusoidal luminance variation by a Gaussian envelope (Field & Tolhurst, 1986). An image of a Gabor contains a set of bars that alternate between high and low luminance, with contrast that is high in the center of the patch and reducing to zero at the edge of the envelope. When the Gabor is used to represent a weighting function, the bright bars represent excitatory regions and the dark bars inhibitory regions. Individual neurons may vary in the spatial extent of the envelope, the orientation of the cosine, the cosine's spatial frequency, and the cosine's position in the envelope (spatial phase). Since these are the early detectors, one common strategy has been to measure the minimum amount of contrast required for these detectors to reach threshold. Typically, cosinusoidal patterns are employed as stimuli, and with these patterns, contrast is defined as $(L_{MAX} - L_{MIN})/(L_{MAX} + L_{MIN})$, where L_{MAX} and L_{MIN} are the maximum and minimum luminance of the image, respectively. With centrally viewed patches of grating of limited envelope size (e.g., 5° of arc) and a large surround (e.g., 30°) of the same average luminance, the function is band limited from approximately 0.5 to 60 cycles per degree (c/deg).

The peak sensitivity (the reciprocal of the contrast at thresholds of 0.0033 and 0.0025) is approximately 300–400 between 3 and 10 c/deg, but this value varies greatly with display properties.

One of the important motivations for this approach to describing visual sensitivity was the power of the tools of Fourier analysis and now wavelet analysis (Press, Teukolsky, Vetterling, & Flannery, 1992). Using these tools, it is possible to describe any image as a collection of sinusoidal gratings varying in orientation, spatial frequency, phase, and contrast. The similarity between the basis functions employed for this analysis and the weighting functions of some visual neurons led to the suggestion that the visual system may be performing a similar analysis. While this now seems unlikely, it is nevertheless valid to argue that the neurons will only respond to that restricted range of spatial frequencies and orientations within an image to which its receptive field is tuned. Thus, if one could determine the sensitivity of a unit and the amount of contrast in the passband of that unit within an image, it should be possible to predict whether that unit will respond to the particular image. In many cases, such prediction is possible (Campbell & Robson, 1968; Graham, 1980). However, the utility of this approach is limited by the lack of generality of the foveal contrast sensitivity function. In the periphery of the visual field, high spatial frequencies are increasingly poorly resolved, and thus a different contrast sensitivity function is obtained at each eccentricity. Contrast sensitivity declines with luminance (Van Ness & Bouman, 1967) and varies with target motion (Kelly & Burbeck, 1980; Robson, 1966), orientation (Mitchell & Ware, 1974), and chromaticity (Metha, Vingrys, & Badcock, 1993; Mitchell & Ware, 1974; Mullen, 1985). While these limitations are significant, the contrast sensitivity function is a more comprehensive measure of visual sensitivity than spatial acuity alone and is used extensively to characterize the performance of the visual system. The influence of temporal modulation on contrast sensitivity (Kelly & Burbeck, 1980; Robson, 1966) is also relevant when assessing the suitability of different display technologies as discussed below.

3.2.2.4 Spatial Position

The ability to accurately localize image features is an important precursor to object recognition, shape discrimination, and interaction with a cluttered environment. The appreciation of a form presupposes the accurate relative localization of elements of the form, and the perception of peripherally viewed stimuli serves to guide eye movements and locomotion. Approaches to the characterization of such ability in the frontal plane involve relative judgments of the spatial position of two or more objects (Badcock, Hess, & Dobbins, 1996; Westheimer & McKee, 1977), bisection of visual space (Levi et al., 1985), and localization of briefly presented target positions (Solman & May, 1990; Watt, 1987). In central vision, relative spatial position thresholds are on the order of 0.05–1.0 arcsec (see vernier acuity previously mentioned), and once again, the threshold depends on the spatial scale of the pattern. If Gaussian luminance increments or Gabor patches are employed, the threshold is proportional to the standard deviation of the Gaussian envelope (Toet, van-Eekhout, Simons, & Koenderink, 1987). These thresholds can be 10 times higher when targets are presented eccentrically by just 10° (Westheimer, 1982), a decline that is substantially more rapid than that obtained for visual acuity estimates (Levi et al.).

Partitioning studies also indicate that, at eccentricities of 10°, position thresholds vary with the size of the area to be partitioned, increasing from about 0.05° at 0.50° separation to 0.11° at 1.1° separation and remaining constant thereafter (Levi & Klein, 1990). Spatial location difference thresholds for briefly presented, single targets increase with eccentricity from about 0.04° at 2° eccentricity to 3.5° at 12° eccentricity (Solman, Dain, Lim, & May, 1995; Solman & May, 1990). Levi and Klein further clarified that when isoeccentric target element separation approximates target eccentricity, then thresholds are 0.01–0.03 of the eccentricity. It is clear that the relative spatial sense measured with two or more objects, or partitioning, is more precise than is the localization of targets with more limited spatial landmarks (Matin, 1986). Apparent position is also influenced by the spatiotemporal proximity of nearby objects. Badcock and Westheimer (1985) showed that objects closer than 3 arcmin result in an apparent shift of the target toward the distractor's spatial position, whereas larger distances cause repulsion in apparent position (see also Fendick, 1983; Rivest & Cavanagh, 1996). A similar phenomenon occurs over large distances where spatial intervals appear smaller if a larger interval is present nearby and vice versa (Burbeck & Hadden, 1993; Hess & Badcock, 1995).

The apparent location of an object is also greatly affected by motion, with its instantaneous position appearing to be farther along a motion trajectory than its veridical location (DeValois & DeValois, 1991; Edwards & Badcock, 2003; McGraw, Walsh, & Barrett, 2004). This motion position illusion can be induced by either motion in adjacent areas (Whitney, 2002) or motion extracted from the global pooling of local motion signals (even when this global motion differs from dominant local directions [Scarfe & Johnston, 2010]).

3.2.3 Depth

Theoretically, detection of targets in depth is limited by the spatial resolution of the visual system. If the target is large enough (e.g., our moon), an individual can easily see it at a great distance (245,000 miles). However, we will often have difficulty discerning the distance of our moon relative to the sun because they both subtend roughly the same visual angle on our retinas (0.5°) and the binocular disparity between them is too small to reach threshold (Gillam, Palmisano, & Govan, 2011). Their depth order is only unambiguous during an eclipse. At near distances, however, an individual has little difficulty determining the relative distances of two objects because of the presence of numerous depth cues. The cues are traditionally discussed in terms of those available with monocular and binocular viewing.

3.2.3.1 Monocular Cues

Monocular cues are those that would be available to an observer using just one eye.

3.2.3.1.1 Pictorial Cues

Pictorial cues are spatial arrangements that convey relative differences in depth. They are employed by artists adept at conveying 3-D impressions with 2-D depictions. They are the following:

1. Relative size: In the absence of information about the absolute size of an object, its apparent size influences judgments of its distance. The deer on the right-hand side of Figure 3.3 appear to be at increasingly greater distances largely because of their progressively smaller sizes. The moved copies on the left appear larger (upper) and smaller (lower) than the originals (on the right) because of the automatic but inappropriate application of depth scaling.
2. Height relative to the horizon: For objects below the horizon (the identical seagulls indicated by solid arrows in Figure 3.4), objects lower in the visual field appear closer than objects higher in the visual field. For objects above the horizon (the identical clouds in Figure 3.4 indicated by dashed arrows), the reverse is true.
3. Interposition or occlusion: When objects are opaque, those nearby occlude the view of parts of those that are farther away and convey an immediate sense of depth. For example, the boats in Figure 3.4 are interposed between the viewer and the city.
4. Shadows and shading: A directed light source will strike the nearest parts of an object in its path and be prevented from striking other features on the same path. Thus, differences in the intensity of the light reflected from object parts can contribute to the appreciation of depth and 3-D shape (see Figure 3.5a and b). In situations where the position of the light source is not detectable from the image, the visual system resolves ambiguous shading cues by assuming that the light source is above the object (see Figure 3.5c). In addition, objects often cast shadows on surfaces near them, adding additional cues to figure ground configurations by occlusion or interpolation. The degree of occlusion of a shadow

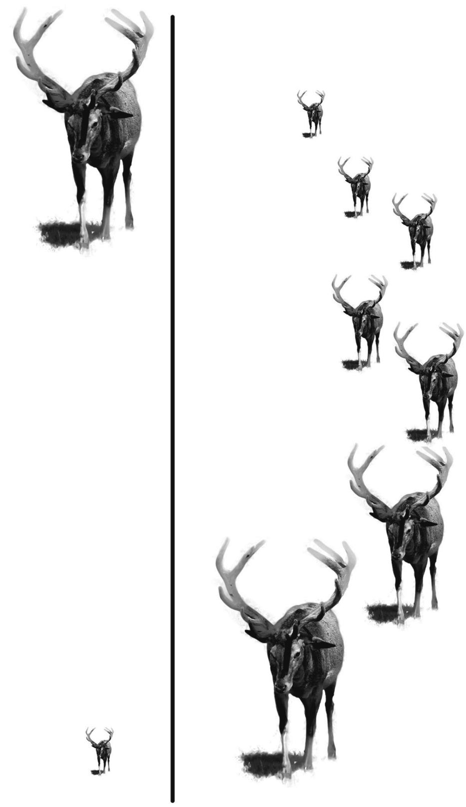

FIGURE 3.3 These images convey differences in apparent depth from decreasing size. The deer on the right are scaled in size so that they vary in apparent distance. The smallest deer appears farther away than the larger ones. The misplaced deer (on the left) appear larger (upper) or smaller (lower) than the identical images on the right (lower and uppermost, respectively) because of inappropriate perceptual depth scaling.

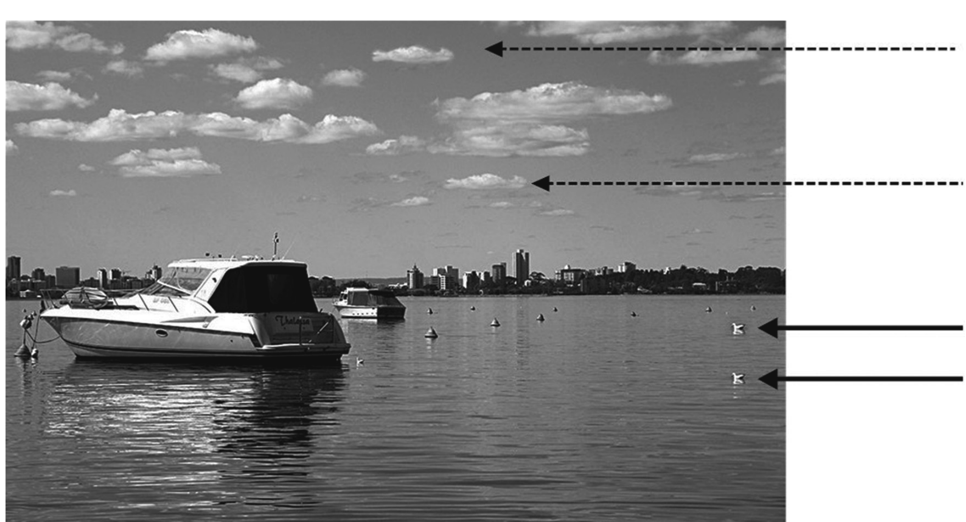

FIGURE 3.4 This scene includes demonstrations of height relative to the horizon and interposition or occlusion cues in determining relative apparent depth (see text for explanation).

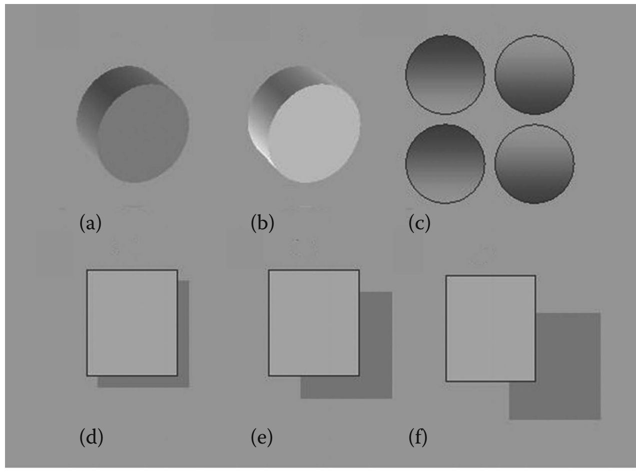

FIGURE 3.5 Brighter objects (b) appear to be somewhat closer than dimmer identical objects (a). (c) Light is assumed to come from above, by default, which means that shadows on the top half of circles produce a percept of concavity while those on the lower half imply convexity. (d–f) show that perceived height of an object, above the picture plane, increases when the object occludes less of its shadow.

by the object casting it can serve as a cue to the distance between an object and the background. These cues are conveyed in 2-D objects by shading (see Figure 3.5d through f).

5. Aerial perspective: The light reflected from objects is both scattered and absorbed by particles in the medium through which it is observed, causing near objects to be perceived as brighter, sharper, and more saturated in color than far objects. Thus, one cue to distance is the brightness of objects (compare objects a and b in Figure 3.5).
6. Linear perspective: When parallel lines recede in distance from an observer (e.g., railroad tracks), their projection on the 2-D retina produces a convergence of the lines as distance increases. The contours do not have to be straight, but merely equally spaced as they become more distant (e.g., a winding roadway). This change in perspective with distance provides a compelling cue for depth and is a consequence of representing 3-D scenes on 2-D surfaces.
7. Texture gradients: Most natural objects are visibly textured. When the image of those objects is projected onto a 2-D surface like the retina, texture elements are distorted and the density of the texture of the surface increases with distance between the surface and the observer. These gradients of texture density are highly salient cues to relative depth. (See Figure 3.6a and c that are warped versions of Figure 3.6b. Figure 3.6b has no texture variation and looks flat. A 180° rotation of Figure 3.6a produces Figure 3.6c. Higher-density regions appear farther away.)

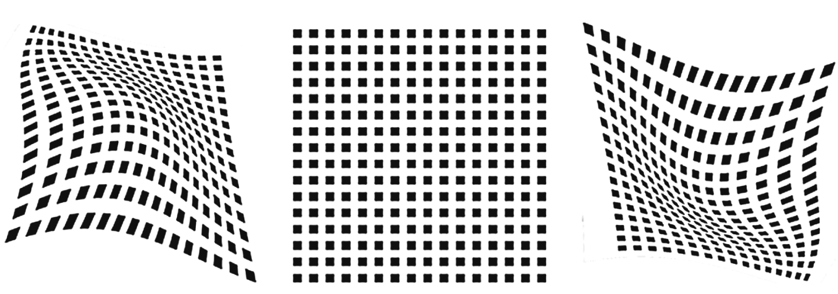

FIGURE 3.6 Texture variation in images can convey a strong impression of relative depth. (b) has no variation and appears flat, whereas the opposite transformations in (a) and (c), which also add a density variation, show that perceived distance increases as density increases.

3.2.3.1.2 Motion Parallax

All of the cues so far discussed have concerned static scenes and a stationary observer, but additional depth information is conveyed when the scene is moving relative to the observer or the observer is moving relative to a static scene. The typical example is an observer on a moving train viewing the landscape as it passes (which generates a pattern of visual motion stimulation known as optic flow). If the observer fixates a point midway between the train and the horizon while the train moves forward, two movement-related phenomena can be observed. First, objects beyond the plane of fixation appear to move in the direction that the observer is moving, while objects closer than the fixation plane move in the direction opposite of the observer's movement. Second, the retinal speed of objects is proportional to the distance of the object from the fixation point. Not only do these phenomena provide important information about the depth of objects in the field, but they also tell us about our movement relative to the environment (Rogers & Graham, 1982; see also Section 3.3.5). A special case of motion parallax, known as the kinetic depth effect, occurs when an individual views a moving 3-D object and the relative motion allows the observer to recover the object's 3-D structure. Imagine a sphere constructed of a meshwork of fine wire. When it is stationary, it may be difficult to appreciate its shape, but when it is rotated, the contours near the observer move in one direction, whereas the contours farther away move in the opposite direction. This shape information can even be recovered from a shadow cast by this sphere onto a 2-D screen.

3.2.3.2 Binocular Cues

Although most people view the world through two eyes, they usually see a single unified view of the world. By viewing a scene alternatively with one eye and then the other, an individual can appreciate the differences in the two views. Ignoring the differences in FOV provided, the individual can also notice slight differences in the relative position of objects within the overlapping regions of the two monocular views. These differences are fused to form a singular view when the individual uses two eyes. The horopter (see Figure 3.7) is a hypothetical surface in space determined by the point in depth for which the eyes are converged (and accommodated). For a given depth of focus, all the points on the horopter are associated with homologous pairs of points on the two retinae. The theoretical horopter defines the curved region in which there is no disparity between the two monocular views. The area in front of the horopter contains disparities that are said to be *crossed* (because it would be necessary to cross the eyes to bring the points onto

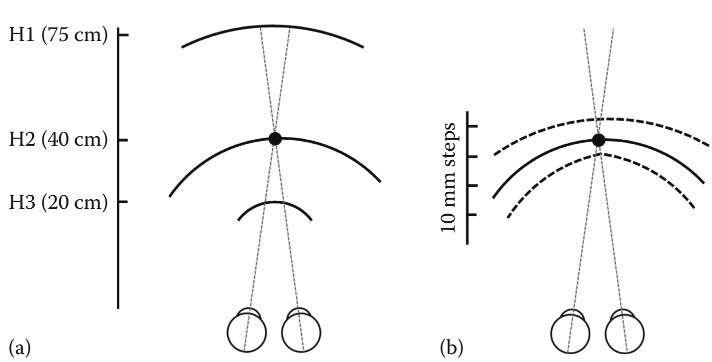

FIGURE 3.7 (a) A depiction of three horopters (H1–H3) associated with fixation in different depth planes. (b) Horopter H2 is replotted with Panum's fusional area, the region of single vision, depicted by dotted lines. (Data replotted from Kalloniatis, M. and Luu, C., Webvision: The organization of the retina and visual system: Perception of depth, 2007. webvision.med.utah.edu/book/part-viii.gabac-receptors/perception-of-depth.)

homologous retinal areas), and the area beyond the horopter contains disparities, which are said to be *uncrossed*. The empirical horopter is shallower in curvature than the theoretical curve, and its curvature decreases with larger viewing distances (see Figure 3.7a). The vertical aspect of the horopter is also tilted, with lower half fields nearer and upper half fields farther away than the point of fixation, a property that may assist when making depth judgments on the ground plane (Schreiber, Hillis, Filippini, Schor, & Banks, 2008).

The region of binocular fusion is a zone approximately centered on the horopter. The reader can appreciate this fact by viewing two objects (e.g., fingers) in different depth planes. If the near object is fixated, attending to the far object reveals double images of that object (diplopia) but a single image of the fixated object. If the far object is fixated, attending to the near object reveals double images of that object (diplopia) but a single image of the fixated object. When fixating 3-D objects, the binocular image contains some areas that fall in front of or behind the horopter and must therefore contain retinal disparities, but these double images are fused and seen as one object. The region in depth over which disparities are fused and singular vision is perceived is termed Panum's fusional area (an example for viewing distance H2 from Figure 3.7a is provided by dotted lines in Figure 3.7b). This area encompasses the horopter and includes areas behind and in front of it.

Studies that have involved dichoptic viewing (presenting separate images to the two eyes) have revealed that subjects will strive to overcome diplopia with vergence eye movements that seek to reduce retinal disparity. If two images can be brought into register with only slight disparity between aspects of the two views, then fusion is achieved and singular vision is experienced. If, however, the disparities are too great, diplopia ensues and binocular rivalry between the two views is often experienced (seeing one view but not the other or seeing part of one view and part of the other). Rivalry can occur at multiple levels in the visual system and can be between the views of different eyes or between different objects, even when part of each object is presented to each eye (Kovacs, Papathomas, Yang, & Feher, 1996; Tong, Meng, & Blake, 2006).

Fusion of disparate images brings with it not only singularity of view but, also more importantly, a vivid sense of depth. For example, imagine viewing dichoptically two separate vertical, square-wave gratings. When fused, they will appear as a single grating in the frontal plane without any depth differences. If the spatial frequency of one of the gratings is shifted slightly lower, the image remains fused but appears to rotate in depth. Retinal disparity is ubiquitously present in our binocular view of the 3-D world, and it may be simulated in 2-D to produce depth information (Howard & Rogers, 1995). The smallest disparity that provides depth information is termed the threshold for stereopsis. This threshold is smallest when the target is viewed foveally and on the horopter (Badcock & Schor, 1985). In the fovea, this threshold may be only a few seconds of arc, but the thresholds increase with the eccentricity of the target to about 300 arcsec at 8°.

In the past, binocular disparity cues were considered useful only for perceiving, and interacting with, very near space (e.g., Arsenault & Ware, 2004; Gregory, 1966). However, recent research conducted in a disused railway tunnel has shown that disparity-based depth is seen well beyond 40 m (Gillam et al., 2011; Palmisano, Gillam, Govan, Allison, & Harris, 2010). With the observers' heads held stationary with complete darkness between the two visible objects in the tunnel, the only cue to depth was binocular disparity. Under these conditions, the estimated depths of objects lying more than 40 m away were still found to increase with their binocular disparity (from 0 to 3 arcmin; corresponding to 0 to 248 m). Interestingly, even when observers were free to move their heads, the contribution of stereopsis to depth appeared to be considerably greater than that of motion parallax.

While stereopsis was long thought to be based simply on binocular disparity cues, this belief has also been challenged by the discovery that stereoscopic depth can still be seen in background regions only visible to one eye (e.g., due to occlusion by a nearer surface). Rather than acting as noise (e.g., by promoting false matches between noncorresponding points in the left and right eyes' images), it is now clear that these types of monocular regions also play important roles in both forming surface representations and binocular depth perception (see Harris & Wilcox, 2009 for a review).

It is also important to recognize that some observers are unable to detect binocular disparity. These stereoanomalous and stereoblind observers, estimated at 3%–5% of the population (Ding & Levi, 2011), with the incidence increasing with age (Rubin et al., 1997), are unable to extract depth signaled purely by disparity differences in modern 3-D display systems and frequently find them unpleasant to view. Studies intending to use 3-D displays should prescreen the observers to ensure their stereoacuity is appropriate for the intended tasks.

3.2.3.3 Accommodation and Vergence

The human visual system contains an elastic lens that can change curvature and refractive index. If an individual looks at an object positioned within about 3 m, the curvature of the lens is adjusted to focus the image sharply on the retina, and the muscular contraction necessary for that adjustment can serve as a cue for the distance of the object. Theoretically, differences in these muscle contractions and differences in sharpness for objects at different distances might also be used as depth cues. With binocular viewing, the two eyes converge when viewing near objects and diverge when viewing objects farther away. Observers are able to use vergence information to assist in judging depth for near objects (Viguier, Clement, & Trotter, 2001), but it is unclear whether the critical information arises from the extrinsic eye muscle tension conveying information about eye position to the brain or, instead, the efferent command to converge the eyes. Accommodation and vergence are normally linked and covary with the depth plane of the target observed. While some doubt the utility of the potential physiological cues to depth, the consensus is that cues associated with accommodation and vergence are viable for near targets (Gillam, 1995) and these variables may be quite important for virtual simulations where depth may be produced by binocular disparity cues presented on screens that are near the eyes. Under these circumstances, the necessary accommodation and convergence may not be congruent with normal visual experience. The consequences of this are discussed in Wann, White, Wilkie, Culmer, Lodge, & Mon-Williams (2014; Chapter 32) and later in this chapter.

3.2.4 Color Vision

In addition to our ability to discern differences in luminance across space and time, humans may also discriminate between the wavelength of light across these dimensions. The history of research on color vision is, perhaps, the most extensive consideration of any human ability and our understanding of the mechanisms by which humans appreciate color differences reflects that impressive effort. The concern here, however, is to define the limits of that human ability and not to explain the mechanisms underlying them. Humans are sensitive to electromagnetic radiation in the range of 370–730 nm (Wandell, 1995). Various light sources (e.g., the sun, tungsten bulbs, fluorescent lights) provide a broad band of radiation across and beyond this visible spectrum. Traditional methods of studying color vision have involved various methods of restricting the spectrum to narrow bandwidths. Transparent devices (e.g., filters, diffraction gratings, prisms) are characterized in terms of spectral transmittance, while opaque media (e.g., paper, paint) are described in terms of their spectral reflectance.

The first question addressed concerning sensitivity to chromaticity involved determining the threshold for detecting the presence of a light composed of only a narrow range of wavelengths. Measures at scotopic levels (rod-mediated vision) revealed that the sensitivity varied with a peak at about 500 nm (the Commission Internationale de l'Eclairage, CIE, $V'(\lambda)$ function; see upper panel of Figure 3.8). At photopic levels, a similar function was observed, with a peak at 555 nm (the CIE $V(\lambda)$ function; see upper panel of Figure 3.8). At mesopic levels, the measured functions are a combination of the other two. It must be noted that these curves only reflect the group-averaged sensitivity to the presence of light of different wavelengths. They do not address the ability to identify or discriminate between colors. Color vision is defined as the ability to discriminate between stimuli of equal brightness on the basis of wavelength alone. Under scotopic conditions, this is not possible as all rods have the same spectral sensitivity, but under photopic conditions (where the three different cone types have different spectral sensitivity), the full extent of color vision is measurable.

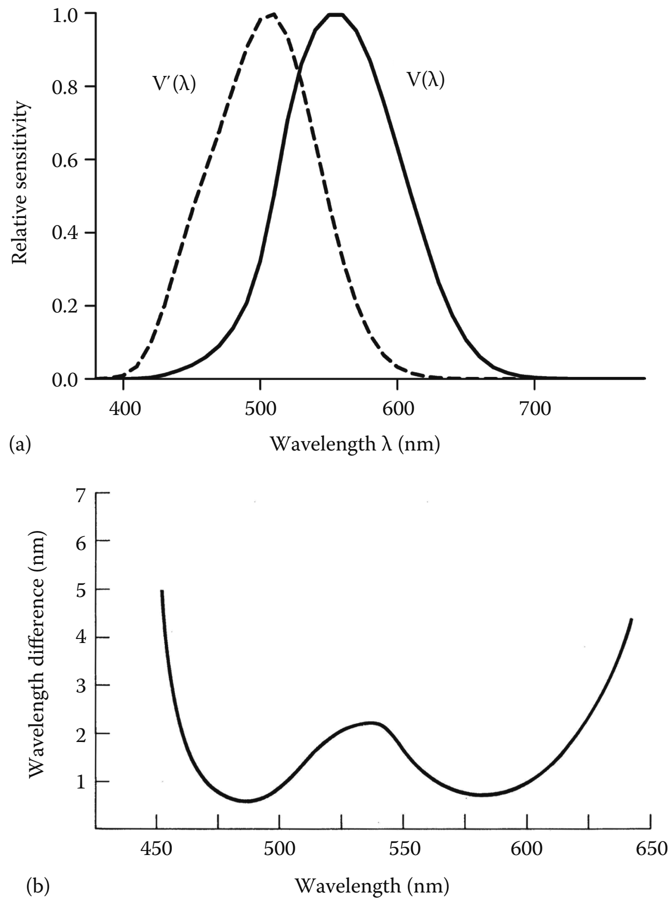

FIGURE 3.8 (a) The upper panel contains the function depicting the spectral sensitivity of the human eye under scotopic [CIE V′(λ)] and photopic [CIE V(λ)] conditions. The two functions have been normalized to remove the overall increased sensitivity of the scotopic system. (b) The lower panel contains the function that describes wavelength discrimination across the spectrum. (From Hurvich, L.M., *Color Vision*, Sunderlund: Sinauer, 1981, p. 210.)

Wavelength discrimination tasks involve asking the observer to vary the wavelength of a comparison stimulus until it can be differentiated from a standard that is of some given wavelength. The increment in wavelength necessary to discern a difference when the stimuli are of equal brightness is plotted against wavelength (see lower panel of Figure 3.8). The wavelength difference varies between 1 and 13 nm, and the function contains two minima in the ranges of 450–500 and 550–600 nm. This kind of discrimination varies with a number of parameters. Chromatic discrimination ability decreases with luminance, especially at lower wavelengths (Brown, 1951; Clarke, 1967; Farnsworth, 1955; Siegal & Siegal, 1972), but remains fairly constant above 100 td. (Trolands are a measure of retinal illuminance that incorporate the effects of changes in pupil diameter with luminance level: I[Trolands] = A [pupil area in mm^2] · L[luminance in cd/m^2].) Discrimination is reduced with field size below 10° (Brown; Wyszecki & Stiles, 1967) but is relatively unaffected by separation between comparison fields (Boynton, Hayhoe, & MacLeod, 1977). If comparison fields are presented successively, discrimination is unaffected until a stimulus onset asynchrony (SOA) of 60 ms but declines at higher intervals (Uchikawa & Ikeda, 1981).

Discrimination for some wavelengths (400 and 580 nm) asymptotes at 100 ms, whereas other wavelengths (480 and 527 nm) require 200 ms for best performance (Regan & Tyler, 1971). Chromatic discrimination is also poor at eccentricities beyond about 8°.

In addition to discrimination of hue, human observers can discern differences in the purity of white light. The threshold for colorimetric purity is defined as the amount of chromatic light that is added to white light to produce a just noticeable difference (JND). Additional JNDs can be observed with increasing steps in the chromatic additive. The number of JNDs observed varies with the wavelength of the additive and is least at about 570 nm. The term *saturation* is used to describe the subjective correlate of colorimetric purity. A highly saturated light appears to have little white light *contamination*.

If the JND is taken as the step size in the dimensions of hue, brightness, and saturation, Gouras (1991) has suggested that there are about 200 JNDs for hue, 500 JNDs for brightness, and 20 JNDs for saturation. This suggests that human color vision capability involves the discrimination of about two million chromatic stimuli ($200 \times 500 \times 20 = 2{,}000{,}000$).

The visual subsystem that processes chromatic information differs in a number of significant ways from the subsystem that encodes luminance variation. Critically, for artificial systems where bandwidth limitations are significant, the chromatic system has substantially poorer spatial (Mullen, 1985) and temporal (Cropper & Badcock, 1994) resolutions, and thus chromatic properties can be rendered more coarsely in both space and time. Mullen estimates that the spatial resolution for red-green gratings is three times worse than for achromatic gratings (10–12 c/deg instead of 60 c/deg), whereas Cropper and Badcock reported a twofold reduction in temporal resolution.

This section has dealt with the detectability of light as a function of wavelength and also the ability to discriminate between lights of differing wavelengths, but the hue perceived is greatly affected by the chromaticity of the surround.

3.2.5 Motion

Humans can see objects moving with respect to themselves, whether they are stationary or moving, and they can detect their own movements through space in a static environment or with objects moving around them. Information about one object's movement relative to another (exocentric motion) is obtained from the image of the objects moving across the retina. An observer may attribute motion to the object that is actually moving or may see movement of a stationary object if the frame around it is displaced (induced movement). Information about object movement relative to an observer's own position or movement (egocentric motion) comes from translations in the retinal image in relation to nonvisual senses (e.g., vestibular and kinesthetic senses) that are involved in body, head, and eye movements. Mere translation of the object on the retina cannot provide information about egocentric information. Humans do not perceive object motion relative to themselves if they move their head or eyes while looking at or away from a stationary object (position constancy).

There are several ways in which one might measure a minimum motion threshold. First, one could measure the minimum distance a feature has to be displaced in order for the direction of motion to be detected. When a visual reference is present, displacements as small as 10 arcsec are sufficient for this judgment (Westheimer, 1978). Second, one could measure the minimum temporal frequency required to detect movement in a temporally extended motion sequence. This threshold depends on both the contrast and the spatial structure of the moving pattern (Derrington & Badcock, 1985). The minimum temporal frequency decreases as contrast increases for low-spatial-frequency periodic sinusoidal grating patterns but is constant for high-spatial-frequency patterns. Third, one could measure the minimum number of dots needed to move in a common direction in a field of randomly moving dots for that direction to be discernible. Edwards and Badcock (1993) have found that as few as 5%–10% of dots moving in a common direction is sufficient either for frontoparallel motion or for expanding/contracting patterns designed to mimic motion in depth, although this threshold does vary with dot step size

(Baker, Hess, & Zihl, 1991). It is also possible to combine signals from dispersed local apertures containing motion into a global rigid solution that may differ from all of the local components (Amano, Edwards, Badcock, & Nishida, 2009). In this latter case, the minimum change in perceived direction or the minimum number of apertures required to be consistent with the rigid solution may be varied, but sensitivity using both measures is high.

While these are impressive abilities, it is also the case that the perception of motion is influenced by a number of stimulus properties. The perceived speed of an object slows if the luminance contrast of the object is reduced (Cavanagh & Favreau, 1985; Thompson, 1982), and observers are unable to see very-high-spatial-frequency repetitive patterns move at all (Badcock & Derrington, 1985). Objects appear to move more slowly through smaller apertures and at greater distances (Rock, 1975).

Finally, even relatively brief periods of exposure to continuous motion (e.g., 30 s to 1 min) produce substantial motion aftereffects, where subsequently viewed stationary patterns appear to drift in the opposite direction to the previously seen physical motion (Mather, Verstraten, & Anstis, 1998). In computer environments, these aftereffects commonly arise due to smooth scrolling of stimulus displays, which can also produce compelling visual illusions of self-motion (see later discussion and Hettinger, 2014; Chapter 18). A common example is observed when trains pull into a station. The continuous visual adaptation to forward motion creates the egocentric impression of rolling backward when the train is stationary.

3.2.6 Motion in Depth

Regan and Beverly have developed the notion that motion in depth is mediated by at least two mechanisms: changing size (Regan & Beverly, 1978a, 1978b) and changing retinal disparity (Regan & Beverly, 1973). Changing size is a monocular mechanism and is characterized by thresholds of about 1 min of arc. The velocity of the size change is proportional to the perceived motion in depth. Changing retinal disparity and inter-ocular velocity differences are stereoscopic cues with approximately the same thresholds for detection as changing size. Figure 3.9 summarizes the movement on the retina caused by monocular and binocular viewing conditions. In Figure 3.9a, the motion of an object toward the viewer results in the edges of the object moving in opposite directions (expansion) on the retina (and vice versa for movement away from the observer, not illustrated). In Figure 3.9b, the motion toward the observer results in both object expansion and

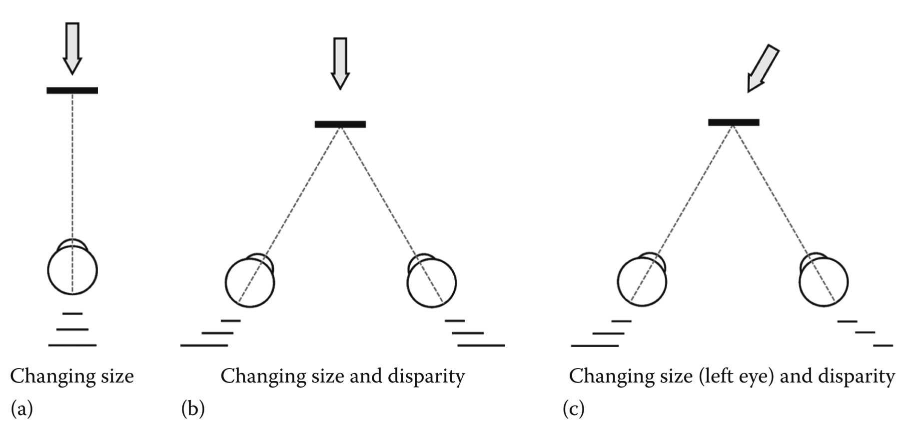

FIGURE 3.9 (a) Indicates how the edges of an object on a single retina move in opposite directions under expansion and contraction conditions produced by an object moving toward or away from the observer. (b) and (c) Indicate how opposite object motion occurs on the two retinas when it is stereoscopically viewed (see text for elaboration).

movement of the edges of the object in opposite directions (translation) on the two retinas. In Figure 3.9c, the motion of an object toward the left eye will result in only expansion in that eye but translation and less expansion in the right eye. In addition to the differences in direction of movement in the binocular case, differences in velocity occur on the retinas when the object is not moving on a trajectory aimed midway between the eyes. Comparisons of monocular and binocular sensitivities to motion in depth reveal the binocular sense to be slightly more sensitive than the monocular case, as in spatial vision (McKee & Levi, 1987). Both mechanisms appear to provide input to a single motion in depth stage, however, because motion perceived in terms of changing size can be *nulled* with conflicting disparity manipulations. This information concerning motion in depth appears to be uniquely derived because this ability cannot be accounted for by static stereoacuity performance and the visual fields for stereomotion in depth are constricted relative to those observed for static stereopsis. However, receptive fields for motion in depth signaled by optic flow without stereo cues are very large compared to those for stereopsis. Burr, Morrone, and Vaina (1998) have shown signal integration with stimuli up to 70° in diameter in psychophysical experiments, and Duffy and Wurtz (1991) have reported single cells with fields up to 100° in diameter in the medial temporal (MT) cortical area. These mechanisms seem ideal for detecting the optic flow produced by locomotion and appear to be centrally involved in its control. Interestingly though, Edwards and Badcock (1993) have shown that observers are more sensitive to contraction than expansion, which is the kind of flow produced by walking backward.

3.3 WHAT DO WE INFER FROM WHAT WE SEE? (SHORTCUTS)

Previously, the various visual capabilities with regard to detection and discrimination along a number of stimulus dimensions were enumerated. That approach to visual perception assumes a relatively passive observer and defines the abilities in terms of the physical limitations of the visual system. But the human observer is rarely passive and usually interacts with the environment with expectations, goals, and purposes. These psychological aspects of perception have been the focus of numerous lines of research and have provided insight into various phenomena that serve to facilitate and disambiguate information-processing tasks. In this section, we consider these propensities and the implications for perception within the simulated environment.

3.3.1 Figure/Ground Organization and Segmentation

Glance at Figure 3.10a and ask yourself what you see first. You might see lots of Ss or Hs made of Ss or an E made of Hs made of Ss. Whatever it was that you perceived first, it is easily possible to see the other possibilities with further examination. While humans rarely think about it, similar perceptual processes are employed in the analysis of all images. An observer might stare at a forest and see its global outline against the sky or look at a particular tree as an element in the forest or attend to a leaf on one of its branches. In each case, the observer can describe the object of their attention as an object perceived against a background (or ground). Since nothing need change in the scene in order for the observer to experience these different perceptual figure–ground organizations, they are forced to conclude that the experiences are not driven by stimulus change but are determined either physiologically (through adaptation) or psychologically, through directed attention and/or perceptual set. Consider Figure 3.10b, the Necker cube. Although the figure is static, an observer can perceive two possible perceptual organizations involving different depth planes, one with the box protruding to the upper right and the other with the box protruding to the lower left. Much of the visual work accomplished is cognitively driven wherein the observer searches for objects or serially processes objects (e.g., reading) against a background. Early Gestalt psychologists (e.g., Koffka, 1922) proposed a set of *laws* that suggest what stimulus configurations facilitate perceptual organization. Recent work has been attempting to determine the algorithmic and physiological bases for

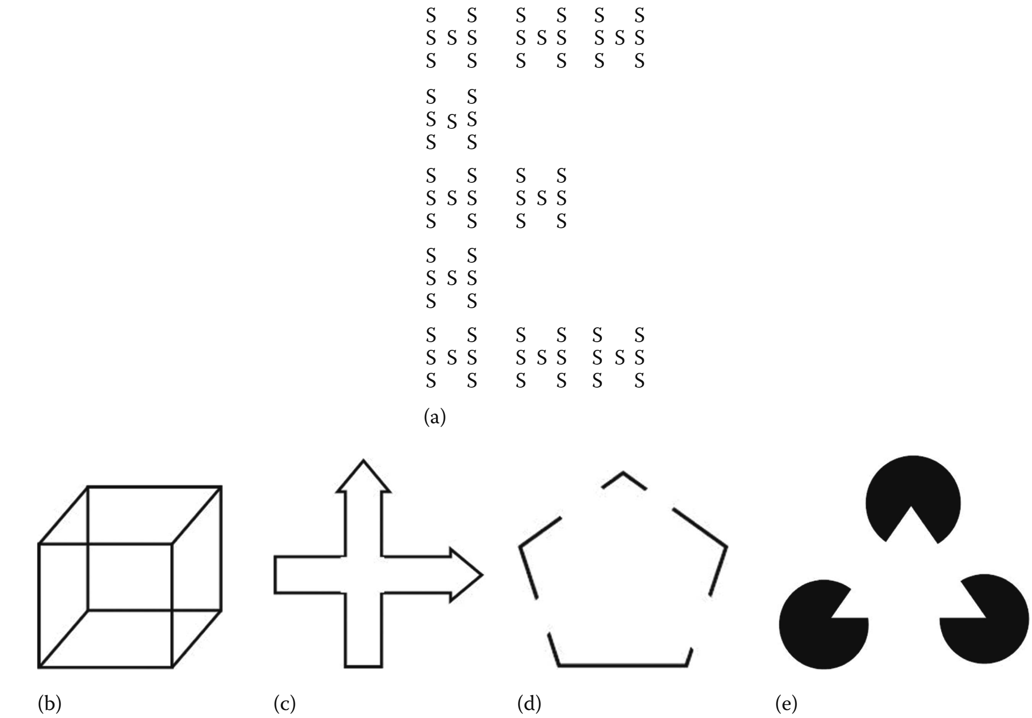

FIGURE 3.10 Stimuli that provide examples of perceptual organization (see text for details). (a) A large letter E composed of mid-level clusters (H), composed of small letters (S). (b) A Necker cube that alternates in 3D interpretation. (c) A compound image interpreted as two arrows crossing because of good continuation. (d) A set of lines form a pentagon through closure. (e) An illusory Kanizsa triangle appears to occlude three circles.

some of these phenomena. They are still poorly understood but have proven to be useful for scene segmentation in computational models. The factors are the following:

1. Proximity: Given an array of elements, those closest together are more likely to be grouped into a global structure. Elements that are close to one another tend to be seen as an object (e.g., the Ss are grouped to form the *H* in Figure 3.10a).
2. Similarity: Similar objects tend to be grouped together to form a global structure (e.g., the *E* in Figure 3.10a).
3. Good continuation: Lines are perceived as continuing with the minimum angular deviation necessary. In Figure 3.10c, the vertical lines of the vertical arrow could deviate sharply along the horizontal arrow but the default grouping avoids this solution. Field, Hayes, and Hess (1993) have described an association field that quantifies the likelihood of binding elements on a smooth path as a function of angular deviation of the path, and the required neural connectivity has been demonstrated in primary visual cortex (Gilbert, Das, Ito, Kapadia, & Westheimer, 1996).
4. Common fate: Elements of an image that undergo a common transformation over time are frequently grouped as one object. The most extensively studied transformation is lateral motion, which produces rapid segmentation of complex scenes. The precise mechanisms are yet to be revealed, but single cells in the MT cortical area have stimulus preferences for dot patterns moving in different directions to their surroundings and thus may form a neural substrate for this segmentation (Allman, Miezin, & McGuinness, 1985). It should be noted that while this segmentation is automatic with moving stimuli, coherent grouping of local texture elements is subject to strong attentional influences (Dickinson, Broderick, & Badcock, 2009).

5. Closure: When a form is drawn with broken or incomplete contours, the perceptual system perceives the form that would result if the contours were complete (Rock, 1975). This is not to say that the breaks in the contours are not detectable, but rather that those breaks do not prevent grouping for object recognition (see Figure 3.10d).

These principles may be applied to images containing a broad array of stimulus attributes. Interestingly, once grouping has occurred, quite distinct boundaries may be formed that are not physically present in the luminance profile of the image. For example, Figure 3.10e appears to be interpreted by the visual system as a white triangle overlapping black circles. An illusory brightness difference is usually perceived between the inferred center of the triangle and the surrounding area. The border between these two zones can appear to be very sharp. Some of these borders may be seen because they stimulate nonlinear texture detection processes within the visual system (Wilson, Ferrera, & Yo, 1992), but no adequate model exists for the full array of illusory contours at this stage. However, there has been substantial progress in understanding how local edge fragments are grouped to define contours (for review, see Loffler, 2008) with demonstrations of explicit global integration processes for both contour (Bell & Badcock, 2008; Loffler, 2008) and texture (Wilson & Wilkinson, 1998), supporting physiological processes in the cortical ventral stream (Connor, 2004) and more detailed computational models (Cadieu et al., 2007).

3.3.2 Perceptual Constancies

Although the 2-D representation of objects on the retina may undergo considerable translations in lightness, color, orientation, size, and position, humans maintain an appreciation for these basic properties of the object. Thus, a white piece of paper (on a gray background) viewed in direct light and in a cast shadow appears to be the same brightness despite the fact that one reflects more light than the other. Likewise, a piece of colored paper (on a gray background) viewed under different colored lights will appear a very similar color despite the differences in hue of the illuminant (although changing the color of just the surround can dramatically alter the perceived hue of the target patch). This page is viewed as rectangular whether viewed from directly above or from a 45° angle, and it is perceived to be the same size whether viewed from 18 in. or 18 ft. The page is seen to be in the same position in space whether an observer looks directly at it or gazes to the side. While the mechanisms whereby these constancies are achieved are still the subject of active research, most explanations emphasize a ratio principle. For lightness and color constancy, it is proposed that the ratio of the lightness or the hue of both the object and the background is extracted, and if the ratio is the same, constancy is achieved. For shape and size constancy, the critical ratio refers to the size and orientation of elements in the object relative to size and orientation in the background under different viewing conditions. See Palmer (1999) for a review of these mechanisms. Position constancy will be elaborated upon later.

3.3.3 Exocentric Motion

Consider a small, stationary point of light in a totally dark room. That object appears to move in a quasirandom fashion, and subjects find it quite difficult to specify its position (by pointing or touching its location). This illusion of exocentric motion is termed the autokinetic effect (Sandström, 1951), and it indicates how difficult it is to determine the motion of an object without a visual frame of reference. The effect is actually caused by implicit eye, head, and body movements in the absence of the stabilizing effect of visual information and is really an egocentric illusion. In a lighted room with referential stimuli, the object is seen as stationary (Mack, 1986). The next question concerns how much that object must be displaced for the observer to detect movement. With a frame of reference, the displacement threshold has been estimated to be 1–2 min, but that value increased tenfold when the reference was removed (Aubert, 1886; Bourdon, 1902; Grim, 1911). With real movement, the moving object is always visible. It moves a given distance (or is visible for a given time), and its movement generally has a fixed velocity (in the simple case). The upper threshold for the detection of movement depends on many factors (luminance, FOV, spatial pattern), but long before an object moves too fast to be seen, the form of the object is obscured. One estimate of this threshold point is termed dynamic visual acuity, and it indicates that form vision as measured by the Landolt C acuity is compromised above about 60°–100°/s (Ludvigh & Miller, 1958). However, very large objects may be detectable up to 3000°/s provided the temporal frequency of the stimuli falls inside the critical flicker-fusion limit (Burr & Ross, 1982). There is a belief that, at moderate velocities, displacements are perceived as movement, but at lower velocities, motion is based on detecting changes in position (Anstis, 1980; Exner, 1875). Some suggest that with brief durations (<0.25 s), only velocity is sensed and only at long durations would velocity thresholds be improved by a framework (Gregory, 1964; Liebowitz, 1955). Velocity thresholds increase in the periphery (Aubert, 1886). There is also a common misconception that the peripheral retina is better at motion detection than the central retina. On the contrary, Tynan and Sekuler (1982) reported minimum movement thresholds to be on the order of 0.03°/s in the fovea, increasing to 0.45°/s at 30° eccentricity, and they become progressively poorer toward the outer limits of the visual field (To, Regan, Wood, & Mollon, 2011).

When two objects are moved in the visual field, the difference threshold for velocity has been estimated to be in the range of 30 arcsec/s (Graham, Baker, Hecht, & Lloyd, 1948) to 2 min/s (Aubert, 1886). But with two objects, induced movement and motion contrast effects come into play. Induced movement is an illusion of motion in which the displacement of some elements results in the attribution of movement in other elements. For example, moving a frame around a stationary object often results in an apparent movement of the object in the direction opposite to frame movement. Motion contrast is a type of induced movement that occurs at the borders of adjacent fields of moving elements. Adjacent elements moving in different directions and/or at different velocities can induce illusory directional and velocity attributes (Levi & Schor, 1984; Whitney, 2002).

With the advent of modern display systems and greater computational power, these findings have now been extended to stimuli covering larger areas of the visual field and more natural scenes. The principles remain, but for comprehensive reviews of recent work, see Nishida (2011) and Burr and Thompson (2011).

3.3.4 Egocentric Motion

In most cases, motion in the visual field due to one's own movement is easily discriminated from motion in the visual field due to object motion. The mechanisms that underlie this ability involve a comparison between the movement sensed on the retina and nonvisual signals associated with one's own movements. When movement of an object on the retina is negatively correlated with eye, head, or body movement (later discussed), position constancy holds and objects are seen as stationary. Any mismatch between the visual and nonvisual information signals movement of the object relative to the observer. The limits of position constancy define the precision of these mechanisms. For head movements, position constancy breaks down with displacement ratios (retinal movement/head movement) of 1.5% (Wallach & Kravitz, 1965). While humans are quite accurate at detecting displacements during fixation (2–3 min), displacements as large as 6 min go undetected if they occur during a saccadic eye movement (Stark, Kong, Schwartz, Hendry, & Bridgeman, 1976). Target displacements of less than 1/3 saccadic displacements are rarely seen (Bridgeman, Hendry, & Stark, 1975). Indeed recent work has shown very substantial compression of perceptual space in a time window from approximately 50 ms prior to a saccade to 50–100 ms after the saccade (Burr & Morrone, 2005). Thus, the resolution of the position constancy mechanism during saccades is degraded relative to fixation. When a moving target is tracked with smooth-pursuit eye movements, its perceived speed has been reported to be 63% of that reported when it is moved through the field when the eyes are fixated (Dichgans, Koener, & Voigt, 1969; Turano & Heidenreich, 1999). In addition, stationary objects in the field appear to move (the Filehne illusion) when the eyes track another moving object (Filehne, 1922). Wertheim (1994) argues that reference signals are generated through visual, oculomotor, and vestibular interactions, and it is the comparison of the retinal and reference signal that mediates the perception of self and object motion. He suggests that while oculomotor-generated signals encode how the eyes move in orbits, vestibular signals must be added to form these reference signals to encode how the eyes move in space, and the neural mechanisms supporting this have now been detected in human parietoinsular vestibular cortex (Cardin & Smith, 2010).

3.3.5 Visually Induced Self-Motion

As we move through the real world, either actively (such as walking) or passively (such as sitting on a moving train), multiple sensory systems are responsible for detecting, processing, and perceiving our self-motion. While nonvisual inputs (from the vestibular system of the inner ear, kinesthesia, somatosensation, and even audition—see Lawson et al. [2014; Chapter 7] and Hettinger [2014; Chapter 18]) often contribute significantly to the conscious perception of this self-motion, visual input appears to play a particularly important role. This is clearly demonstrated by the occurrence of visually induced illusions of self-motion, known as vection (Fischer & Kornmüller, 1930). As far back as the time of Helmholtz (1867/1925), it was known that compelling vection could be induced by viewing a quickly moving river from a bridge. Passengers sitting on a stationary train often experience a similar illusion of self-motion when the train on the next track pulls out from the station (Mach, 1875). This vection is a potentially useful tool for enhancing the user experience in many virtual reality applications, such as vehicle simulation, telepresence, and architecture walkthroughs.

Early laboratory research confirmed that large optic flow patterns can generate robust translational and rotational illusions of self-motion in physically stationary observers (Brandt, Dichgans, & Koenig, 1973; Johansson, 1977; Mach, 1875; Tschermak, 1931). In one study, Mach (1922) presented subjects with a large, endless belt (covered with an alternating black-and-white stripe pattern) moving horizontally across two rollers. This induced linear vection (LV) in the opposite direction to the belt's motion. In another study, Mach (1875) placed subjects inside a large rotating drum and found that this induced circular vection (CV) in the opposite direction to the drum's motion. Later research showed that this CV can be subjectively indistinguishable from real self-rotation (e.g., produced by rotating the subject's chair—Dichgans & Brandt, 1978). In fact, tilting one's head during CV produces powerful *pseudo*-Coriolis effects (which include illusions of apparent tilt, dizziness, disorientation, drowsiness, and nausea—Brandt, Wist, & Dichgans, 1971). Interestingly, these symptoms occur even though the complex (Coriolis cross-coupled) vestibular stimulation typically thought to be responsible for them is absent.

Vection is generally perceived in the opposite direction to the motion of the visual background (Ito & Shibata, 2005), except in relatively rare cases of inverted vection where foreground motion dominates the experience (Nakamura & Shimojo, 2000). Typically, there is a 3–12 s delay between the initial exposure to optic flow and vection onset, during which the observer perceives object/scene motion (Brandt et al., 1973; Bubka, Bonato, & Palmisano, 2008; Telford & Frost, 1993). This delay was thought to reflect the time required for the stationary observer to resolve the following sensory conflict: his/her visual input indicates self-motion, but the expected nonvisual inputs for this self-motion are absent (Johnson, Sunahara, & Landolt, 1999; Zacharias & Young, 1981). According to this type of explanation, LV and CV should be reduced/enhanced by maximizing/minimizing this sensory conflict, respectively. However, upright observers still perceive compelling head-over-heels CV when placed inside a fully furnished tumbling room, despite the extreme conflicts between their visual and nonvisual inputs (with the former indicating head-over-heels self-rotation and the latter indicating that the observer is physically stationary—Allison, Howard, & Zacher, 1999; Palmisano, Allison, & Howard, 2006). Another challenge to this explanation has been posed by findings that self-motion displays containing simulated viewpoint jitter/oscillation (thought to generate significant and sustained sensory conflicts) actually induce vection more quickly and intensely than comparable nonjittering self-motion displays (thought to generate only minimal/transient sensory conflict—see Kim, Palmisano, & Bonato, 2012; Palmisano, Allison, Kim, & Bonato, 2011; Palmisano, Gillam, & Blackburn, 2000; Palmisano et al., 2008).

Over the years, parametric research has provided insights into how self-motion might be best simulated with purely visual displays. These studies have shown that LV and CV will generally be more compelling when the optic flow (1) has a faster speed of translation/rotation (up to an optimal velocity; e.g., yaw CV speed is proportional to the stimulus speed up to 90°/s—Allison et al., 1999; Brandt et al., 1973), (2) has more and larger moving elements (e.g., Reason et al., 1982), (3) stimulates a larger retinal area (e.g., Held et al., 1975), and (4) is perceived to be from the background (as opposed to the foreground—Nakamura & Shimojo, 1999; Ohmi & Howard, 1988; Ohmi, Howard, & Landolt, 1987; Seno, Ito, & Sunaga, 2009).

While large fields of view and peripheral motion stimulation are often desirable when simulating self-motion scenarios, they are not necessary to induce vection; for example, Andersen and Braunstein (1985) showed that LV can be induced with as little as 7.5° of central motion stimulation. Both central and peripheral motion stimulation can induce compelling vection (Post, 1988). However, the optimal stimulus spatial frequency for vection appears to depend on the retinal eccentricity of the motion stimulation. Palmisano and Gillam (1998) found that CV was more compelling when higher-spatial-frequency motion stimulation was presented to central vision (where acuity is higher) and low-spatial-frequency motion stimulation was presented peripherally (where the eye's spatial resolution is also lower). Such findings relax the requirement for high-resolution imagery in peripheral display regions (unless the simulation requires the user to look/attend to these regions).

Stereoscopic presentation of the vection-inducing stimulus has also been shown to facilitate CV and LV (Lowther & Ware, 1996; Palmisano, 1996). Consistent stereoscopic cues not only reduce vection onsets and increase vection durations, but they also increase vection speeds and the perceived distances traveled (Palmisano, 2002; Zikovitz, Jenkin, & Harris, 2001). With stereoscopic presentation/display technology becoming increasingly affordable and available, stereoscopic cues can now be used to enhance many self-motion simulations (particularly where the simulated environment is by necessity sparse/basic).

More recent research has also provided evidence that color information plays a role in vection. Bonato and Bubka (2006) found that a chromatic grating (consisting of blue, green, red, yellow, black, and white stripes) is more effective at inducing vection than a gray-shade version of this same grating (consisting of black, white, and four different shades of gray stripes; closely matched in luminance to the stripes in the chromatic condition). However, color manipulations should be used with caution, as research suggests that the use of red backgrounds and perceived changes in scene illumination may inhibit vection (Nakamura, Seno, Sunaga, & Ito, 2010; Seno, Sunaga, & Ito, 2010).

Interestingly, there now appears to be mounting evidence that vection is determined simply not only by physical display parameters but also by how we look at, attend to, perceive, and interpret the optic flow:

1. *Gaze effects*: Palmisano and Kim (2009) have shown that peripheral looking and gaze shifting both significantly improve the LV produced by displays simulating constant-velocity forward self-motion or low frequency self-acceleration (relative to stable central gaze conditions). Displays simulating sideways or rotary self-motions often generate optokinetic nystagmus (OKN—see Section 3.4.1.2), regular eye-movement patterns in which smooth pursuits in one direction are interleaved with rapid return flicks. Suppressing this OKN by fixating a stationary target appears to reduce CV onset and increase CV speed (Becker, Raab, & Jürgens, 2002).
2. *Attentional effects*: Vection appears to be dominated by unattended (as opposed to attended) display motion and is weakened when attentional resources are divided by increasing the task load (Kitazaki & Sato, 2003; Seno, Ito, & Sunaga, 2011a).
3. *Role of perception/interpretation*: In one such study, Seno and his colleagues (2009) used a motion-defined Rubin's vase as a vection stimulus, where the subject's figure/ground perceptions alternated between seeing either two faces (looking at each other) or a vase. Irrespective of which motion region was seen to be the figure, they found vection was always dominated by the motion of the perceived ground.

Recently, other higher-level factors have also been shown to be important in vection induction. As is demonstrated by many theme park rides, the potential for physical observer motion

(Riecke, 2009; Wright, DiZio, & Lackner, 2006), the naturalism (Riecke, Schulte-Pelkum, Avraamides, Heyde, & Bülthoff, 2006), and ecological validity of the display (Bubka & Bonato, 2010) also appear to be important for vection. For example, Bubka and Bonato reported that first-person video taken with a handheld camera when walking through a corridor induces faster vection onsets and longer vection durations than video shot from a rolling cart. They also found that full-color versions of these videos were more effective at inducing vection than their grayscale versions. In another study, Riecke et al. (2006) generated yaw CV using a rotating VE created from either a naturalistic roundshot photo of a city scene or scrambled/inverted versions of this same photograph. Based on the physical parameters of each of these displays (e.g., the numbers of high-contrast edges), the scrambled controls were predicted to produce the best vection. Instead, the naturalistic self-motion stimulus produced superior CV to both the scrambled and inverted controls.

3.4 HOW DO WE LOOK AT AND MANIPULATE OBJECTS? (MOVEMENT OF THE OBSERVER)

Interaction with the real world is rarely passive but instead involves purposive movement of the eyes, head, arms, legs, and body. The way in which an individual accomplishes this involves complex sensory–motor systems that are characterized by reflexive and volitional components. These processes must be considered when simulating the environment, especially when attempting to incorporate simulation of the results of bodily movement within a VE. Some of the systems involved in visual inspection of the world and their implications for man–machine interface are now considered.

3.4.1 Eye Movements

There are a variety of eye movements (for a detailed review, see Schutz, Braun, & Gegenfurtner, 2011), and they may be described generally as being conjunctive or disjunctive. Conjunctive eye movements involve the eyes moving in the same direction (OKN, smooth pursuit, and saccades), whereas disjunctive eye movements involve the eyes moving in opposite directions (convergence and divergence). A major consideration in ocular control concerns the ability to control the eyes when examining objects. While this is often done while the head and object are stationary (fixation), it is also done while the object moves (smooth pursuit) or when the head moves (see the vestibulo-ocular reflex—VOR—Lawson et al., 2013; Chapter 7). Visual capabilities are quite different during these oculomotor behaviors.

3.4.1.1 Fixation

When observers look at objects in the environment, they position their eyes so that the object of interest falls on the area of the retina that has the best visual acuity (the central 5° around the optic axis of the eye termed the macula). When this is achieved, the observer is said to have fixated the object. With a stationary target, fixation stability may be defined as the degree to which the eye is stable with reference to some fixation point on the object. Fixation stability is surprisingly poor. The eye may drift as much as a degree without the observer being aware of the drift in fixation. Even with well-controlled fixation, the eye is in constant motion, with microsaccades (movements) of many arcsec (Cherici, Kuang, Poletti, & Rucci, 2012; Riggs, Armington, & Ratliff, 1954). These ocular tremors are necessary for continuous viewing of an object. Artificial conditions that render the stimulus stabilized on the retina result in disappearance of the stimulus (Kelly & Burbeck, 1980). Attempts at maintaining fixation during image movement (see smooth pursuit, Section 3.4.1.3) or head movement involve interactive inputs to and from the vestibular system.

3.4.1.2 Optokinetic Nystagmus

If a contoured visual field is rotated before a stationary observer, OKN is elicited. The eyes drift in the direction of field rotation and then snap back in the opposite direction, and this cyclic pattern of eye movements is repeated in bursts, interrupted by short periods of relative gaze stability.

The initial tracking response (slow phase) is seen as a period of smooth pursuit and the compensatory snap back (fast phase) as a corrective saccade (see Section 3.4.1.4). The existence of this reflex is seen as evidence that there are visual mechanisms that take into account the effect of movement of stationary aspects of the visual field on the retina as the head or the eyes move and that these mechanisms can provide, together with the VOR, additional complementary information about how the eye should move during head movements to provide gaze stability. Unlike the VOR, this response does not abate with continuous stimulation. While the VOR is thought to adapt because the vestibular end organs cease to respond at constant velocities, OKN is seen as a compensatory response for this failing. Another difference between the two responses is that the VOR has a short latency and a slow decay, while OKN has a long latency and slow buildup. While OKN is seen as a reflex, it also can be suppressed with fixation of a stationary target.

3.4.1.3 Smooth Pursuit

Observers are quite capable of tracking moving objects with the head held stationary. This is generally considered a voluntary response engaging the smooth-pursuit system, which must calculate the speed of moving objects to maintain fixation (Turano & Heidenreich, 1999). This system requires a moving object. Smooth-pursuit eye movements cannot be made in total darkness or in a visual field devoid of movement. It is assumed that this system requires volition, because it is also possible to fixate a point when other moving objects are present. The upper limit of this response is about 100°/s.

3.4.1.4 Saccades

Another response, which has generally been considered voluntary, is the saccade. This is a ballistic movement between one fixation and another. Its latency ranges from about 150 to 200 ms, and its velocity is proportional (within limits) to the distance the eye must be moved. Saccades can occur with speeds up to 900°/s. Whereas saccades usually are made from one target to another in visual space, saccades can be executed with high degree of precision to spatial positions defined cognitively (signals in other modalities, verbal commands, memories of spatial locations). Whereas this suggests voluntary control, aspects of saccades (direction, velocity, or amplitude) cannot be changed after they are initiated. However, saccadic latency can be shortened significantly if a presaccadic fixation target is removed shortly before the saccadic target is presented, indicating that disengagement from one target is necessary prior to a saccade (Abrams, Oonk, & Pratt, 1998; Forbes & Klein, 1996). Another characteristic of vision during saccades is that the visibility of achromatic, but not chromatic, stimuli is suppressed throughout the eye movement (Burr, Morrone, & Ross, 1994). Chromatic stimuli are usually not perceived, however, because the retinal velocities exceed the resolution limits of the chromatic pathways.

3.4.1.5 Vergence/Accommodation

All of the eye movements so far considered are conjugate, but vergence movements are also necessary to focus on objects in depth. Accommodation of the crystalline lens is linked to vergence in normal observers, and changes in the refractive power of the eyes are correlated such that objects are maintained in sharp focus and registered with retinal disparities on the order of min of arc. As noted in Section 3.2.3.3, these accommodative and vergence-based responses are not only important for focusing, they may also be important for the perception of the scene's 3-D layout.

3.4.2 Head and Body Movements

One of the most important aspects of contemporary virtual simulations is the ability to provide shifts in the scene contingent on head, hand, or body movement. Position tracking technology provides a major contribution to sense of presence (i.e., the subjective experience of being in the VE—see Chertoff & Schatz, 2013; Chapter 34). The aim of the present section is to review the implications of the visual forcing functions on head and body movement that might be expected to influence the ability to provide realistic scene translations. However, it should be recognized that eye-movement-based adjustments are also needed to fully control scene properties for the impact of observer motions that impact on the visual field.

Postural stability is maintained through the vestibular reflexes acting on the neck and limbs. These reflexes are under the control of three classes of sensory inputs: muscle proprioceptors, vestibular receptors, and visual inputs. As previously discussed, a comparison of the vestibular and visual inputs is necessary to determine if the observer or the environment is moving. Visual inputs that give rise to vection have been shown to result in bodily sway and even falling in young children (Lee, 1980). It is reasonable to expect that such stimulation may result in reflexive neck muscle activity that can lead to head movements. When all or a large part of the visual field is filled with moving contours, standing observers lean in the direction of scene motion (Lee & Lishman, 1975; Lestienne, Soechting, & Berthoz, 1977). There is a compensatory leaning in the opposite direction upon the cessation of stimulation that may last for many seconds (Reason, Wagner, & Dewhurst, 1981). If the virtual scene is driven by head movement, such unintentional reflex head and body movements (and the aftereffects thereof, see Wann, White, Wilkie, Culmer, Lodge, & Mon-Williams, 2014 and DiZio & Lackner, 2013; Chapters 32 and 33) may serve to impede or interfere with scene shifts intended to simulate voluntary attempts at navigation through the VE. In some cases, reflex movements can be suppressed, but this may imply that some learning must occur to achieve efficient and unnatural adjustments that are necessary to facilitate optimal performance.

3.5 HOW DO WE DEPICT THE THREE-DIMENSIONAL WORLD IN TWO-DIMENSIONAL DISPLAYS? (HARDWARE REQUIREMENTS)

Modern VE technology is an extension of all past endeavors to depict renditions of the visual world. There has been an orderly progression in the mastery of monocular cues for depth in 2-D static displays to the movement and disparity cues for depth that are so compelling in dynamic displays. The advent of computer-controlled displays has provided the possibility of greater user interaction with these virtual depictions, and with it comes an illusory sense of presence (see Chertoff & Schatz, 2013; Chapter 34) and some adaptational problems for the user (Welsh & Mohler, 2013; Chapter 25).

3.5.1 Static Two-Dimensional Displays

A first approximation to simulating the 3-D world is the traditional 2-D display ubiquitously embodied in drawings, paintings, and photographs. What can be provided relative to the capabilities of the human visual system? Two-dimensional displays can provide fine spatial detail, variations in contrast, differences in hue, the spatial position of objects in the frontal plane, and indications of 3-D using the pictorial cues for depth (relative size, height in the field, occlusion, shading, texture, and linear and aerial perspectives). Using a single image on paper, canvas, or film, one cannot provide the stimulus conditions to provoke motion parallax, kinetic depth, retinal disparity, or, normally, motion (although Akiyoshi Kitaoka provides many examples where illusory motion, induced by oculomotor instability, is perceived in static images. See www.ritsumei.ac.jp/~akitaoka/index-e.html). The FOV is usually restricted and viewed within a 3-D framework provided by the rest of the real world. Some have argued that because of human sensitivity to the actual flatness of the 2-D plane on which 3-D simulations are viewed, the pictorial cues are in conflict with other cues provided simultaneously in the 3-D world and hence are rendered less effective (Ames, 1925; Nakayama, Shimojo, & Silverman, 1989).

As previously noted, in real 3-D views, the ability to focus and converge on different depth planes may provide reafferent information for depth judgments, but these are missing in a depiction of 3-D in only one depth plane. Can these other important visual capabilities be recruited with other manipulations of static 2-D displays? This is possible only if stereoscopic viewing conditions are provided—presenting two versions of the same scene with fusible retinal disparities. This is done in various ways. The first stereo displays were achieved with stereoscopes that allowed viewing of two disparate pictures superimposed by means of prismatic lenses. The two pictures are frequently obtained by simultaneous photography of the same scene with two cameras in slightly different positions and then presented to the appropriate eyes. An anaglyphic approach renders the two members of the stereo pair in different wavelengths (e.g., red and green) and lets the observer view the two disparate scenes with glasses that segregate the visual images with different chromatic filters in front of the eyes. This separation is rarely perfect, with unwanted ghost images appearing in each eye's view that were meant for the opposite eye—referred to as *crosstalk*. However, typically the resulting *left* and *right* eye images can still be fused to provide rather striking depth information and limited motion parallax. Unfortunately, one loses the ability to render color appropriately with this technique (binocular color rivalry is also introduced). A similar technique involves the use of polarizing filters. Two disparate scenes are presented through orthogonally polarized filters. Each eye of the observer views through a polarizer matched to only one of the scenes. The visual system combines the two views and is able to detect the disparity cues. This method allows more natural color rendition; however, it also suffers from crosstalk between the left and right eyes' images (which increases dramatically as the observer rotates his/her head away from vertical). The stereoscopic images so popular in the comics section of Sunday papers use an anaglyphic approach, which does not require chromatic glasses, but require the observer to converge or diverge the eyes to achieve fusion. The motivation to attempt such viewing is to find the hidden figure not visible in the monocular or nonfused binocular view. More recently, with the advent of new 3-D display technology, there has been a revival of interest in displays using either parallax barriers or lenticular prisms. Images are created by alternating left and right eye views in columns of pixels that, after printing, are overlaid with a sheet that either directs the alternate columns of the images to left and right eyes or provides a barrier that alternately blocks the left or right eye view for the appropriate columns. With these displays, disparities are rendered to the two eyes to create compelling 3-D scenes without the need for additional glasses, but in most cases, they must be viewed from a relatively specific angle. Outside these angles, double images may be detectable. In this sense, they are more restrictive than the systems using filter lenses to direct the separate views to each eye.

Modern televisions also support 3-D displays. They may also require glasses, employing a crossed polarization method or more frequently frame interleaving. The polarization method is as described earlier, but frame interleaving requires temporally alternating frames that display left and right eye views, synchronized with glasses that alternately allow only the left or right eye to view the scene. Usually, this is achieved using liquid-crystal display (LCD) technology and is therefore quite slow, but provided frame alternation rates exceed 120 Hz, each eye receives a 60 Hz image sequence and relatively flicker-free rendition of 3-D scenes is achieved. Again, crosstalk is common with these shutter glasses. The polarization reduces the luminance of the images but natural colors can be rendered.

Head-mounted displays (HMDs) can also render 3-D scenes. They provide a separate screen for each eye that can be readily fused to create the retinal disparities needed to perceive depth. This method allows more natural color and perfect separation between the left and right eye images. It does not require frame alternation but usually employs LCD screens with quite slow refresh rates (approximately 60 Hz). While the gain in level of quality of 3-D simulation provided by stereo viewing is impressive, the static 2-D display lacks the vividness that scene motion conveys.

3.5.2 Dynamic Two-Dimensional Displays

In the early part of the last century, considerable amusement was provided to those fortunate enough to own a stereoscope. The excitement generated by such devices paled in comparison to the first moving pictures, however. Movement in nature is an extremely compelling sensation, providing immediate impressions of one's position relative to other objects in the world. The simulation of movement in static displays was first accomplished with a deck of cards, wherein an ordered sequence of pictures denoting differences in scene position was presented in rapid succession by sleight of hand. The deck was bent and individual cards were allowed to slip off the thumb to reveal a simulation of motion. The motion picture projector was developed using the same rudimentary principle, rapidly sequencing a series of still pictures. These pictures were captured on filmstrips with cameras capable of rapid photography. The current standard for film projection is 24 frames/s, and each frame is transilluminated for about 30 ms. A rotary shutter blade is interposed while the film is advanced (taking about 10 ms), yielding a frame rate of about 24 frames/s. This simple arrangement provides a series of still pictures but also produces considerable detectable flicker. By increasing the rate of flicker (to 50–60 Hz), the achromatic critical flicker-fusion limit for the human observer is surpassed and an acceptable sensation of smooth movement is obtained (Cropper & Badcock, 1994).

While these developments added realistic movement to simulations of the real world, stereoscopic and motion parallax cues were still not available. In the 1950s, the first 3-D movies were screened, and this technology combined the anaglyphic technique for stereo vision with the cinematic technique for motion, providing quite compelling perception of motion in depth and some limited motion parallax information. In the last decade, 3-D television has become readily available. Displays using either anaglyphic techniques with passive polarizing glasses, frame alternation with active LCD glasses, or lenticular screens provide ready access to moving 3-D images sequences in quite natural color. Modern cinematography has become an ubiquitous source of entertainment and education in this century, but it does not approach a true simulation of our normal experiences with the 3-D world. In most cases, the observer is quite passive and relatively immobile, with the 3-D point of focus specified by the cinematographer rather than the viewer. That is also the case with standard video simulations, which, although technically more advanced electronically, suffer from the same inadequacies. Video, however, affords the possibility to interface with sophisticated computer systems that offer the possibility to incorporate user movements into the simulated environment.

3.5.3 Electronic Displays and the User Interface

The cathode ray tube (CRT) began as a method of rendering simple electronic activity immediately visible. For example, oscilloscopes are often used to display voltage changes over time. The early versions involved an electron *gun* aimed at a phosphorescent surface. Deflection plates or coils controlled the direction of the gun, and the intensity of fluorescence was varied via accelerating plates that determined the rate of electron flow. Originally, control of the gun was accomplished with *vector scanning* that addressed only the points (pixels) on the screen necessary to *draw* the image in question. This proved adequate for simple geometric shapes, but more complex scenes required the more elaborate *raster scanning* approach wherein every pixel was addressed. In raster scanning, the stream of electrons is swept across the screen horizontally from left to right and then snapped back and down one line in order to paint the next row. This procedure is repeated until the entire screen has been painted and the gun is turned off (blanked) during retrace (snap back). One complete painting of the screen is termed a frame, and frame rates above the human flicker-fusion point (50–60 Hz) result in a flicker-free sequence of pictures. This technology provided the basis for black-and-white television. This relatively simple system was expanded to provide color television by including three independent guns that selectively stimulate colored (red, green, and blue) phosphors laid down as triplets on the screen. One problem with CRT displays is that the tube length increases as the tube width and height increase, resulting in large displays that are quite bulky. For this reason, they have fallen out of favor with television and computer monitor manufacturers in spite of advances in display technology that reduced the size of the apparatus and increased the spatial and temporal resolution. The last decade has seen these displays largely replaced by liquid-crystal, plasma, and electroluminescent digital displays that are far less bulky but have disadvantages associated with resolution, color, or expense; these shortcomings are rapidly being reduced. The switch from the analog technology used in CRTs to digital displays is currently associated with a much poorer ability to render luminous intensity. Advanced graphics systems are available with digital-to-analog converters (DACs) capable of 14-bit intensity resolution (16,384 levels) on each gun that are all available with CRTs, but LCD panels are currently 10 bits (1024) or more frequently 8 (256 levels) and finer gradations of their input signals will not alter the step size between levels (Wang & Nikolic, 2011). The discussion of relevant parameters for meeting user specifications will be undertaken with regard to CRT displays, but the principles will apply to all display technologies, and promising new technologies will be noted in the discussion. The parameters of importance for meeting the user requirements are provided in Table 3.1.

3.5.3.1 Spatial Parameters

The FOV depends on viewing distance (FOV = atan[screen size/viewing distance]), but at a fixed distance, the FOV for modern displays varies widely, from the tiny screens produced for miniature television to the panoramic screens employed in IMAX (Image MAXimum) and movie theaters. The small screens have been incorporated into HMDs, with optics designed to allow sharp focus of images within the near point of vision and to provide a large FOV. Video projection systems can fill the human FOV but often present problems of spatial resolution, require sophisticated warping software to prevent image distortion, and are limited to relatively low luminance levels (although this is rapidly improving). CRT-based HMDs have a limited FOV (<80°) and, essentially now replaced by lighter LCD displays, offer somewhat larger fields of view (105°) and simpler spatial image specification. Although the peripheral retina has poor spatial resolution, wide field of view provides valuable information for visually guided behavior and for producing the illusion of vection (see Hettinger, 2014; Chapter 18).

The spatial resolution of most displays does not approach the limits of human visual acuity (approximately 1 arcmin/pixel) or spatial positioning ability (interpolated displacements of 5–30 arcsec) unless viewing distance is substantial (frequently several meters), which, of course, reduces FOV. This trade-off between FOV and spatial resolution is an important limitation for visual simulation, but there has been steady improvement in spatial resolution over time with all technologies and this is likely to continue. The major cost of less than optimal resolution is a sacrifice of texture and the sudden appearance of objects at simulated distance. This latter characteristic is potentially very serious. Castiello, Badcock, and Bennett (1999) have shown that when observers are reaching to grasp an object, a suddenly illuminated distractor object changes the motor aspects of reaching behavior. This change only occurs if the illumination is sudden; gradual onset has no effect, and if a spotlight suddenly illuminates the peripheral scene but there is no object, the motor movement is not affected. Thus, in VE setups, suddenly occurring peripheral objects would be expected to be the most disruptive of all possible object appearance modes if an observer is trying to make motor movements in the setup. Higher-resolution renditions are needed to overcome this problem.

A second critical aspect with regard to spatial resolution is the uniformity and stability of the pixel matrix over time. In this regard, LCD and the newer organic light emitting diode (OLED) displays are superior to CRTs. The latter use magnetic deflection to spread the image across the screen and this can be influenced by local magnetic effects (e.g., northern vs. southern hemisphere or magnets placed nearby), whereas the LCD and OLED matrices have a physical basis and do not move. CRTs may also perform differently when cold, such as when first turned on (Metha et al., 1993).

To match or exceed the contrast sensitivity of the human visual system, the luminance steps (gray scale) must be small enough to provide differences in luminance to which humans are insensitive. For many natural scenes, 24-bit (8 bits on each of the red, green, and blue guns) graphics systems are adequate. However, if fine contrast discrimination is required, most observers will require either higher-contrast resolution or that the full range of steps be compressed into the near-threshold range of contrasts being presented. The latter has the consequence of restricting the maximum contrast

TABLE 3.1
Relevant Display Factors Relating to Human Visual Abilities

Dimension of Vision	Critical Display Parameters[a]			
Spatial vision	Pixel uniformity: OLED and LCD displays have structured pixel arrays. CRT can be distorted by magnetic interference. This is important for displays that require precise geometry.	Pixel size (screen size ÷ number): finer spatial resolution allows greater positional precision and more accurate shape rendition, particularly for sharp-edged figures that cannot be interpolated across pixels.[b] Optimal positional thresholds can subtend a few seconds of arc. If sharp-edged stimuli are employed, the pixels need to match this resolution. Interpolation allows finer positional adjustment and a tolerance for lower display resolution.	Luminance uniformity: OLED/LCD are more uniform than CRTs, which are brightest in the center. This means object intensity and hue can vary with location when using CRTs, unless corrections are applied when generating the stimuli.	Intensity resolution (bit depth): more bits allow finer control of contrast and larger range and support better antialiasing for rendering sharp edges at small angles to pixel array orientation. CRTs are superior because of analog control. Graphics card sets the limit (best 14 bits, 16,384 levels, per R, G, and B guns). Most LCD and OLED are limited to 8-bit (256 levels) intensity per R, G, or B channel.
Color vision	Display type: OLED has much larger gamut than typical CRT or LCD (Figure 3.11), which allows a bigger range of colors to be represented veridically. The gamut is, however, still smaller than that of normal human vision.	Phosphor types: phosphors, LCD channels, and OLEDs may vary in their exact location in CIE space and thus alter the available gamut (Figure 3.11). It is therefore important to ensure that the desired colors can be represented by the display system.	Gun independence and stability: CRT gun interactions complicate calibration of intensity ranges. High-end CRTs ensure minimal interactions when intensities are high for one gun and low for the others but can require extensive warm-up periods. OLEDs should also be excellent here.	Intensity resolution (bit depth): 8 bits/R, G, or B channel is insufficient for measuring chromatic thresholds unless the intensity range covered by the 256 levels is compressed (only possible for analog displays). If a large color range is required simultaneously with small changes in color, then systems should be chosen with as many bits/gun as possible.

(*continued*)

TABLE 3.1 (continued)
Relevant Display Factors Relating to Human Visual Abilities

Dimension of Vision	Critical Display Parameters[a]			
Image motion	Display type: LCDs have slow response rates making them poor for displaying moving images veridically; CRTs and OLEDs can both provide the rapid responses needed.	Raster rate and phosphor decay: smooth slow motion requires fine temporal resolution with little temporal smear. LCDs are unsuitable for this purpose with low refresh rates and slow responses. CRTs with >120 Hz refresh rates are preferable.	Pixel size and spacing: moving images are more realistically represented when both small spatial and temporal displacements can be produced. This is particularly important for representing very slow motion. However, as noted in Footnote b, interpolation of smoothly varying luminance profiles can give subpixel precision. The motion of sharp-edged stimuli moving slowly will be quantized by the pixel size.	Intensity resolution (bit depth): the interpolation procedure described in Footnote b is limited by intensity resolution. Very slow motion may require small intensity changes for each pixel. Eight-bit systems are often sufficient for motion of natural scenes, but to measure the limits of motion sensitivity requires finer adjustment of intensity in many cases. Sharp edges can only be moved from one pixel column to the next and are not directly affected by intensity resolution.
Stereopsis	Interlaced frames: flicker-free viewing requires >60 Hz for each eye or >120 Hz if using a single monitor. CRTs can achieve this. Anaglyph methods: >60 Hz sufficient.	Monocular monitors: haploscopic systems using either a different monitor for each eye or different half screens present challenges for matching of intensity ranges. CRT half-screen systems will have intensity gradients falling away from the screen center in opposite directions. Contrast differences can produce distortions in perceived disparity.	Frame disparity: in all displays, large disparities will produce diplopia. Small ranges of disparities[c] within a scene are desirable to avoid this issue.	Intensity resolution (bit depth): as discussed previously, rendering of fine spatial changes of location for the right and left eye image components to produce small disparity differences may require interpolation (or resampling) of the image for a new position relative to the pixel matrix. Fine luminance resolution allows for smaller image changes and more precise interpolation.

Vection (visually simulated self-motion induced in both active and passive observers)	Scene update rate (lag): display lags up to 160 ms impair vection and generate simulator sickness symptoms. With longer lags, the visual system appears to override or downplay the visual–vestibular conflict produced by the observer's head motion.	FOV: while it is possible to induce vection with display sizes as small as 7.5° of visual angle, larger displays are required for compelling vection experiences. Recommended display sizes for vection are between 60° and 120° of visual angle.	Frame rate: LCD, DLP,[d] OLED, and CRT projectors/displays are all capable of inducing compelling vection. Vection can be induced by displays with refresh rates less than 30 Hz.	Intensity resolution (bit depth): limitations in spatiotemporal resolution generate *jaggies* (or pixel creep) in computer-generated self-motion displays. These artifacts are problematic when vection displays are presented only to the observer's central visual field (where acuity is the highest) and simulate self-motion with respect to a 3-D (as opposed to a 2-D) environment.

[a] This table lists the parameters that are critical for a variety of different tasks. At this point, no display technology is optimal for all applications. The intention here is to indicate what aspects of performance may be limited by the parameters so that the reader can choose the optimal display system for the tasks they intend to implement. In most cases, measuring human performance limits on a particular dimension will require choosing a display that offers the best available resolution on that dimension. Given that is the case, we have not tried to indicate preferred choices of display technology.

[b] The spatial resolution of the pixel array is more critical for sharp-edged stimuli than those with gradual luminance variation at their contours. Sharp edges can only be moved from one pixel location to the next, and thus the pixel size sets a limit on the minimum change but smoothly changing luminance gradients can be resampled at different locations on the pixel array, allowing for positional changes that are smaller than the pixel size, provided the intensity resolution is high enough to adequately sample the luminance waveform. This interpolation procedure is applicable for a wider array of natural images.

[c] There are significant individual differences in sensitivity to diplopia, for example, reported diplopia thresholds have ranged from 2 to 20 arcmin for horizontal disparity in the fovea (Duwaer & Van Den Brink, 1981).

[d] DLP (digital light projector) is a digital micromirror technology is produced in two forms. Both are very fast and capable of bright, high-resolution displays with up to 10-bit (1024 levels) intensity resolution per color. However, one method achieves separate color channels by rotating a colored filter through R, G, and B sectors in the otherwise white light path. This presents R, G, and B components of images at separate points in time, and if an observer makes an eye movement during the presentation, objects are seen as different colored replicas smeared across the scene. The second method uses a prismatic process that presents all three color signals simultaneously and should be preferred.

range available at any instant but does allow small contrast steps to be presented. Computers that boast 16-bit DACs for each gun are preferable, but expensive and uncommon. Many commercially available monitors provide adequate luminance (up to 100 cd/m^2) and support photopic viewing levels. Thus, the current limitation in intensity resolution in CRTs is usually provided by the DACs employed. However, LCDs and OLED displays are digital, and step sizes in intensity for each R, G, and B component are fixed, as noted previously. Ten-bit (1024 levels, 1920 × 1080 pixels, 60–85 Hz, 100 cd/m^2) OLED displays and 12-bit (4096 levels, 1920 × 1200 pixels, 120 Hz refresh, 250 cd/m^2) LCD displays are already available.

3.5.3.2 Color

Many modern monitors are capable of providing high spatial resolution (1920 × 1200 pixels or more) and moderate temporal resolution (60–150 Hz frame rates) while also providing a colored image. The CRT monitors contain three electron guns, which are each aimed at a different phosphor type so that in near spatial proximity on the screen, red, green, and blue signals can all be produced, while LCD and OLED displays have individual elements that can be switched on or off to display the color at a particular location. There are a number of important monitor characteristics that are required for high-fidelity color rendition.

The range of achievable colors depends on the chromaticity coordinates of the phosphors used. However, current monitors are only capable of producing a small part of the full range of chromaticities detectable by the human visual system. Figure 3.11 depicts the full-color gamut specified in the CIE 1931 xy chromaticity space for normal human vision, with the more restricted gamuts available for different display systems marked as triangles. The spaces shown are EBU (European Broadcast Union) that corresponds to the PAL television standard, ITU-R (the High Definition Television [HDTV] standard), SMPTE-C (the current American broadcast standard, a variant of NTSC),

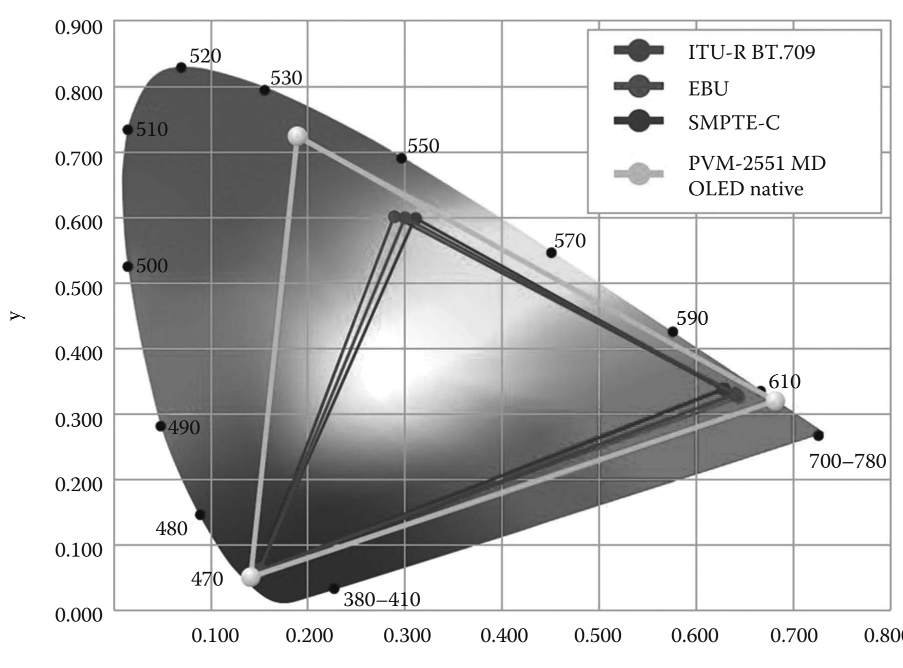

FIGURE 3.11 **(See color insert.)** The CIE (1931) xy chromaticity space for normal human vision, with a 2° field, is depicted as the colored region, with the gamut of color available with the typical RGB monitor broadcast standards depicted as triangles. The perceptual locus of narrowband spectral lights is indicated by numbers specifying their wavelengths. (Adapted from Sony PVM-2551MD Medical OLED monitor catalog, 2011.)

and PVM-2551 MD (an OLED native gamut used in the Sony Medical OLED monitors). The OLED display represents a significant improvement in the available color gamut.

Within the available gamut, accurate color rendition in CRTs requires high gun stability over time and gun independence, but both of these factors need to be rated against the intended use of the display. Gun stability is often a problem during the first hour a monitor is turned on. During this period, the luminance and chromaticity of a display will vary, even in very good monitors. Metha et al. (1993) provide some representative data and a method for evaluating the visual salience of these changes.

Gun interactions are more problematic. On less expensive CRT monitors, the intensity of the output from one gun will vary as the output of the other guns increases. The consequence is that there is not a fixed relationship between a given intensity level specified in the driving software and the output of a particular gun. The output will vary with different images. Observers are quite tolerant of these changes for noncritical applications (e.g., word processing), but the interactions are likely to be destructive when accurate color discriminations are required. Expensive monitors are more likely to avoid this problem, but the only way to be sure is to measure the light output for each gun independently and compare the output for several levels of activation of the other guns (see Metha et al., 1993). This check is critical as many manufacturers of CRT monitors deliberately build in a gun interaction that affects the upper half of the intensity range to minimize the amount of flashing when users move pointers from one window to another in graphical user interfaces.

The final issue of significance for rendering color is the variation in light output across the CRT monitor screen. Some variation is due to nonuniformity of the phosphor deposition during manufacture, but most of it is due to the design characteristics of the shadow mask technology. All such monitors are brightest in the center of the screen and luminance drops considerably toward the periphery. In part, this is due to the smaller effective aperture in the shadow mask at shallower angles of incidence, but the loss is also due to the polarity of the emitted light. The output is much more intense in the direction continuing the path of the electron beam, a direction that points to one side of an observer aligned with the center of the screen. Metha et al. (1993) found this loss to affect luminance more than chromaticity, but ideally stimuli requiring accurate luminance and chrominance discriminations should be small and placed centrally on a CRT screen to minimize spatial variations due to the monitor.

Fortunately, many of these issues are minimized in the OLED displays. The color gamut is larger, and because the pixels are separate elements illuminated in identical manner, screen uniformity is much higher. Stability over time is also claimed to be high but remains to be tested for particular displays.

3.5.3.3 Image Motion

Two factors that have a strong influence on simulated motion when using CRTs are phosphor persistence and frame rate. Phosphor persistence refers to how quickly the phosphorescence fades after it has been energized. Long-persistence phosphors reduce flicker but will also result in image smear when image movement is simulated. Short-persistence phosphors require higher update rates for pixels that convey the presence of static parts of the image, and if the rate is too low, flicker may be a problem. In practice, it is quite easy to determine whether the quantization of motion will be detectable. If the temporal frequency components introduced by the quantization fall within the range detectable by an observer, then the observer will be able to discriminate between smooth and sampled motion sequences (Cropper & Badcock, 1994; Watson, Ahumada, & Farrell, 1986). Other factors may be critical if either speed difference judgments or very slow speeds need to be produced. In both these cases, the achievable steps in speed will depend on a combination of the frame rate of the monitor and the spatial resolution. The quantization in both time and space limits the minimum speed and minimum difference in speed achievable.

The new display technologies differ in their suitability to display moving images. LCDs have slow temporal responses and therefore often produce artifacts, such as motion smear. Manufacturers have tried to remedy this limitation using deblurring algorithms but these alter the specified image sequence and can create detectable quantization artifacts. OLED displays, on the other hand, employ a very fast technology to illuminate a pixel and are therefore able to avoid motion-dependent spatial artifacts and maintain full image intensity for moving images.

3.5.3.4 Stereopsis

Stereoscopic vision is currently enjoying a boom and is supported by either providing separate monitors for each eye (e.g., HMDs and boom-mounted devices), shutter glasses, anaglyphic glasses, or projectors and autostereoscopic displays, including the new arrays of 3-D television systems. Dichoptic presentation using two monitors presents two views of the same scene, one to each eye, and depends on binocular fusion to yield stereopsis. Electronic shutter glasses present alternating views of the display synchronized to the frame rate such that one interleaved frame in each pair is presented to each eye. The display provided in each set of odd or even frames contains a disparity between the two eyes and binocular fusion provides stereopsis. Anaglyphic glasses filter the images provided to the two eyes with chromatic filters, usually red and green in older systems, and the display contains disparate images rendered in the two colors; modern systems usually employ orthogonally polarized filters with full-color images.

Autostereoscopic systems involve lenses imposed between the viewer and display or use a haploscopic arrangement in which dichoptic stimulation is achieved using mirrors to align a different display with each eye. While most would agree that stereoscopic vision is extremely important for the feeling of immersion and presence (see Chertoff & Schatz, 2013; Chapter 34) that the VE seeks to convey, all current techniques involve unwanted attributes. HMDs are bulky, have insufficient FOV, and can cause eyestrain if the optics create competition between the accommodative and vergence systems (see Wann, White, Wilkie, Culmer, Lodge, & Mon-Williams, 2014; Chapter 32). Electronic shutter glasses are less cumbersome, but are also restricted in FOV, suffer from crosstalk, and require higher frame rates to avoid flicker. With flicker-fusion thresholds near 60 Hz at current display luminances, a frame rate of 120 Hz or greater is required for flicker-free vision. The older anaglyphic techniques are compromised because color is used to separate the signals for the two eyes (often not perfectly) and cannot simultaneously render a true-color image. They are also limited by the restrictions of FOV imposed by glasses. Most autostereoscopic systems offer only small display sizes.

However, the modern large screen 3-D televisions (also described in Sections 3.5.1 and 3.5.2) offer improvements in FOV, provided high spatial resolution is not simultaneously required. Stereoscopic capability is desirable because it provides additional binocular depth cues through motion parallax and motion in depth, and this technology is likely to continue to improve given the current resurgence in manufacturer interest. Advances using lenticular displays allow for glasses-free viewing and provide a more natural viewer experience, in addition to providing unobstructed views of the observer that enhances the ability to readily monitor eye movements. This combined with OLED technology would provide a good combination of spatial resolution, color gamut, and temporal response for many research purposes and appears to be achievable with current technology.

3.6 WHAT ARE THE SUCCESSES AND FAILURES OF OUR ATTEMPTS TO SIMULATE THE *REAL WORLD*? (THE STATE OF THE ART AND BEYOND)

3.6.1 Successes

As observed above (3.5.3.1), the spatial aspects of visual simulation seem to be adequate for representing small regions of real-world scenes. HMDs have benefited in the last decade from reduced mass and improved spatial and temporal resolution with less peripheral distortion. Emerging technology promises even higher resolution and more encompassing FOV. As previously discussed, the depiction of color only approximates the limits of human abilities with new OLED technology promising a wider color gamut and more uniformity across displays. The temporal characteristics of most displays are adequate and might be expected to improve with the faster responses and higher intensities of OLEDs that are starting to become available. These improvements will support more realistic depiction of object motion and smooth scene translation. As simulation of motion profiles improves, more pronounced vection and a more realistic sense of presence can be expected. Improving the sense of presence, however, might produce more rather than fewer problems for users.

A significant, recent improvement has been the ability to incorporate eye-movement monitoring in HMDs. Without this, observer-generated changes in fixation within a scene could not be compensated for, precluding the representation of natural changes of the visual field under free viewing.

3.6.2 Failures

One of the most problematic areas in VE system design concerns the simulation of environmental motion relative to the observer as their head or body moves in the simulator. VE setups using head-coupled or head-slaved tracking systems all contain unavoidable display lag. *End-to-end* system lag refers to the time it takes to track the observer's physical movements and update them into the graphical display (typically by shifting the observer's virtual viewpoint). This lag depends on a number of factors, including the computational power of the stimulus generation computer, the speed of the particular tracking system, and often the required graphical detail of the scenes to be depicted. A decade ago, the best VE systems boasted 40 ms update rates but achieve this by sacrificing resolution; modern systems are substantially faster but are still unable to render a full field high-resolution display with rapid updates. In most systems, end-to-end system lag ranges between 40 and 250 ms (Moss, Muth, Tyrrell, & Stephens, 2010). The resulting display lag can have detrimental effects on perceptual stability (Allison, Harris, Jenkin, Jasiobedzka, & Zacher, 2001), presence (Meehan, Razzaque, Whitton, & Brookes, 2003), simulator sickness (Draper, Viirre, Furness, & Gawron, 2001), simulator fidelity (Adelstein, Lee, & Ellis, 2003; Mania, Adelstein, Ellis, & Hill, 2004), and virtual task performance (Frank, Casali, & Wierwille, 1988; So & Griffin, 1995—see also Chapters 23 through 26, this book). These detrimental effects are often due to the mismatch between visual and vestibular cues to motion. The lag in updating the scene means that a vestibular cue to head movement is followed by, rather than being coincident with, retinal-image motion. This particular problem can potentially be overcome by increased computational power that may eventually allow the time lag to become imperceptibly small. Estimates of the minimum detectable lag can vary markedly depending on the task. Some studies suggest that display lag has to be at least 150 ms to be detectable (Moss et al., 2010) with as little as 60 ms impairing perceptual stability (Allison et al., 2001). By contrast, other studies suggest that display lags as short as 15 ms are detectable (Mania et al., 2004).

A much more difficult problem is the need to provide vestibular input more generally. When observers are free to navigate around a scene, there is often a mismatch between visually indicated changes and cues from the vestibular system. In the extreme, vision may indicate locomotion, while the vestibular system indicates that the body is stationary. This mismatch frequently produces feelings of nausea. The relationship between visual and vestibular inputs is also under constant recalibration. Thus, periods of mismatch can lead to recalibration, which will maintain feelings of nausea in the natural environment until the normal relationship has been relearned (see Welsh & Mohler, 2013; Chapter 25). Improving visual display technology will not reduce this particular problem; however, providing appropriate vestibular inputs might reduce this (see Lawson et al., 2013; Chapter 7).

While current projection displays do afford FOVs that approach the limits of human observers, HMDs are limited by the size of small monitors relative to the interocular distance and the optics necessary to allow clear vision within the near point. Many optical compensatory arrangements result in eye strain and visual aftereffects. Research on the aftereffects generated with the use of HMDs has indicated another conflict situation in apparent motion situations. Stereoscopic displays that require near viewing often give rise to induced binocular stress. Mon-Williams, Wann, and Rushton (1993) reported loss of visual acuity and eye muscle imbalance after only a 10 min exposure to such displays, with some of the symptoms associated with motion sickness. They suggest that these aftereffects stem from adaptation of the accommodation–vergence system due to the disparity between the stereoscopic depth cues provided and the image focal depth. They later reported that displays that preserve the concordance between vergence and accommodation do not produce these aftereffects with viewing times of 30 min (Rushton, Mon-Williams, & Wann, 1994).

Unlike distance estimation in the real world, which is quite accurate up to 20 m (e.g., measured by having people walk blindfolded to previously seen targets—Loomis, Fujita, Da Silva, & Fukusima, 1992), distances tend to be underestimated in VEs. When viewing VEs through HMDs, helmet weight, low display resolution and level of scene detail, limited FOV and reduction in depth cues have all been identified as possible causes of this distance underestimation. Over the last decade, a number of studies have been conducted to examine the effects of these different factors, both in isolation and in combination (e.g., by having subjects view the real world through mock HMDs with similar weights and FOVs to real HMDs). While their findings have been somewhat mixed, it appears that both the mechanical and optical properties of HMDs contribute to distance underestimation in VEs (Bodenheimer et al., 2007; Knapp & Loomis, 2004; Messing & Durgin, 2005; Thompson et al., 2004; Willemsen, Colton, Creem-Regehr, & Thompson, 2009). This distance underestimation appears to be less likely when using large screen immersive displays (Plumert, Kearney, Cremer, & Recker, 2005). Interestingly, there is evidence that distance estimation in augmented reality (when virtual targets are viewed in the context of a real surrounding environment using a see-through HMD) is similar to that obtained when viewing real targets in a real environment (Jones, Swan, Singh, Kolstad, & Ellis, 2008).

3.7 NEW RESEARCH DIRECTIONS

The development of VE systems has proceeded without an abiding concern for the intricacies of human sensory systems. As these factors are considered in simulator design, the dearth of knowledge of many aspects of human perception is highlighted. In many respects, efforts to create realistic VEs serve to drive and motivate basic research efforts, but they also point to the limitations of classical research settings. One major problem in the approach to the study of human abilities centered on the way in which efforts have been compartmentalized (Kelly & Burbeck, 1984). The use of VE technology has forced, and will continue to force, a broadening in the scope of interests and considerations in the study of sensory motor interactions (e.g., Calvert, Spence, & Stein, 2004). As researchers begin to delve more deeply into these problems, the apparatus necessary to address these issues will require higher resolution on both spatial and temporal dimensions and become more expensive. This may dictate that research goals must be narrowed. It may be laudable, but not practical, to attempt to generate realistically natural displays that are valid in every VE application although improvements in this desire have arisen with the incorporation of augmented reality systems that include cameras mounted on HMDs to combine computer-generated displays with the currently viewed scene (Lee, Oakley, & Ryu, 2012).

One way that research tasks may be limited is to address perceptual problems of the particular device in question. While one can predict, to some extent, how well a display might meet the requirements of the human user, a more empirical approach might be to address the adequacy of the instrument with psychophysical experiments in the simulator at the outset. This approach may also reveal idiosyncratic problems that emerge through the use of the device and belie concerns based on theoretical predictions.

3.7.1 Limitations of Classical Research

Much of the research reviewed previously was generated in laboratories developed for the study of a particular visual capability (e.g., color, motion) without regard to how that ability might be compromised when other sensory and motor systems are in operation. In many experimental situations, the aim was to study the ability in isolation to avoid sources of psychological interference (e.g., arousal, attention, expectation). The aim in VE, however, is often quite different. The user is often actively engaged in tasks where many systems are employed.

Perhaps the greatest research impetus from the use of VE technology has been to emphasize sensory–sensory and sensory–motor interactions. There is a pressing need for increased information with regard to visual–vestibular interactions. In part, this is to alleviate some of the negative effects of using VE technology (e.g., nausea) but also because vestibular inputs can change basic properties of visual neurons, such as their preferred orientation of line stimuli (Kennedy, Magnin, & Jeannerod, 1976). It will be important to determine to what extent vision is identical when vestibular inputs are decorrelated.

Several recent studies have reported that consistent head movements (Ash et al., 2011) and consistent biomechanical cues (generated when walking on a treadmill—Seno, Ito, & Sunaga, 2011b) both appear to facilitate the vection induced by optic flow. Similarly, presenting consistent auditory cues and air flow to the observer's face has also been shown to improve the vection induced by optic flow (Riecke, Väljamäe, & Schulte-Pelkum, 2009; Seno, Ogawa, Ito, & Sunaga, 2011). However, other research suggests that we can be remarkably tolerant to expected Ash, Palmisano, Govan & Kim, 2011 conflicts between visual, vestibular, and nonvisual self-motion inputs (Ash & Palmisano, 2012; Ash, Palmisano, Govan, & Kim, 2011; Kim & Palmisano, 2008). Unfortunately, while treadmills are one of the most widely used (and researched) techniques to allow walking in VEs, they are known to generate problematic visual illusions—decreases in the perceived optic flow speed during treadmill walking (Durgin, Gigone, & Scott, 2005) and increases in perceived real-world walking speeds directly afterward (Pelah & Barlow, 1996).

3.7.1.1 Sensory and Sensory–Motor Interactions

The synthesis of classical theories of color vision into the current view of multistage processing is perhaps a model for how researchers will proceed in elaborating knowledge of visual processing relative to nonvisual sensory and motor processes. While considerable psychophysical and physiological evidence supported both the trichromatic and opponent process views, the controversy cooled when we learned how these mechanisms were staged and integrated. In similar fashion, our understanding of vision during various eye movements (previously discussed) is a step toward understanding how vision is compromised during such activity, but we have yet to fully appreciate how visual scene movement might elicit reflexive eye, head, and body reactions. Becker (1989) reported that head movements are a regular feature of gaze shifts greater than 20°. Some of the studies that have addressed coincident eye, head, and hand movements have done so while the subjects performed a volitional task (e.g., Pelz, Ballard, & Hayhoe, 1993), and they reveal the sequential ordering of gaze, head, and hand movements. With the use of head trackers, incorporating eye-movement monitoring and other pointing devices, it may be the case that these systems are requiring learned modifications to the natural interaction of the sensory and motor systems.

3.7.1.2 Esoteric Devices and Generalization

If it is the case that users are forced to adapt to the idiosyncrasies of a particular VE, then the traditional laboratory work from which researchers hope to predict visual performance may not provide the basis for legitimate generalization. Studies of basic human visual function may not tell about the human capability to cope with unique sensory–motor rearrangements. Investigations designed to study these abilities (e.g., Welsh & Mohler, 2013; Chapter 25) cannot anticipate all the kinds of interactive adaptations that might be required by new technology. For this reason, systematic research on basic visual function and sensory–motor interactions should be incorporated into the design and testing of new VE systems at the outset.

3.7.2 Basic Research in Simulators

Many of the more sophisticated VE systems in use today were developed to address practical problems in training and performance. Various vehicular simulators, for example, are used to train operators initially and to help them maintain their skills. Since this represents the primary mission of the apparatus, little or no time is allotted for investigation of many of the human-interface problems that might exist unless they severely compromise that mission. It is important for the successful development of future systems to document such problems in existing systems and to identify causal variables and usage protocols to prevent them. The work of Kennedy and others (see Stanney et al., 2013; Chapter 31) serves as an exemplary first step in this direction.

3.7.2.1 Machine or Observer Limited

The basic question asked of vision scientists concerning VE systems is: will this system meet or exceed the visual requirements of the human user? This chapter has been concerned with the answer to that question and has attempted to review visual capabilities and relate them to technical specifications. While such an approach has considerable heuristic value, many assumptions are made and conclusions based on such analyses are, in a sense, still quite theoretical. A more direct and empirical approach is to measure visual function within the VE system in question. Some things are readily predictable from extant data. For example, one might ask if the contrast scaling of a display is fine enough to exceed the contrast discrimination capacity of the user. However, the relationship between perceived distance traveled using visual information and the simulated distance traveled is less predictable (Redlick, Harris, & Jenkin, 1999). The most straightforward way to answer that question would be to measure the function of interest inside the VE system and compare the results to measures obtained with more standard methods. If similar findings are obtained with the two methods, the system could be described as observer limited and adequate in that particular dimension. If performance on that measure was poorer in the VE system than by conventional testing, performance could be termed machine limited, and this finding would indicate that the system falls short of the user's capability. It is an open question at the moment as to whether VE systems need always be observer limited to be adequate for a given application. Many visual tasks are not performed with stimuli at the limits of visual capability, but this is sometimes not clear until the tasks are performed with observer-limited and machine-limited systems.

3.7.2.2 Adaptation and Emerging Problems

Many early and continuing studies of human perception indicate that considerable adaptation occurs under conditions of perceptual rearrangement (Webster, 2011). Such adaptation can result in a rescaling of the perceptual systems and to corresponding changes in perceptual–motor responses. With brief exposures (<10 min) to such stimuli, adaptation occurs but recovery is relatively fast. With prolonged exposures (>10 min to several days), such adaptation might be more permanent and require readaptation over an extended period to restore normal perception and performance (Kohler, 1972; see also Welsh & Mohler, 2013; Chapter 25). One notion is that as more realistic visual simulations are developed, less adaptation might occur and this may alleviate the problems associated with VE aftereffects (Stanney et al., 1998). While this seems a reasonable hypothesis for visual adaptation, subtle variations from natural image statistics are still likely to produce adaptation (Falconbridge & Badcock, 2006; Webster, 2011); it may also not hold true for adaptation to mismatches between the senses or the perceptual–motor linkages.

In traditional psychophysical experiments, it is often the case that the variance of the measurements is initially quite high and reduces with continued data collection. It is recognized that subjects often need experience in the paradigm to adopt the best perceptual set and response strategy. Thus, some forms of adaptation to the VE might be expected to occur through perceptual–motor learning and some through changing cognitive problem-solving strategies. These modifications in perceptual set and response strategy would not be expected to result in debilitating aftereffects and might best be considered VE-specific effects. Little research has been carried out on these important but indirect effects on visual processing in VE systems.

REFERENCES

Abrams, R. A., Oonk, H. M., & Pratt, J. (1998). Fixation point offsets facilitate endogenous saccades. *Perception & Psychophysics*, *60*, 201–208.

Adelstein, B. D., Lee, T. G., & Ellis, S. R. (2003, October 13–17). Head tracking latency in virtual environments: Psychophysics and a model. In *Proceedings of the Human Factor Ergonomics Society 47th Annual Meeting* (pp. 2083–2087), HFES, Denver, CO.

Allison, R. S., Harris, L. R., Jenkin, M., Jasiobedzka, U., & Zacher, J. E. (2001, March 13–17). Tolerance of temporal delay in virtual environments. In *Proceedings of the IEEE Virtual Reality Conference* (pp. 247–254), Yokohama, Japan.

Allison, R. S., Howard, I. P., & Zacher, J. E. (1999). Effect of field size, head motion, and rotational velocity on roll vection and illusory self-tilt in a tumbling room. *Perception*, *28*, 299–306.

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, *14*, 105–126.

Amano, K., Edwards, M., Badcock, D. R., & Nishida, S. (2009). Adaptive pooling of visual motion signals by the human visual system revealed with a novel multi-element stimulus. *Journal of Vision*, *9*(3), 1–25.

Ames, A. Jr. (1925). Illusion of depth from single pictures. *Journal of the Optical Society of America*, *10*, 137–148.

Andersen, G. J., & Braunstein, M. L. (1985). Induced self-motion in central vision. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(2), 122–132.

Anstis, S. M. (1980). The perception of apparent movement. In H. C. Longuet-Higgins & N. S. Sutherland (Eds.), *The psychology of vision*. London, U.K.: Royal Society.

Arakawa, Y. (1953). Quantitative measurements of visual fields for colors with a direct-current method. *American Journal of Ophthalmology*, *36*, 1594–1601.

Arsenault, R., & Ware, C. (2004). The importance of stereo, eye coupled perspective and touch for eye-hand coordination. *Presence: Teleoperators and Virtual Environments*, *13*, 549–559.

Ash, A., & Palmisano, S. (2012). Vection during conflicting multisensory information about the axis, magnitude and direction of self-motion. *Perception*, *41*, 253–267.

Ash, A., Palmisano, S., Govan, G., & Kim, J. (2011). Effect of visual display lag on vection in active observers. *Aviation, Space and Environmental Medicine*, *82*, 763–769.

Ash, A., Palmisano, S., & Kim, J. (2011). Perceived self-motion induced by consistent and inconsistent multisensory stimulation. *Perception*, *40*, 155–174.

Aubert, H. (1886). Die Bewegungsempfindung. *Archiv fur die Gesamte Physiologie*, *39*, 347–370.

Badcock, D. R., & Derrington, A. M. (1985). Detecting the displacement of periodic patterns. *Vision Research*, *25*, 1253–1258.

Badcock, D. R., Hess, R. F., & Dobbins, K. (1996). Localization of element clusters: Multiple cues. *Vision Research*, *36*, 1467–1472.

Badcock, D. R., & Schor, C. (1985). The depth increment detection function within spatial channels. *Journal of the Optical Society of America*, *A2*, 1211–1216.

Badcock, D. R., & Westheimer, G. (1985). Spatial location and hyperacuity: The centre/surround localization contribution function has two substrates. *Vision Research*, *25*, 1259–1267.

Baker, C. L., Hess, R. F., & Zihl, J. (1991). Residual motion perception in a "motion-blind" patient, assessed with limited-lifetime random dot stimuli. *The Journal of Neuroscience*, *11*(2), 454–461.

Becker, W. (1989). Metrics. In M. E. Goldburg & R. H. Wurtz (Eds.), *The neurobiology of saccadic eye movements*. Amsterdam, the Netherlands: Elsevier Science Publishers.

Becker, W., Raab, S., & Jürgens, R. (2002). Circular vection during voluntary suppression of optokinetic reflex. *Experimental Brain Research*, *144*(4), 554–557.

Bell, J., & Badcock, D. R. (2008). Luminance and contrast cues are integrated in global shape detection with contours. *Vision Research*, *48*(21), 2336–2344.

Bodenheimer, B., Meng, J., Wu, H., Narasimham, G., Rump, B., McNamara, T. P., Rieser, J. J. (2007). Distance estimation in virtual and real environments using bisection. In C. Wallraven & V. Sunstedt (Eds.), *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization* (pp. 35–40). Tubingen, Germany: ACM.

Boff, K. R., Kaufman, L., & Thomas, J. P. (Eds.). (1986). *Handbook of human perception* (2 Vols.). New York, NY: John Wiley & Sons.

Boff, K. R., & Lincoln, J. E. (1988). *Engineering data compendium: Human perception and performance*. Wright-Patterson Airforce Base, OH: Armstrong Aerospace Medical Research Laboratory.

Bonato, F., & Bubka, A. (2006). Chromaticity, spatial complexity, and self-motion perception. *Perception, 35*(1), 53–64.

Borish, I. (1970). *Clinical refraction* (3rd ed.). Chicago, IL: Professional Press.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178.

Bourdon, B. (1902). *La perception visuelle de l'espace*. Paris, France: Schleicher Freres.

Boynton, R. M., Hayhoe, M. M., & MacLeod, D. I. A. (1977). The gap effect: Chromatic and achromatic visual discrimination as affected by field separation. *Optica Acta*, *24*, 159–177.

Brandt, T., Dichgans, J., & Koenig, E. (1973). Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research*, *16*, 476–491.

Brandt, T., Wist, E. R., & Dichgans, J. (1971). Optisch induzierte pseudo-Coriolis-effekte und circularvektion: Ein beitrag zur optisch-vestibulären interaktion. *Archiv Für Psychiatrie und Nervenkrankheiten*, *214*, 365–389.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*, 719–722.

Brown, W. R. J. (1951). The influence of luminance level on visual sensitivity to color differences. *Journal of the Optical Society*, *41*, 684–688.

Bubka, A., & Bonato, F. (2010). Natural visual-field features enhance vection. *Perception*, *39*, 627–635.

Bubka, A., Bonato, F., & Palmisano, S. (2008). Expanding and contracting optical flow patterns and vection. *Perception*, *37*, 704–711.

Burbeck, C. A., & Hadden, S. (1993). Scaled position integration areas: Accounting for Weber's law for separation. *Journal of the Optical Society of America*, *A10*, 5–15.

Burr, D., & Morrone, M. C. (2005). Eye movements: Building a stable world from glance to glance. *Current Biology*, *15*(20), R839–R840.

Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, *371*, 511–513.

Burr, D. C., Morrone, M. C., & Vaina, L. M. (1998). Large receptive fields for optic flow detection in humans. *Vision Research*, *38*, 1731–1743.

Burr, D. C., & Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Research*, *22*, 479–484.

Burr, D. C., & Thompson, P. (2011). Motion psychophysics: 1985–2010. *Vision Research*, *51*(13), 1431–1456.

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and invariance. *Journal of Neurophysiology*, *98*(3), 1733–1750.

Calvert, G., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes*. New York, NY: MIT Press.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology [London]*, *197*, 551–566.

Cardin, V., & Smith, A. T. (2010). Sensitivity of human visual and vestibular cortical regions to egomotion-compatible visual stimulation. *Cerebral Cortex*, *20*, 1964–1973.

Castiello, U., Badcock, D. R., & Bennett, K. (1999). Sudden and gradual presentation of distractor objects: Differential interference effects. *Experimental Brain Research*, *128*(4), 550–556.

Cavanagh, P., & Favreau, O. E. (1985). Color and luminance share a common motion pathway. *Vision Research*, *25*, 1595–1601.

Chalupa, L. M., & Werner, J. S. (2004). *The visual neurosciences* (Vols. 1 & 2). Cambridge, MA: MIT Press.

Cherici, C., Kuang, X., Poletti, M., & Rucci, M. (2012). Precision of sustained fixation in trained and untrained observers. *Journal of Vision*, *12*(6:31), 1–16.

CIE (1932). *Commission internationale de l'Eclairage proceedings, 1931*. Cambridge, U.K.: Cambridge University Press, 1932.

Clarke, F. J. J. (1967, April). The effect of field-element size on chromaticity discrimination. In *Proceedings of the Symposium on Colour Measurement in Industry, 1*, The Colour Group, London, U.K.

Connor, C. E. (2004). Shape dimensions and object primitives. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (Vol. 2, pp. 1080–1089). London, U.K.: Bradford.

Craik, K. J. W. (1939). The effect of adaptation upon visual acuity. *British Journal of Psychology*, *29*, 252–266.

Cropper, S. J., & Badcock, D. R. (1994). Discriminating smooth from sampled motion: Chromatic and luminance stimuli. *Journal of the Optical Society of America*, *11*, 515–530.

Derrington, A. M., & Badcock, D. R. (1985). Separate detectors for simple and complex grating patterns? *Vision Research*, *25*, 1869–1878.

DeValois, R. L., & DeValois, K. K. (1991). Vernier acuity with stationary moving Gabors. *Vision Research*, *31*, 1619–1626.

Dichgans, J., & Brandt, T. (1978). Visual-vestibular interaction: Effects on self-motion perception and postural control. In *Perception, handbook of sensory physiology* (pp. 756–804). New York, NY: Springer.

Dichgans, J., Koener, F., & Voigt, K. (1969). Verleichende sdalierung des afferenten und efferenten bewegungssehen beim menschen: Lineare funktionen mit verschiedener ansteigssteilheit. *Psychologische Forschung*, *32*, 277–295.

Dickinson, J. E., Broderick, C., & Badcock, D. R. (2009). Selective attention contributes to global processing in vision. *Journal of Vision*, *9*(2), 1–8.

Ding, J., & Levi, D. M. (2011). Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision. *Proceedings of the National Academy of Sciences*, *108*(37), E733–E741.

Draper, M. H, Viirre, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors*, *43*, 129–145.

Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, *65*, 1329–1345.

Durgin, F. H., Gigone, K., & Scott, R. (2005). Perception of visual speed while moving. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 339–353.

Duwaer, A. L., & van den Brink, G. (1981). Diplopia thresholds and the initiation of vergence eye-movements. *Vision Research*, *21*(12), 1727–1737.

Edwards, M., & Badcock, D. R. (1993). Asymmetries in the sensitivity to motion in depth: A centripetal bias. *Perception*, *22*, 1013–1023.

Edwards, M., & Badcock, D. R. (2003). Motion distorts perceived depth. *Vision Research*, *43*, 1799–1804.

Exner, S. (1875). Uber das sehen von bewegungen und die theorie des zusammengesetzen auges. *Sitzungsberichts Akademie Wissen Schaft Wein*, *72*, 156–190.

Falconbridge, M. S., & Badcock, D. R. (2006). Implicit exploitation of regularities: Novel correlations in images quickly alter visual perception. *Vision Research*, *46*, 1331–1335.

Farnsworth, D. (1955). Tritanomalous vision as a threshold function. *Die Farbe*, *4*, 185–197.

Fendick, M. G. (1983). Parameters of the retinal light distribution of a bright line which correspond to its attributed spatial location. *Investigative Ophthalmology and Visual Science*, *24*, 92.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research*, *33*, 173–193.

Field, D. J., & Tolhurst, D. J. (1986). The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proceedings of the Royal Society of London B: Biological Sciences*, *228*, 379–400.

Filehne, W. (1922). Ueber das optische wahenehmen von bewegungen. *Zeitschrift fur Sinnesphysiologie*, *53*, 134–145.

Fischer, M. H., & Kornmüller, A. E. (1930). Optokinetisch ausgelöste bewegungswahrehmungen und optokinetischer nystagmus. *Journal für Psychologie und Neurologie*, *41*, 273–308.

Forbes, B., & Klein, R. M. (1996). The magnitude of the fixation offset effect with endogenously and exogenously controlled saccades. *Journal of Cognitive Neuroscience*, *8*, 344–352.

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

Gilbert, C. D., Das, A., Ito, M., Kapadia, M., & Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proceedings of the National Academy of Science*, *93*(2), 615–622.

Gillam, B. (1995). The perception of spatial layout from static optical information. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 23–70). New York, NY: Academic Press.

Gillam, B., Palmisano, S., & Govan, D. G. (2011). Motion parallax and binocular disparity for depth interval estimation beyond interaction space. *Perception*, *40*, 39–49.

Gouras, P. (1991). The perception of color. *Vision and visual dysfunction* (Vol. 6). London, U.K.: Macmillan.

Graham, C. H., Baker, K. E., Hecht, M., & Lloyd, V. V. (1948). Factors influencing thresholds for monocular movement parallax. *Journal of Experimental Psychology*, *38*, 205–223.

Graham, N. (1980). Spatial frequency channels in human vision: Detecting edges without edge detectors. In C. Harris (Ed.), *Visual coding and adaptability* (pp. 215–262). Hillsdale, NJ: Lawrence Erlbaum Associations.

Gregory, R. L. (1964). Human perception. *British Medical Bulletin*, *20*, 21–26.

Gregory, R. L. (1966). *Eye and brain*. London, U.K.: World University Library.

Grim, K. (1911). Uber diie genauigkeit der wahrnehmung und ausfuhrung von augenbeweg unger. *Zeitschrift fur Sinnesphysiologie*, *45*, 9–26.

Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and surface perception. *Vision Research*, *49*, 2666–2685.

Hayhoe, M. M., Levin, M. E., & Koshel, R. J. (1992). Subtractive processes in light adaptation. *Vision Research*, *32*(2), 323–333.

Held, R., Dichgans, J., & Bauer, J. (1975). Characteristics of moving visual areas influencing spatial orientation. *Vision Research*, *15*, 357–365.

Helmholtz, H. (1867/1925). *Physiological optics* (3rd ed., Vol. 3). Menasha, WI: The Optical Society of America.

Hess, R. F., & Badcock, D. R. (1995). Metric for separation discrimination by the human visual system. *Journal of the Optical Society of America*, *A12*, 3–16.

Hettinger, L. G. (2014). Illusory self-motion in virtual environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments* (2nd ed., pp. 437–468). New York: CRC Press.

Hood, D. C., & Finkelstein, M. A. (1979). Comparison of changes in sensitivity and sensation: Implications for the response-intensity function of the human photopic system. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 391–405.

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

Hurvich, L. M. (1981). *Color vision*, p. 210. Sunderland: Sinauer.

Ito, H., & Shibata, I. (2005). Self-motion perception from expanding and contracting optical flows overlapped with binocular disparity. *Vision Research*, *45*, 397–402.

Johansson, G. (1977). Studies on visual-perception of locomotion. *Perception*, *6*, 365–376.

Johnson, W. H., Sunahara, F. A., & Landolt, J. P. (1999). Importance of the vestibular system in visually induced nausea and self-vection. *Journal of Vestibular Research – Equilibrium & Orientation*, *9*, 83–87.

Jones, J. A., Swan, J. E., Singh, G., Kolstad, E., & Ellis, S. R. (2008). The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In S. Creem-Regehr & K. Myszkowski (Eds.), *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization* (pp. 9–14). Los Angeles, CA: ACM.

Kalloniatis, M., & Luu, C. (2007). Webvision: The organization of the retina and visual system: Perception of depth. webvision.med.utah.edu/book/part-viii.gabac-receptors/perception-of-depth.

Kelly, D. H., & Burbeck, C. A. (1980). Motion and vision. III. Stabilized pattern adaptation. *Journal of the Optical Society of America*, *70*, 1283–1289.

Kelly, D. H., & Burbeck, C. A. (1984). Critical problems in spatial vision. *Critical Reviews in Biomedical Engineering*, *10*, 125–177.

Kennedy, H., Magnin, M., & Jeannerod, M. (1976). Receptive field response of LGB neurons during vestibular stimulation in awake cats. *Vision Research*, *16*, 119–120.

Kim, J., & Palmisano, S. (2008). Effects of active and passive viewpoint jitter on vection in depth. *Brain Research Bulletin*, *77*, 335–342.

Kim, J., Palmisano, S., & Bonato, F. (2012). Simulated angular head oscillation enhances vection in depth. *Perception*, *41*, 402–414.

Kitazaki, M., & Sato, T. (2003). Attentional modulation of self-motion perception. *Perception*, *32*, 475–484.

Knapp, J. M., & Loomis, J. M. (2004). Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators and Virtual Environments*, *13*, 572–577.

Koenderink, J. J., Bouman, M. A., Bueno de Mesquita, A. E., & Slappendel, S. (1978). Perimetry of contrast detection thresholds of moving spatial sine wave patterns II. The far peripheral visual field (eccentricity 0°–50°). *Journal of the Optical Society of America*, *68*, 850–854.

Koffka, K. (1922). Perception: An introduction to Gestalt theorie. *Psychological Bulletin*, *19*, 531–585.

Kohler, I. (1972). Experiments with goggles. In R. Held & W. Richards (Eds.), *Perception: Mechanisms and models* (p. 390). San Francisco, CA: W. H. Freeman & Co.

Kovacs, I., Papathomas, T. V., Yang, M., & Feher, A. (1996). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences*, *93*, 15508–15511.

Lawson, B. D., & Riecke, B. E. (2014). Perception of body motion. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments* (2nd ed., pp. 163–196). New York, NY: CRC Press.

Lee, C., Oakley, I., & Ryu, J. (2012). Exploring the impact of visual-haptic registration accuracy in augmented reality. In P. Isokoski & J. Springare (Eds.), *Haptics: Perception, devices, mobility, and communication* (pp. 85–90). Heidelberg, Germany: Springer Verlag.

Lee, D. N. (1980). The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London*, *290*, 169–179.

Lee, D. N., & Lishman, J. R. (1975). Visual proprioceptive control of stance. *Journal of Human Movement Studies*, *1*, 89–95.

Leibowitz, H. W. (1955). The relation between the rate threshold for the perception of movement and luminance for various durations of exposure. *Journal of Experimental Psychology*, *49*, 209–214.

Lestienne, F., Soechting, J., & Berthoz, A. (1977) Postural readjustments induced by linear motion of visual scenes. *Experimental Brain Research*, *28*, 363–384.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654.

Levi, D. M., & Klein, S. A. (1990). The role of separation and eccentricity in encoding position. *Vision Research*, *30*, 557–585.

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, *25*, 963–977.

Levi, D. M., & Schor, C. M. (1984). Spatial and velocity tuning of processes underlying induced motion. *Vision Research*, *24*, 1189–1195.

Loffler, G. (2008). Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research*, *48*(20), 2106–2127.

Loomis, J. M., Fujita, N., Da Silva, J. A., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 906–921.

Lowther, K., & Ware, C. (1996). Vection with large screen 3D imagery. In *ACM CHI '96* (pp. 233–234). New York, NY: ACM.

Ludvigh, E., & Miller, J. W. (1958). Study of visual acuity during the ocular pursuit of moving test objects I. Introduction. *Journal of the Optical Society of America*, *48*, 799–802.

Mach, E. (1875). *Grundlinien der Lehre von der Bewegungsempfindungen*. Leipzig, Germany: Engelmann.

Mach, E. (1922). *Die Analyse der Empfindungen [The analysis of sensations]*. Gena: Gustav Fischer.

Mack, A. (1986). Perceptual aspects of motion in the frontal plane. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Sensory processes and perception* (Vol. 1), pp. 17.1–17.38. New York, NY: John Wiley & Sons.

Mania, K., Adelstein, B. D., Ellis, S. R., & Hill, M. I. (2004). Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In V. Interrante & A. McNamara (Eds.), *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* (pp. 39–48). Los Angeles, CA: ACM.

Masland, R. H., & Albright, T. D. (2008). *The senses: A comprehensive reference* (Vols. 1 and 2). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-012370880-9.09003-4.

Mather, G., Verstraten, F., & Anstis, S. (Eds.). (1998). *The motion aftereffect: A modern perspective*. London, U.K.: MIT Press.

Matin, L. (1986). Visual localization and eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of human perception and performance* (Vol. 1, pp. 1–45). New York, NY: John Wiley & Sons.

McCann, J. J., & Hall, J. A. (1980). Effects of average-luminance surrounds on the visibility of sine-wave gratings. *Journal of the Optical Society of America*, *70*(2), 212–219.

McGraw, P. V., Walsh, V., & Barrett, B. T. (2004). Motion-sensitive neurones in V5/MT modulate perceived spatial position. *Current Biology*, *14*, 1090–1093.

McKee, S. P., & Levi, D. M. (1987). Dichoptic hyperacuity: The precision of nonius alignment. *Journal of the Optical Society of America*, *A4*, 1104–1108.

Meehan, M., Razzaque, S., Whitton, M. C., & Brookes, F. P. Jr. (2003). Effect of latency on presence in stressful virtual environments. In *VR '03: Proceedings of the IEEE Conference on Virtual Reality* (pp. 141–148). Washington, DC: IEEE Computer Society.

Messing, R., & Durgin, F. H. (2005). Distance perception and the visual horizon in head-mounted displays. *ACM Transactions of Applied Perception*, *2*, 234–250.

Metha, A. B., Vingrys, A. J., & Badcock, D. R. (1993). Calibration of a colour monitor for visual psychophysics. *Behavior Research Methods, Instruments, and Computers*, *25*(3), 371–383.

Millodot, M., Johnson, C. A., Lamont, A., & Leibowitz, H. A. (1975). Effect of dioptrics on peripheral visual acuity. *Vision Research*, *15*, 1357–1362.

Mitchell, D. E., & Ware, C. (1974). Interocular transfer of a visual after-effect in normal and stereoblind humans. *Journal of Physiology [London]*, *236*, 707–721.

Mon-Williams, M., Wann, J. P., & Rushton, S. (1993). Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics*, *13*, 387–391.

Moss, J. D., Muth, E. R., Tyrrell, R. A., & Stephens, B. R. (2010). Perceptual thresholds for display lag in a real visual environment are not affected by field of view or psychophysical technique. *Displays*, *31*, 143–149.

Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *Journal of Physiology*, *359*, 381–400.

Nakamura, S., Seno, T., Ito, H., & Sunaga, S. (2010). Coherent modulation of stimulus colour can affect visually induced self-motion perception. *Perception*, *39*, 1579–1590.

Nakamura, S., & Shimojo, S. (1999). Critical role of foreground stimuli in perceiving visually induced self-motion (vection). *Perception*, *28*, 893–902.

Nakamura, S., & Shimojo, S. (2000). A slowly moving foreground can capture an observer's self-motion a report of new motion illusion: Inverted vection. *Vision Research*, *40*, 2915–2923.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping and the recognition of occluded objects. *Perception*, *18*, 55–68.

Nandy, A. S., & Tjan, B. S. (2012). Saccade-confounded image statistics explain visual crowding. *Nature Neuroscience*, *15*(3), 463–469.

Nishida, S. (2011). Advancement of motion psychophysics: Review 2001–2010. *Journal of Vision*, *11*(5:11), 1–53.

Ohmi, M., & Howard, I. P. (1988). Effect of stationary objects on illusory forward self-motion induced by a looming display. *Perception*, *17*, 5–11.

Ohmi, M., Howard, I. P., & Landolt, J. P. (1987). Circular vection as a function of foreground-background relationships. *Perception*, *16*, 17–22.

Palmer, S. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Palmisano, S. (1996). Perceiving self-motion in depth: The role of stereoscopic motion and changing-size cues. *Perception & Psychophysics*, *58*, 1168–1176.

Palmisano, S. (2002). Consistent stereoscopic information increases the perceived speed of vection in depth. *Perception*, *31*, 463–480.

Palmisano, S., Allison, R. S., & Howard, I. P. (2006). Illusory scene distortion occurs during perceived self-rotation in roll. *Vision Research*, *46*, 4048–4058.

Palmisano, S., Allison, R. S., Kim, J., & Bonato, F. (2011). Simulated viewpoint jitter shakes sensory conflict accounts of self-motion perception. *Seeing and Perceiving*, *24*, 173–200.

Palmisano, S., Allison, R. S., & Pekin, F. (2008). Accelerating self-motion displays produce more compelling vection in depth. *Perception*, *37*, 22–33.

Palmisano, S., & Gillam, B. J. (1998). Stimulus eccentricity and spatial frequency interact to determine circular vection. *Perception*, *27*, 1067–1078.

Palmisano, S., Gillam, B. J., & Blackburn, S. (2000). Global perspective jitter improves vection in central vision. *Perception*, *29*, 57–67.

Palmisano, S., Gillam, B. J., Govan, D., Allison, R. S., & Harris, J. (2010). Stereoscopic perception of real depths at large distances. *Journal of Vision*, *10*(6), 1–16.

Palmisano, S., & Kim, J. (2009). Effects of gaze on vection from jittering, oscillating and purely radial optic flow. *Attention, Perception & Psychophysics*, *71*, 1842–1853.

Pelah, A., & Barlow, H. B. (1996). Visual illusion from running. *Nature*, *6580*, 381.

Pelz, J. B., Ballard, D. H., & Hayhoe, M. M. (1993). Memory use during the performance of natural visuomotor tasks. *Investigative Ophthalmology and Visual Science [Suppl.]*, *34*(4), 1234.

Plumert, J. M., Kearney, J. K., Cremer, J. F., & Recker, K. (2005). Distance perception in real and virtual environments. *ACM Transactions on Applied Perception*, *2*, 216–233.

Post, R. B. (1988). Circular vection is independent of stimulus eccentricity. *Perception*, *17*, 737–744.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing*. Cambridge, England: Cambridge University Press.

Reason, J. T., Mayes, A. R., & Dewhurst, D. (1982). Evidence for a boundary effect in roll vection. *Perception & Psychophysics*, *31*(2), 139–144.

Reason, J., Wagner, H., & Dewhurst, D. (1981). A visually driven postural aftereffect. *Acta Psychologica*, *48*, 241–251.

Redlick, F. P., Harris, L. R., & Jenkin, M. (1999). Active motion reduces the perceived self-displacement created by optic flow. *Investigative Ophthalmology & Visual Science*, 40, S798–S798.

Regan, D., & Beverley, K. I. (1973). The dissociation of sideways movements from movements in depth: Psychophysics. *Vision Research*, *13*, 2403–2415.

Regan, D., & Beverley, K. I. (1978a). Illusory motion in depth: Aftereffect of adaptation to changing size. *Vision Research*, *18*, 209–212.

Regan, D., & Beverley, K. I. (1978b). Looming detectors in the human visual pathway. *Vision Research*, *18*, 415–421.

Regan, D., & Tyler, C. W. (1971). Temporal summation and its limit for wavelength changes: An analog of Bloch's law for color vision. *Journal of the Optical Society of America*, *61*, 1414–1421.

Riecke, B. E. (2009). Cognitive and higher-level contributions to illusory self-motion perception ("vection"): Does the possibility of actual motion affect vection? *Japanese Journal of Psychonomic Science*, *28*, 135–139.

Riecke, B. E., Schulte-Pelkum, J., Avraamides, M. N., Heyde, M. V. D., & Bülthoff, H. H. (2006). Cognitive factors can influence self-motion perception (vection) in virtual reality. *ACM Transactions on Applied Perception*, *3*, 194–216.

Riecke, B. E., Väljamäe, A., & Schulte-Pelkum, J. (2009). Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Transactions on Applied Perception*, *6*, 1–27.

Riggs, L. A., Armington, J. C., & Ratliff, F. (1954). Motions of the retinal image during fixation. *Journal of the Optical Society of America*, *44*, 315–321.

Rivest, J., & Cavanagh, P. (1996). Localizing contours defined by more than one attribute. *Vision Research*, *36*, 53–66.

Robson, J. G. (1966). Spatial and temporal contrast-sensitivity functions of the visual system. *Journal of the Optical Society of America*, *56*, 1141–1142.

Rock, I. (1975). *An introduction to perception*. London, U.K.: Macmillan.

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, *22*, 261–270.

Rubin, G. S., West, S. K., Munoz, B., Bandeen-Roche, K., Zeger, S., Schein, O., & Fried, L. P. (1997). A comprehensive assessment of visual impairment in a population of older Americans. The SEE study. Salisbury Eye Evaluation Project. *Investigative Ophthalmology & Visual Science*, *38*(3), 557–568.

Rushton, S., Mon-Williams, M., & Wann, J. (1994). Binocular vision in a bi-ocular world: New generation head-mounted displays avoid causing visual deficit. *Displays*, *15*, 255–260.

Sandström, C. I. (1951). *Orientation in the present space*. Stockholm, Sweden: Almgvist & Wiksell.

Scarfe, P., & Johnston, A. (2010). Motion drag induced by global motion Gabor arrays. *Journal of Vision*, *10*(5:14), 1–15.

Schreiber, K. M., Hillis, J. M., Filippini, H. R., Schor, C. M., & Banks, M. S. (2008). The surface of the empirical horopter. *Journal of Vision*, *8*(3:7), 1–20.

Schutz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A review. *Journal of Vision*, *11*(5:9), 1–30.

Seno, T., Ito, H., & Sunaga, S. (2009). The object and background hypothesis for vection. *Vision Research*, *49*, 2973–2982.

Seno, T., Ito, H., & Sunaga, S. (2011a). Attentional load inhibits vection. *Attention, Perception & Psychophysics*, *73*, 1467–1476.

Seno, T., Ito, H., & Sunaga, S. (2011b). Inconsistent locomotion inhibits vection. *Perception*, *40*, 747–750.

Seno, T., Ogawa, M., Ito, H., & Sunaga, S. (2011). Consistent air flow to the face facilitates vection. *Perception*, *40*, 1237–1240.

Seno, T., Sunaga, S., & Ito, H. (2010). Inhibition of vection by red. *Attention, Perception, & Psychophysics*, *72*, 1642–1653.

Sony PVM-2551MD Medical OLED monitor (2011). http://pro.sony-asia.com/product/resources/en_BA/images/Brochure/Medical/BP000460_pvm-2551md.pdf

Shlaer, S. (1937). The relation between visual acuity and illumination. *Journal of General Physiology*, *21*, 165–188.

Siegal, M. H., & Siegal, A. B. (1972). Hue discrimination as a function of luminance. *Perception & Psychophysics*, *12*, 295–299.

So, R. H., & Griffin, M. J. (1995). Head-coupled virtual environment with display lag. In K. Carr & R. England (Eds.), *Simulated and virtual realities: Elements of perception* (pp. 103–111). London, U.K.: Taylor & Francis Group.

Solman, R. T., Dain, S. J., Lim, H. S., & May, J. G. (1995). Reading-related wavelength and spatial frequency effects in visual spatial location. *Ophthalmic and Physiological Optics*, *15*, 125–132.

Solman, R. T., & May, J. G. (1990). Spatial localization discrepancies: A visual deficiency in poor readers. *American Journal of Psychology*, *103*(2), 243–263.

Stanney, K. M., Salvendy, G., Deisigner, J., DiZio, P., Ellis, S., Ellison, E., … Witmer, B. (1998). Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. Report sponsored by the Life Sciences Division at NASA Headquarters. *International Journal of Human-Computer Interaction*, *10*(2), 135–187.

Stark, L., Kong, R., Schwartz, S., Hendry, D., & Bridgeman, B. (1976). Saccadic suppression of image displacement. *Vision Research*, *16*, 1185–1187.

Telford, L., & Frost, B. J. (1993). Factors affecting the onset and magnitude of linear vection. *Perception & Psychophysics*, *53*, 682–692.

Thomas, J. P. (1975). Spatial resolution and spatial interaction. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 5). New York, NY: Academic Press.

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, *22*, 377–380.

Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging the distance in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, *13*, 560–571.

To, M. P. S., Regan, B. C., Wood, D., & Mollon, J. D. (2011). Vision out of the corner of the eye. *Vision Research*, *51*, 203–214.

Toet, A., van-Eekhout, M. P., Simons, H. L., & Koenderink, J. J. (1987). Scale invariant features of differential spatial displacement discrimination. *Vision Research*, *27*, 441–451.

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, *10*(11), 502–511.

Tschermak, A. (1931). Optischer raumsinn (optical sense of space). In A. Bethe, G. Bergnann, G. Emden, & A. Ellinger (Eds.), *Handbuch der normalen und pathologischen physiologie* (pp. 824–1000). Leipzig, Germany: Springer.

Turano, K. A., & Heidenreich, S. M. (1999). Eye movements affect the perceived speed of visual motion. *Vision Research*, *39*, 1177–1188.

Tynan, P., & Sekuler, R. (1982). Motion processing in peripheral vision: Reaction time and perceived velocity. *Vision Research*, *22*, 61–68.

Uchikawa, K., & Ikeda, M. (1981). Temporal deterioration of wavelength discrimination with successive comparison method. *Vision Research*, *21*, 591–595.

van den Berg, R., Roerdink, J. B. T. M., & Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation and hue compared to orientation. *Journal of Vision*, *7*(2), 1–11.

Van Ness, F. L., & Bouman, M. A. (1967). Variation of contrast sensitivity with luminance. *Journal of the Optical Society of America*, *57*, 401–406.

Viguier, A., Clement, G., & Trotter, Y. (2001). Distance perception within near visual space. *Perception*, *30*(1), 115–124.

Wallach, H., & Kravitz, J.-H. (1965). Rapid adaptation in the constancy of visual direction with active and passive rotation. *Psychonomic-Science*, *3*(4), 165–166.

Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates.

Wang, P., & Nikolic, D. (2011). An LCD monitor with sufficiently precise timing for research in vision. *Frontiers in Human Neuroscience*, *5*(85), 1–10.

Wann, J. P., White, A. D., Wilkie, R. M., Culmer, P. R., Lodge, J. P. A., & Mon-Williams, M. (2014). Measurement of visual aftereffects following virtual environment exposure. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of Virtual Environments* (2nd ed., pp. 809–852). Boca Raton, FL: CRC Press.

Watson, A. B., Ahumada, A. J., & Farrell, J. E. (1986). Window of visibility: A psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A-Optics and Image Science*, *3*, 300–307.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America*, *A4*, 2006–2021.

Watt, R. J., Morgan, M. J., & Ward, R. M. (1983). The use of different cues in vernier acuity. *Vision Research*, *23*, 991–995.

Webster, M. A. (2011). Adaptation and visual coding. *Journal of Vision*, *11*(5:3), 1–23.

Wertheim, A. H. (1994). Motion perception during self-motion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences*, *17*, 293–355.

Westheimer, G. (1978). Spatial phase sensitivity for sinusoidal grating targets. *Vision Research*, *18*, 1073–1074.

Westheimer, G. (1982). The spatial grain of the perifoveal visual field. *Vision Research*, *22*, 157–162.

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research*, *17*, 941–947.

Whitney, D. (2002). The influence of visual motion on perceived position. *Trends in Cognitive Sciences*, *6*(5), 211–216.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168.

Willemsen, P., Colton, M. B., Creem-Regehr, S. H., & Thompson, W. B. (2009). The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Transactions on Applied Perception*, *6*, 1–14.

Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, *9*, 79–97.

Wilson, H. R., & Wilkinson, F. (1998). Detection of global structure in Glass patterns: Implications for form vision. *Vision Research*, *38*, 2933–2947.

Wright, W. G., DiZio, P., & Lackner, J. R. (2006). Perceived self-motion in two visual contexts: Dissociable mechanisms underlie perception. *Journal of Vestibular Research*, *16*, 23–28.
Wyszecki, G., & Stiles, W. S. (1967). *Color science: Concepts and methods, quantitative data and formulae*. New York, NY: John Wiley & Sons.
Zacharias, G. L., & Young, L. R. (1981). Influence of combined visual and vestibular cues on human perception and control of horizontal rotation. *Experimental Brain Research*, *41*, 159–171.
Zikovitz, D. C., Jenkin, M., & Harris, L. R. (2001). Comparison of stereoscopic and non-stereoscopic optic flow displays. *Journal of Vision*, *1*(3), 317.

This page intentionally left blank

4 Virtual Auditory Displays

Michael Vorländer and Barbara Shinn-Cunningham

CONTENTS

4.1 INTRODUCTION

Auditory processing is often given little attention when designing virtual environments (VEs) or simulations. This lack of attention is unfortunate because auditory cues play a crucial role in everyday life. Auditory cues increase awareness of surroundings, cue visual attention, and convey a variety of complex information without taxing the visual system. The entertainment industry has long recognized the importance of sound to create ambience and emotion, aspects that are often lacking in VEs. Placing someone in a virtual world with an improperly designed auditory interface is equivalent to creating a *virtual* hearing impairment for the user, making them less aware of their surroundings, and contributing to feelings of isolation.

Auditory perception, especially localization, is a complex phenomenon affected by physiology, expectation, and even the visual interface. This chapter will consider different methods for creating auditory interfaces. As will be discussed, spatialized audio using headphones or transaural systems is necessary to create compelling sound, and spatialized sound offers the sound engineer the greatest amount of control over the auditory experience of the listener. Multichannel audio systems can produce virtual sound events in 3D to some extent, but they require complex equipment and signal processing. For many applications, especially using projection screens, standard stereo speaker systems may be simpler to implement and provide benefits not available to headphone systems. Properly designed speaker systems, especially using subwoofers, may contribute to emotional context. The positives and negatives associated with each option will be discussed.

It is impossible to include everything that needs to be known about designing auditory interfaces in a single chapter. The current aim is to provide a starting point, laying out the essential theory behind implementing sound in a VE without overwhelming the novice designer. Instead of trying to review all perceptual and technical issues related to creating virtual auditory displays, this chapter focuses on fundamental aspects of spatial auditory perception and the generation of spatial auditory cues in VEs. Specifically, the chapter begins by introducing basic properties of sound and discussing the perception of these sound properties (psychoacoustics), with a special emphasis on spatial hearing. General techniques for producing auditory stimuli, both with and without spatial cues, are then considered (see Letowski et al., 2001, for a lexicon for understanding auditory displays). Unlike the visual channel, very little effort has been put into formulating theories concerning the creation of synthetic sound sources in VEs; the question of how to generate realistic sounds (rather than using sources from some precomputed, stored library of source sounds) is beyond the scope of this chapter. In addition, the technology involved in producing spatialized audio is rapidly changing, with new products introduced all the time as others disappear, so that any specific recommendations would quickly be dated. However, an overview of current technology and solutions is presented at the conclusion of the chapter.

4.1.1 Why Are Virtual Auditory Interfaces Important?

4.1.1.1 Environmental Realism and Ambience

If it does nothing else, an auditory interface should convey basic information about the VE to the user. For instance, in the real world, pedestrians walking through a shopping area are aware of everything from the sound of their own footsteps to the sounds of other shoppers to the mechanical sounds from cash registers, scanners, escalators, and other machines. In *control room* situations such as nuclear power plants, air traffic control centers, or the bridge of a ship, sounds such as alarms, switches being toggled, and verbal communications with other people in the room (including sounds of stress or uncertainty) provide vital information for participants. The location of these voices, switches, and alarms also provides information concerning their function and importance. In the absence of these basic auditory cues, situational awareness is severely degraded. The same is true in VEs.

The entertainment industry has recognized that sound is a vital aspect of creating ambience and emotion for films. George Lucas, best recognized by the public for stunning visual effects in his movies, has stated that sound is 50% of the movie experience (THX Certified Training Program, 2000). In VEs, the argument is often erroneously made that sound is secondary, since the visual image of a police car chasing down a city street can be compelling on its own. However, without appropriate sounds (squealing tires, a police siren, the tortured breathing of the driver, etc.), the emotional impact of a simulation is muted. The sound quality of footsteps depends on whether you are in grass, on pavement, or in a hallway. Likewise, the sound of one's own voice differs depending on whether you are inside a room or in an open field. These are the types of things that create ambience and feeling in film; the same is true in VEs.

4.1.1.2 Presence/Immersion and Perceived Simulation Quality

Presence (Chertoff & Schatz, 2013; Chapter 34, this book) can be defined as the "sense of being immersed in a simulation or virtual environment." Such a nebulous concept is difficult to quantify. Although definitive evidence is lacking, it is generally believed that the sense of presence is dependent on auditory, visual, and tactile fidelity (Sheridan, 1996). Referring back to the previous section, it can be inferred that as environmental realism increases, the sense of presence increases. However, although realism probably contributes to the sense of presence, it is not necessarily true that an increased sense of presence results in a greater sense of realism. Specifically, although virtual or spatial audio does not necessarily increase the perceived realism of a VE, it does increase the sense of presence (Hendrix & Barfield, 1996). Thus, if implemented properly, appropriately designed audio increases the overall sense of presence in a VE or simulation. Indeed, using medium- and high-quality auditory displays can enhance the perceived quality of visual displays. Inversely, using low-quality auditory displays reduces the perceived quality of visual displays (Storms, 1998).

4.1.1.3 Selective Auditory Attention

In a multisource sound environment, it is easier to segregate, attend to, and comprehend sound sources if they are separated in space, something known as the *cocktail party effect* (Cherry, 1953; Shinn-Cunningham, 2008; Yost, 2006). This ability to direct selective auditory attention, which enables a listener to process whatever sound source is most important at a given moment, is critical in many common situations such as teleconferencing (Begault, 1999) or multichannel radio communications (Begault, 1993; Begault & Wenzel, 1992; Haas, Gainer, Wightman, Couch, & Shilling, 1997). Even when spatial sound cues are imperfect (which can degrade sound localization accuracy), they can improve communication in multichannel situations (Drullman & Bronkhorst, 2000; Shinn-Cunningham, Ihlefeld, & Satyavarta, 2005).

4.1.1.4 Spatial Auditory Displays

While graphical displays are an obvious choice for displaying spatial information to a human operator (particularly after considering the spatial acuity of the visual channel), the visual channel is often overloaded, with operators monitoring a myriad of dials, gauges, and graphic displays. In these cases, spatial auditory cues can provide invaluable information to an operator, particularly when the visual channel is saturated (Begault, 1993; Bronkhorst, Veltman, & van Breda, 1996; Shilling & Letowski, 2000). Spatial auditory displays are also being developed for use in applications for which visual information provides no benefit, for instance, in limited field-of-view (FOV) applications or when presenting information to the blind. In command/control applications, the primary goal is to convey unambiguous information to the human operator. In such situations, realism, per se, is not useful, except to the extent that it makes the operator's task easier (i.e., reduces the workload); however, spatial resolution is critical. In these applications, signal-processing schemes that could enhance the amount of information transferred to the human operator may be useful, even if the result is *unnatural*, as long as the user is able to extract this information (e.g., see Durlach, Shinn-Cunningham, & Held, 1993). It should be noted that when designing spatialized auditory displays for noisy environments such as cockpits, electronic noise cancellation technology should be employed and user's hearing loss taken into account to make certain that the displayed information is perceptible to the user (Begault, 1996). Also, for high-g environments, more work needs to be conducted to discover the contribution of g-forces to displacements in sound localization (e.g., Clark & Graybiel, 1949; DiZio, Held, Lackner, Shinn-Cunningham, & Durlach, 2001).

4.1.1.5 Cross-Modal Interactions

The importance of multimodal interactions involving the auditory system cannot be ignored (Popescu et al., 2014, Chapter 17; Simpson Cowgill, Gilkey, & Weisenberger, 2014, Chapter 13). A whole range of studies show that judgments about one sensory modality are influenced by information in other sensory modalities. For instance, localized auditory cues reduce response times to visual targets (Frens, van Opstal, & van der Willigen, 1995; Perrott, Saberi, Brown, & Strybel, 1990; Perrott, Sadralodabai, Saberi, & Strybel, 1991). Similarly, the number of auditory events affects the perceived number of visual events occurring at that time (e.g., see Shams, Kamitani, & Shimojo, 2004). Even noninformative auditory cues can improve accuracy of perception of visual motion (Kim, Peters, & Sham, 2012), demonstrating the power of cross-modal perceptual effects. Auditory cues also augment or even substitute for tactile and/or visual information about events that are difficult to perceive in these other modalities, such as visual information outside a limited FOV. Through such cross-modal interactions, auditory cues can play an important role in conveying information that may, superficially, seem to be more naturally communicated through some other sensory channel.

4.2 PHYSICAL ACOUSTICS

4.2.1 Properties of Sound

Sound is a pressure wave produced when an object vibrates rapidly back and forth. The diaphragm of a speaker produces sound by pushing against molecules of air, thus creating an area of high pressure (*condensation*). As the speaker's diaphragm returns to its resting position, it creates an area of low pressure (*rarefaction*). This localized disturbance travels through the air as a wave of alternating low pressure and high pressure at approximately 344 m/s or 1128 ft/s (at 70°F), depending on temperature and humidity.

4.2.1.1 Frequency

If the musical note "A" is played as a pure sinusoid, there will be 440 condensations and rarefactions per second. The distance between two adjacent condensations or rarefactions, typically represented by the symbol λ, equals the wavelength of the sound wave. The velocity at which the sound wave is traveling is denoted as c. The time one full oscillatory cycle (condensation through rarefaction) takes is called the frequency (f) and is expressed in Hertz or cycles per second. The relationship between frequency, velocity, and wavelength is given by $f = c/\lambda$.

From a modeling standpoint, this relationship is important when considering the Doppler shift. As a sound source is moving toward a listener, the perceived frequency increases because the wavelength is compressed as a function of the velocity (v) of the moving source. This compression can be explained by the equation $\lambda = (c - v)/f$. For negative velocities (i.e., for sources moving away), this expression describes a relative increase in the wavelength (and a concomitant decrease in frequency).

4.2.1.2 Strength

The amplitude of the waveform determines the intensity of a sound stimulus. It should not be confused, however, with the sound intensity defined in physics as the sound energy propagating per second through a reference area. The meaning of intensity in the context of this chapter is simply the meaning of strength. It is measured in decibels (dB). Decibels give the level of sound (on a logarithmic scale) relative to some reference level. One common reference level is 2×10^{-5} N/m^2. Decibels referenced to this value are commonly used to describe sound intensity expressed in units of dB sound pressure level (SPL). The sound level in dB SPL can be computed by the following equation:

$$\text{dB SPL} = 20 \log_{10}\left(\frac{\text{RMS sound pressure}}{20 \times 10^{-6}\ \text{N/m}^2}\right)$$

The threshold of hearing is in the range of 0–10 dB SPL for most sounds, although the actual threshold depends on the spectral content of the sound. When measuring sound strength in the *real world*, it is measured with a sound pressure meter. Most sound pressure meters allow one to collect sound-level information using different scales that weight energy in different frequencies differently in order to approximate the sensitivity of the human auditory system to sound at low-, moderate-, or high-intensity levels. These scales are known as A, B, and C weighted scales, respectively. The B scale is rarely used; however, the C scale (dBC) is useful for evaluating noise levels in high-intensity environments such as traffic noise and ambient cockpit noise. The frequency response of the dBC measurement is closer to an unfiltered (flat) response than dBA. In fact, when conducting *sound surveys* in a complex noise environment, it is prudent to measure sound level in both dBA and flat response (or dBC) to make an accurate assessment of the audio environment.

Frequency, intensity, and complexity are physical properties of an acoustic waveform. The perceptual analogs for frequency, intensity, and complexity are pitch, loudness, and timbre, respectively. Although the distinction between physical and perceptual measures of sound properties is an important one, both physical and perceptual descriptions are important when designing auditory displays.

4.3 PSYCHOPHYSICS

The basic sensitivity of the auditory system is reviewed in detail in a number of textbooks (e.g., see Gelfand, 1998; Moore, 1997; Yost, 2006; Zwicker & Fastl, 2007). This section provides a brief overview of some aspects of human auditory sensitivity that are important to consider when designing auditory VEs.

4.3.1 Frequency Analysis in the Auditory System

In the cochlea, acoustic signals are broken down into constituent frequency components by a mechanical Fourier-like analysis. Along the length of the cochlea, the frequency to which that section of the cochlea responds varies systematically from high to low frequencies. The strength of neural signals carried by the auditory nerve fibers arrayed along the length of the cochlea varies with the mechanical displacement of the corresponding section of the cochlea. As a result, each nerve fiber can be thought of as a frequency channel that conveys information about the energy and timing of the input signal within a restricted frequency region. At all stages of the auditory system, these multiple frequency channels are evident.

Although the bandwidth changes with the level of the input signal and with input frequency, to a crude first-order approximation, one can think of the frequency selectivity of the auditory system as constant on a log-frequency basis (approximately one-third octave wide). Thus, a particular auditory nerve responds to acoustic energy at and near a particular frequency.

Humans are sensitive to acoustic energy at frequencies between about 20 and 22,000 Hz. Absolute sensitivity varies with frequency. Humans are most sensitive to energy at frequencies around 2000 Hz and are less sensitive for frequencies below and above this range.

The fact that input waveforms are deconstructed into constituent frequencies affects all aspects of auditory perception. Many behavioral results are best understood by considering the activity of the auditory nerve fibers, each of which responds to energy within about a third of an octave of its particular *best frequency*. For instance, the ability to detect a sinusoidal signal in a noise background degrades dramatically when the noise spectrum is within a third octave of the sinusoid frequency. When a noise is spectrally remote from a sinusoidal target, it causes much less interference with the detection of the sinusoid. These factors are important when one considers the spectral content of different sounds that are to be used in an auditory VE. For instance, if one must monitor multiple kinds of alerting sounds, choosing signals that are spectrally remote from one another will improve a listener's ability to detect and respond to different signals.

4.3.2 Intensity Perception

Listeners are sensitive to sound intensity on a logarithmic scale. For instance, doubling the level of a sound source causes roughly the same perceived change in the loudness independent of the reference level. This logarithmic sensitivity to sound intensity gives the auditory system a large dynamic range. For instance, the range between just detectable sound levels and sounds that are so loud that they cause pain is roughly 110–120 dB (i.e., an increase in sound pressure by a factor of a million). The majority of the sounds encountered in everyday experience span a dynamic intensity range of 80–90 dB. Typical sound reproduction systems use 16 bits to represent the pressure of the acoustic signal (providing a useful dynamic range of about 90 dB), which is sufficient for most simulations.

While sound intensity (a physical measure) affects the loudness of a sound (a perceptual measure), loudness does not grow linearly with intensity. In addition, the same decibel increase in sound intensity can result in different increments in loudness, depending on the frequency content of the sound. Thus, intensity and loudness, while closely related, are not equivalent descriptions of sound.

4.3.3 Masking Effects

As mentioned earlier, when multiple sources are presented to a listener simultaneously or in rapid succession, the sources interfere with one another in various ways. For instance, a tone that is audible when played in isolation may be inaudible when a loud noise is presented simultaneously. Such effects (known as *masking* effects) arise from a variety of mechanisms, from physical interactions of the separate acoustic waves impinging on the ear to high-level, cognitive factors. For a more complete description of these effects than is given in the following, see Yost (2006, pp. 153–167) or Moore (1997, pp. 111–120).

Simultaneous masking occurs when two sources are played concurrently. However, signals do not have to be played simultaneously for them to interfere with one another perceptually. For instance, both *forward* masking (in which a leading sound interferes with perception of a trailing sound) and *backward* masking (in which a lagging sound interferes with perception of a leading sound) occur. Generally speaking, many simultaneous and forward-masking effects are thought to arise from peripheral interactions that occur at or before the level of the auditory nerve. For instance, the mechanical vibrations of the basilar membrane are nonlinear, so that the response of the membrane to two separate sounds may be less than the sum of the response to the individual sounds. These nonlinear interactions can suppress the response to what would (in isolation) be an audible event.

Other, more central factors influence masking as well. For instance, backward masking may reflect higher-order processing that limits the amount of information extracted from an initial sound in the presence of a second sound. The term *informational masking* refers to all masking that cannot be explained by peripheral interactions in the transduction of sound by the auditory periphery. Most such effects can be traced to problems with segregating a source of interest from other, competing sources, problems with identifying which source in a sound mixture is the most important (*target*) source, or some combination thereof (Shinn-Cunningham, 2008). These failures of selective auditory attention can have significant impact on perception, even for sounds that are clearly audible. For instance, perceptual sensitivity in discrimination and detection tasks is often degraded when there is uncertainty about the characteristics of a target source (e.g., see Yost, 2006, pp. 219–220).

4.3.4 Pitch and Timbre

Just as sound intensity is the physical correlate of the percept of loudness, source frequency is most closely related to the percept of pitch. For sound waves that are periodic (including pure sinusoids, for instance), the perceived *pitch* of a sound is directly related to the inverse of the period of the sound signal. Thus, sounds with low pitch have relative long periods and sounds with high pitch generally have short periods. Many real-world sounds are not strictly periodic in that they have a temporal pattern that repeats over time but has fluctuations from one cycle to the next. Examples of such pseudo-periodic signals include the sound produced by a flute or a vowel sound spoken aloud. The perceived pitch of such sounds is well predicted by the average period of the cyclical variations in the stimulus.

The percept of pitch is not uniformly strong for all sound sources. In fact, nonperiodic sources such as noise do not have a salient pitch associated with them. For relatively narrow sources that are aperiodic, or for band-limited noise, a weak percept of pitch can arise that depends on the center frequency or the cutoff frequency of the spectral energy of the signal, respectively. In fact, perceived pitch is affected by a wide variety of stimulus attributes, including temporal structure, frequency content, harmonicity, and even loudness. Although the pitch of a pure sinusoid is directly related to its frequency, there is no single physical parameter that can predict perceived pitch for more complex sounds. Nonetheless, for many sounds, pitch is a very salient and robust perceptual feature that can be used to convey information to a listener. For instance, in music, pitch conveys melody. In speech, pitch conveys a variety of information (ranging from the gender of a speaker to paralinguistic, emotional content of a speech). Pitch is also a very important cue for segregating competing sound sources and allowing a listener to focus selective auditory attention on a target source (e.g., see Carlyon, 2004).

The percept of timbre is the sound property that enables a listener to distinguish an oboe from a trumpet. Like pitch, the percept of timbre depends on a number of physical parameters of sound, including spectral content and temporal envelope (such as the abruptness of the onset and offset of sound). Like pitch, timbre is an important property for enabling listeners to identify a target source and thus can be used to convey information through an auditory display (e.g., see Brewster, Wright, & Edwards, 1993). However, sounds with different timbres have different perceptual weight, a factor that should be considered in designing discrete sounds for an auditory display (e.g., see Chon & McAdams, 2012). As with pitch, timbre is a feature that allows listeners to direct selective auditory attention to a desired target amidst competing sounds, enabling them to extract desired information from that source despite the presence of interfering information (e.g., Maddox & Shinn-Cunningham, 2011).

4.3.5 Temporal Resolution

The auditory channel is much more sensitive to temporal fluctuations in sensory inputs than either visual (Badcock, Palmisano, & May, 2013, Chapter 3) or proprioceptive (Dindar, Tekalp, & Basdogan, 2013, Chapter 5; Lawson & Riecke, 2014, Chapter 7) channels. For instance, the auditory system can detect amplitude fluctuations in input signals up to 50 Hz (i.e., a duty cycle of 20 ms) very easily (e.g., see Yost, 2006, pp. 146–149). Sensitivity degrades slowly with increasing modulation rate, so that some sensitivity remains even as the rate approaches 1000 Hz (i.e., temporal fluctuations at a rate of 1 per ms). The system is also sensitive to small fluctuations in the spectral content of an input signal for roughly the same modulation speeds. Listeners not only can detect rapid fluctuations in an input stimulus, but they can react quickly to auditory stimuli. For instance, reaction times to auditory stimuli are faster than visual reaction times by 30–40 ms (an improvement of roughly 20%; e.g., see Welch & Warren, 1986).

4.3.6 Spatial Hearing

Spatial acuity of the auditory system is far worse than that of the visual (Badcock et al., 2013, Chapter 3) or proprioceptive (Dindar et al., 2013, Chapter 5; Lawson & Riecke, 2013, Chapter 7) systems (for a review, see Middlebrooks & Green, 1991). For a listener to detect an angular displacement of a source from the median plane, the source must be displaced laterally by about a degree. For a source directly to the side, the listener does not always detect a lateral displacement of 10°. Auditory spatial acuity is even worse in other spatial dimensions. A source in the median plane must be displaced by as much as 15° for the listener to perceive the directional change accurately. While listeners can judge relative changes in source distance, absolute distance judgments are often surprisingly inaccurate even under the best of conditions (e.g., see Zahorik, Brungart, & Bronkhorst, 2005).

Functionally, spatial auditory perception is distinctly different from that of the other *spatial* senses of vision and proprioception. For the other spatial senses, position is neurally encoded at the most peripheral part of the sensory system. For instance, the photoreceptors of the retina are organized topographically so that a source at a particular position (relative to the direction of gaze) excites a distinct set of receptors (Badcock et al., 2013, Chapter 3). In contrast, spatial information in the auditory signals reaching the left and right ears of a listener must be computed from the peripheral neural representations. The way in which spatial information is carried by the acoustic signals reaching the eardrums of a listener has been the subject of much research. This section provides a brief review of how acoustic attributes convey spatial information to a listener and how the perceived position of a sound source is computed in the brain (for more complete reviews, see Blauert, 1997; Middlebrooks & Green, 1991; Mills, 1972; Wightman & Kistler, 1993).

4.3.6.1 Head-Related Transfer Functions

The pairs of spatial filters that describe how sound is transformed as it travels through space to impinge on the left and right ears of a listener are known as head-related transfer functions (HRTFs). HRTFs describe how to simulate the direct sound reaching the listener from a particular position but do not generally include any reverberant energy. Empirically measured HRTFs vary mainly with the direction from head to source but also vary with source distance (particularly for sources within reach of the listener). For sources beyond about a meter away, the main effect of distance is just to change the overall gain of the HRTFs. In the time domain, the HRTF pair for a particular source location provides the pressure waveforms that would arise at the ears if a perfect impulse were presented from the spatial location in question. Often, HRTFs are represented in the frequency domain by taking the Fourier transform of time-domain impulse responses.

HRTFs contain most of the spatial information present in real-world listening situations. In particular, binaural cues are embodied in the relative phase and magnitude (respectively) of the linear filters for the left and right ears. Spectral cues and source intensity are present in the absolute frequency-dependent magnitudes of the two filters.

Figure 4.1 shows two HRTF pairs from a human subject in the time domain (left side of figure) and in the frequency domain (magnitude only, right side of figure). All panels are for a source at

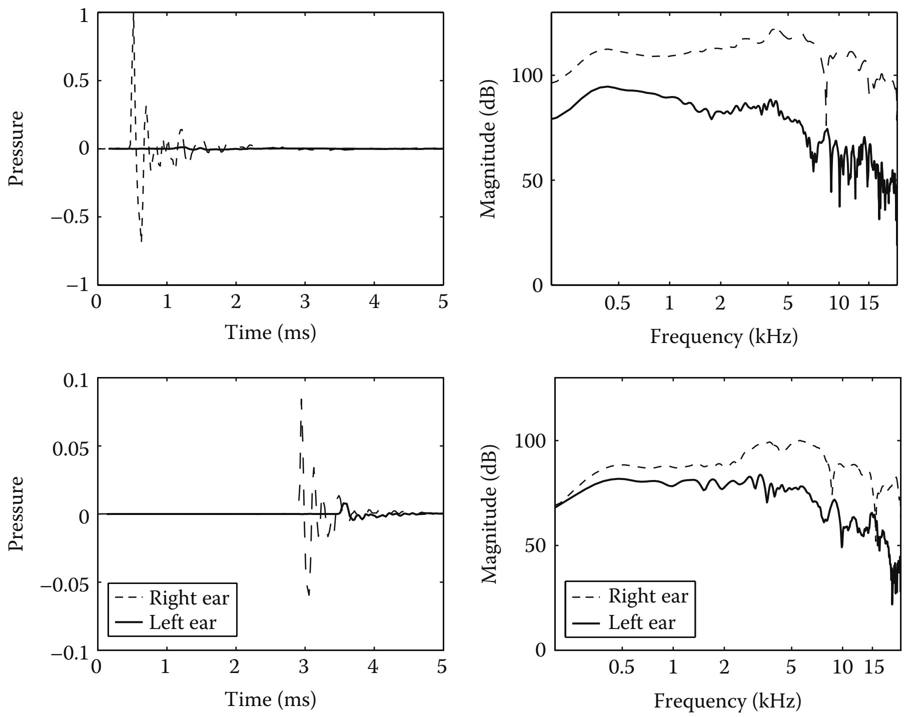

FIGURE 4.1 Time-domain (left panels) and magnitude spectra (right panels) representations of anechoic HRTFs for a human subject. All panels show a source at 90 azimuth and 0 elevation. Top panels are for a source at 15 cm, and bottom panels for a source at 1 m.

azimuth 90 and elevation 0. The top two panels show the HRTF for a source very close to the head (15 cm from the center of the head). The bottom two panels show the HRTF for a source 1 m from the head. In the time domain, it is easy to see the interaural differences in time and intensity, while the frequency-domain representation shows the spectral notches that occur in HRTFs, as well as the frequency-dependent nature of the interaural level difference (ILD). The ILDs are larger in all frequencies for the nearer source (top panels) as expected. In the time domain, the 1 m source must traverse a greater distance to reach the ears than the near source, resulting in additional time delay before the energy reaches the ears (note time-onset differences in the impulse responses in the left top and left bottom panels).

4.3.6.2 Binaural Cues

The most robust cues for source position depend on differences between the signals reaching the left and right ears. Such *interaural* or *binaural* cues are robust specifically because they can be computed by comparing the signals reaching each ear. As a result, binaural cues allow a listener to factor out those acoustic attributes that arise from source content from those attributes that arise from source position.

Depending on the angle between the interaural axis and a sound source, one ear may receive a sound earlier than the other. The resulting interaural time differences (ITDs) are the main cue indicating the laterality (left/right location) of the direct sound. The ITD grows with the angle of the source from the median plane; for instance, a source directly to the right of a listener results in an ITD of 600–800 μs favoring the right ear. ITDs are most salient for sound frequencies below about 2 kHz but occur at all frequencies in a sound. At higher frequencies, listeners use ITDs in signal *envelopes* to help determine source laterality but are insensitive to differences in the interaural phase of the signal.

Listeners can reliably detect ITDs of 10–100 μs (depending on the individual listener), which grossly correspond to ITDs that would result from a source positioned 1°–10° from the median plane. Sensitivity to changes in the ITD deteriorates as the reference ITD gets larger. For instance, the smallest detectable change in ITD around a reference source with an ITD of 600–800 μs (corresponding to the ITD of a source far to the side of the head) can be more than a factor of 2 larger than for a reference with zero ITD.

At the high end of the audible frequency range, the head of the listener reflects and diffracts signals so that less acoustic energy reaches the far side of the head (causing an *acoustic head shadow*). Due to the acoustic head shadow, the relative intensity of a sound at the two ears varies with the lateral location of the source. The resulting ILDs generally increase with source frequency and angle between the source and median plane. ILDs are perceptually important for determining source laterality for frequencies above about 2 kHz.

When a sound source is within reach of a listener, extralarge ILDs (at all frequencies) arise due to differences in the relative distances from source to left and right ears (e.g., see Brungart & Rabinowitz, 1999; Duda & Martens, 1998). These additional ILDs are due to differences in the relative distances from source to left and right ears and help to convey information about the relative distance and direction of the source from the listener (Shinn-Cunningham, Santarelli, & Kopco, 2000). Other low-frequency ILDs that may arise from the torso appear to help determine the elevation of a source (Algazi, Avendano, & Duda, 2001). Most listeners are able to detect ILDs of 0.5–1 dB, independent of source frequency.

The perceived location of a sound source usually is consistent with the ITD and ILD information available. However, there are multiple source locations that cause roughly the same ITD and ILD cues. For sources more than a meter from the head, the locus of such points is approximately a hyperbolic surface of rotation symmetric about the interaural axis that is known as the *cone of confusion* (see left side of Figure 4.2). When a sound is within reach of the listener, extralarge ILDs provide additional robust, binaural information about the source location. For a simple spherical

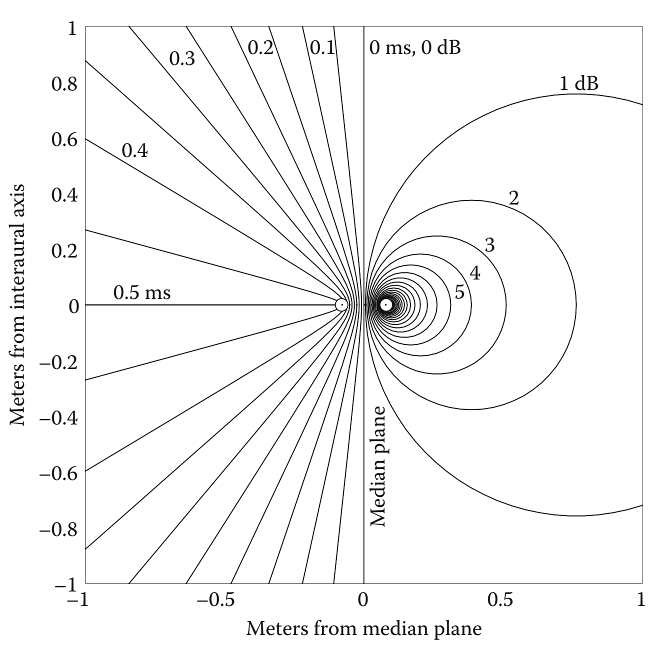

FIGURE 4.2 Iso-ITD (left side of figure) and iso-ILD (right side of figure) contours for sources near the head. On the left, sources at each location along a contour give rise to nearly the same ITD. On the right, sources at each location along a contour give rise to nearly the same unique near-field ILD component.

head model, low-frequency ILDs are constant on spheres centered on the interaural axis (see right side of Figure 4.2). The rate at which the extralarge ILD changes with spatial location decreases as sources move far from the head or near the median plane. In fact, once a source is more than a meter or so from the head, the contribution of this *near-field* ILD is perceptually insignificant (Shinn-Cunningham et al., 2000). In general, positions that give rise to the same binaural cues (i.e., the intersection of constant ITD and ILD contours) form a circle centered on the interaural axis (Shinn-Cunningham et al.). Since ITD and ILD sensitivity is imperfect, the locus of positions that cannot be resolved from binaural cues may be more accurately described as a *torus of confusion* centered on the interaural axis. Such tori of confusion degenerate to the more familiar cones of confusion for sources more than about a meter from a listener.

4.3.6.3 Spectral Cues

The main cue to resolve source location on the torus of confusion is the spectral content of signals reaching the ears. These spectral cues arise due to interactions of the outer ear (pinna) with the impinging sound wave that depend on the relative position of sound source and listener's head (Batteau, 1967). Spectral cues only occur at relatively high frequencies, generally above 6 kHz. Unlike interaural cues for source location, spectral cues can be confused with changes in the spectrum of the source itself. Perhaps because of this ambiguity, listeners are more likely to make localization errors in which responses fall near the correct torus of confusion but are not in the right direction. Individual differences in spectral filtering of the pinnae are large and are important when judging source direction (e.g., see Wenzel, Arruda, Kistler, & Wightman, 1993).

4.3.6.4 Anechoic Distance Cues

In general, the intensity of the direct sound reaching a listener (i.e., sound that does not come off of reflective surfaces like walls) decreases with the distance of the source. In addition, the atmosphere absorbs energy in high, audible frequencies as a sound propagates, causing small changes in the spectrum of the received signal with changes in source distance. If a source is unfamiliar, the intensity and spectrum of the direct sound are not robust cues for distance because they can be confounded with changes in the intensity or spectral content (respectively) of the signal emitted from the source. However, even for unfamiliar sources, overall level and spectral content provide relative distance information (Mershon, 1997).

4.3.6.5 Reverberation

Reverberation (acoustic energy reaching a listener from indirect paths, via walls, floors, etc.) generally has little affect on or degrades the perception of source direction (e.g., see Begault, 1993; Hartmann, 1983; Shinn-Cunningham, 2000b). However, it actually aids in the perception of source distance (e.g., see Mershon, Ballenger, Little, McMurty, & Buchanan, 1989; Shinn-Cunningham, Kopco, & Santarelli, 1999). At least grossly, the intensity of reflected energy received at the ears is independent of the position of the source relative to the listener (although it can vary dramatically from one room to another). As a result, the ratio of direct to reverberant energy provides an absolute measure of source distance for a given listening environment.

Reverberation not only provides a robust cue for source distance, but it also provides information about the size and configuration of a listening environment. For instance, information about the size and *spaciousness* of a room can be extracted from the pattern of reverberation in the signals reaching the ears. While many psychophysical studies of sound localization are performed in anechoic (or simulated anechoic) environments, reverberation is present (in varying degrees) in virtually all everyday listening conditions. Anechoic environments (such as those used in many simulations and experiments) seem subjectively unnatural and strange to naive listeners. Conversely, adding reverberation to a simulation causes all sources to seem more realistic and provides robust information about relative source distance (e.g., see Begault, Wenzel, Lee, & Anderson, 2012; Brungart & D'Angelo, 1995). While reverberation may improve distance perception and improve the realism of a display, it can decrease accuracy in directional perception, albeit slightly, and may interfere with the ability to extract information in a source signal (e.g., degrade speech) and to attend to multiple sources (e.g., see Section 4.3.6.8).

4.3.6.6 Dynamic Cues

In addition to *static* acoustic cues like ITD and ILD, changes in spatial cues with source or listener movement also influence perception of source position and help to resolve torus-of-confusion ambiguities (e.g., see Wallach, 1940; Wightman & Kistler, 1989). For instance, a source either directly in front or directly behind a listener would cause near-zero ITDs and ILDs; however, a leftward rotation of the head results in ITDs and ILDs favoring either the right ear (for a source in front) or the left ear (for a source behind).

While the auditory system generally has good temporal resolution, the temporal resolution of the binaural hearing system is much poorer. For instance, investigations into the perception of moving sound sources imply that binaural information averaged over a time window lasting 100–200 ms results in what has been termed *binaural sluggishness* (e.g., see Grantham, 1997; Kollmeier & Gilkey, 1990).

4.3.6.7 Effects of Stimulus Characteristics on Spatial Perception

Characteristics of a source itself affect auditory spatial perception in a number of ways. For instance, the bandwidth of a stimulus can have a large impact on both the accuracy and precision of sound localization. As a result, one must consider how nonspatial attributes of a source in a VE will impact spatial perception of a signal. In cases where one can design the acoustic signal (i.e., if the signal is a warning signal or some other arbitrary waveform), these factors should be taken into consideration when one selects the source signal.

For instance, the spectral filtering of the pinnae cannot be determined if the sound source does not have sufficient bandwidth. This makes it difficult to unambiguously determine the location of a source on the torus of confusion for a narrowband signal. Similarly, if a source signal does not have energy above about 5 kHz, spectral cues will not be represented in the signals reaching the ears and errors along the torus of confusion are more common (e.g., Gilkey & Anderson, 1995).

Ambiguity in narrowband source locations arises in other situations as well. For instance, narrowband, low-frequency signals in which ITD is the main cue can have ambiguity in their heard location because the auditory system is only sensitive to interaural phase. Thus, a low-frequency sinusoid with an ITD of half cycle favoring the right ear may also be heard far to the left side of the head. However, binaural information is integrated across frequency so that ambiguity in lateral location is resolved when interaural information is available across a range of frequencies (Brainard, Knudsen, & Esterly, 1992; Stern & Trahiotis, 1997; Trahiotis & Stern, 1989). When narrowband sources are presented, the heard location is strongly influenced by the center frequency of the source (Middlebrooks, 1997).

While spectral bandwidth is important, temporal structure of a source signal is also important. In particular, onsets and offsets in a signal make source localization more accurate, particularly when reverberation and echoes are present. A gated or modulated broadband noise will generally be more accurately localized in a reverberant room (or simulation) than a slowly gated broadband noise (e.g., see Rakerd & Hartmann, 1985, 1986).

4.3.6.8 Top-Down Processes in Spatial Perception

Experience with or knowledge of the acoustics of a particular environment also affects auditory localization, and implicit learning and experience affects performance (e.g., see Clifton, Freyman, & Litovsky, 1993; Shinn-Cunningham, 2000b). In other words, spatial auditory perception is not wholly determined by stimulus parameters but also by the state of the listener. Although such effects are not due to conscious decision, they can measurably alter auditory localization and spatial perception. For instance, when localizing a sound followed by a later *echo* of that sound, the influence of the later sound diminishes with repetition of the sound pairing, as if the listener learns to discount the lagging echo (Freyman, Clifton, & Litovsky, 1991).

4.3.6.9 Benefits of Binaural Hearing

Listeners benefit from receiving different signals at the two ears in a number of ways. As discussed earlier, ITD and ILD cues allow listeners to determine the location of sound sources. However, in addition to allowing listeners to locate sound sources in the environment, binaural cues allow the listener to selectively attend to sources coming from a particular direction. This ability is extremely important when there are multiple competing sources in the environment (e.g., see Shinn-Cunningham, 2008).

Imagine a situation in which there is both a speaker (whom the listener is trying to attend) and a competing source (that is interfering with the speaker). If the speaker and competitor are both directly in front of the listener, the competitor degrades speech reception much more than if the competitor is off to one side, spatially separated from the speaker. This *binaural advantage* arises in part because when the competitor is off to one side of the head, the energy from the competitor is attenuated at the far ear. As a result, the signal-to-noise ratio at the far ear is larger than when the competitor is in front. In other words, the listener has access to a cleaner signal in which the speaker is more prominent when the speaker and noise are spatially separated. However, the advantage of the spatial separation is even larger than can be predicted on the basis of energy.

A homologous benefit can be seen under headphones. In particular, if one varies the level of a signal until it is just detectable in the presence of a masker, the necessary signal level is much lower when the ITD of the signal and masker are different than when they are the same. The difference between these thresholds, referred to as the masking level difference (MLD), can be as large as 10–15 dB for some signals (e.g., see Durlach & Colburn, 1978; Zurek, 1993).

The binaural advantage affects both signal detection (e.g., see Gilkey & Good, 1995) and speech reception (e.g., see Bronkhorst & Plomp, 1988). It is one of the main factors contributing to the ability of listeners to monitor and attend multiple sources in complex listening environments (i.e., the *cocktail party effect*; see, e.g., Shinn-Cunningham, 2008; Yost, 1997). Thus, the binaural advantage is important for almost any auditory signal of interest. In order to get these benefits of binaural hearing, signals reaching a listener must have appropriate ITDs and/or ILDs.

4.3.6.10 Adaptation to Distorted Spatial Cues

While a naive listener responds to ITD, ILD, and spectral cues based on their everyday experience, listeners can learn to interpret cues that are not exactly like those that occur naturally. For instance, listeners can learn to adapt to unnatural spectral cues when given sufficient long-term exposure (Hofman, Van Riswick, & Van Opstal, 1998). Short-term training allows listeners to learn how to map responses to spatial cues to different spatial locations than normal (Shinn-Cunningham, Durlach, & Held, 1998). These studies imply that for applications in which listeners can be trained, *perfect* simulations of spatial cues may not be necessary. However, there are limits to the kinds of distortions of spatial cues to which a listener can adapt (e.g., see Shinn-Cunningham, 2000a).

4.3.6.11 Intersensory Integration of Spatial Information

Acoustic spatial information is integrated with spatial information from other sensory channels (particularly vision) to form spatial percepts (e.g., see Welch & Warren, 1986). In particular, auditory spatial information is combined with visual (and/or proprioceptive) spatial information to form the percept of a single, multisensory event, especially when the inputs to the different modalities are correlated in time (e.g., see Popescu et al., 2013; Warren, Welch, & McCarthy, 1981; Chapter 17). When this occurs, visual spatial information is much more potent than that of auditory information, so the perceived location of the event is dominated by the visual spatial information (although auditory information does affect the percept to a lesser degree, e.g., see Pick, Warren, & Hay, 1969;

Welch & Warren, 1980). *Visual capture* refers to the perceptual dominance of visual spatial information, describing how the perceived location of an auditory source is captured by visual cues.

Summarizing these results, it appears that the spatial auditory system computes source location by combining all available acoustic spatial information. Perhaps even more importantly, a priori knowledge and information from other sensory channels can have a pronounced effect on spatial perception of auditory and multisensory events.

4.3.7 Auditory Scene Analysis

Listeners in real-world environments are faced with the difficult problem of listening to many competing sound sources that overlap in both time and/or frequency. The process of separating out the contributions of different sources to the total acoustic signals reaching the ears is known as *auditory scene analysis* (e.g., see Bregman, 1990; Carlyon, 2004).

In general, the problem of grouping sound energy across time and frequency to reconstruct each sound source is governed by a number of basic (often intuitive) principles. For instance, naturally occurring sources are often broadband, but changes in the amplitude or frequency of the various components of a single source are generally correlated over time. Thus, comodulation of sound energy in different frequency bands tends to group these signal elements together and cause them to fuse into a single perceived source. Similarly, temporal and spectral proximity both tend to promote grouping so that signals close in time or frequency are grouped into a single perceptual source (sometimes referred to as a stream). Spatial location can also influence auditory scene analysis such that signals from the same or similar locations are grouped into a single stream. Other factors affecting streaming include (but are not limited to) harmonicity, timbre, and frequency or amplitude modulation (Bregman, 1990; Darwin, 1997).

For the development of auditory displays, these grouping and streaming phenomena are very important because they can directly impact the ability to detect, process, and react to a sound. For instance, if a masker sound is played simultaneously with a target sound, the ability to process the target is significantly worse if the target is heard as just one component of a single sound source comprised of the masker plus the target; when the target is heard as a distinct sound source, a listener is much better at both detecting the target's presence and extracting meaning from the target. Such *grouping* effects cannot be explained solely by peripheral mechanisms, since many times, the target sound is represented faithfully in activity on the auditory nerve. Instead, such effects arise from *central* limitations (e.g., see Shinn-Cunningham, 2008).

4.3.8 Speech Perception

Arguably the most important acoustic signal is that of speech. The amount of information transmitted via speech is larger than any other acoustic signal. For many applications, accurate transmission of speech information is the most critical component of an auditory display.

Speech perception is affected by many of the low-level perceptual issues discussed in previous sections. For instance, speech can be masked by other signals, reducing a listener's ability to determine the content of the speech signal. Speech reception in noisy environments improves if the speaker is located at a different position than the noise source(s), particularly if the speaker and masker are at locations giving rise to different ILDs. Speech reception is also affected by factors that affect the formation of auditory streams, such as comodulation, harmonic structure, and related features. However, speech perception is governed by many high-level, cognitive factors that do not apply to other acoustic signals. For instance, the ability to perceive a spoken word improves dramatically if it is heard within a meaningful sentence rather than in isolation. Speech information is primarily conveyed by sound energy between 200 and 5000 Hz. For systems in which speech communication is critical, it is important to reduce the amount of interference in this range of frequencies or it will impede speech reception.

4.4 SPATIAL SIMULATION

Spatial auditory cues can be simulated using headphone displays or loudspeakers. Headphone displays generally allow more precise control of the spatial cues presented to a listener, both because the signals reaching the two ears can be controlled independently and because there is no indirect sound reaching the listener (i.e., no echoes or reverberation). However, headphone displays are generally more expensive than loudspeaker displays and may be impractical for applications in which the listener does not want to wear a device on the head. While it is more difficult to control the spatial information reaching a listener in a loudspeaker simulation, loudspeaker-based simulations are relatively simple and inexpensive to implement and do not physically interfere with the listener.

Simulations using either headphones or speakers can vary in complexity from providing no spatial information to providing nearly all naturally occurring spatial cues. This section reviews both headphone and speaker approaches to creating spatial auditory cues.

4.4.1 Room Modeling

HRTFs generally do not include reverberation or echoes, although it is possible to measure binaural transfer functions (known as binaural room transfer functions) that incorporate the acoustic effects of a room. While possible, such approaches are generally not practical because such filters vary with listener and source position in the room as well as the relative position of listener and source to produce a combinatorially large number of transfer functions. In addition, such filters can be an order of magnitude longer than traditional HRTFs, increasing both computational and storage requirements of the system.

There has been substantial effort devoted to developing computational models for room reverberation, including high-quality commercial software packages (e.g., see www.odeon.dk or www.catt.se). The required computations are quite intensive; in order to simulate each individual reflection, one must calculate the distance the sound wave has traveled, how the waveform was transformed by every surface on which it impinged, and the direction from which it is arriving at the head. The resulting waveform must then be filtered by the appropriate anechoic HRTF based on the direction of incidence with the head (Vorländer, 2008, pp. 141–146).

If one looks at the resulting reflections as a function of time from the initial sound, the number of reflections in any given time slice increases quadratically with time. At the same time, the level of each individual reflection decreases rapidly, both due to energy absorption at each reflecting surface and increased path length from source to ear. Moreover, those reflections lose their coherence due to surface scattering and edge diffraction. Although second- or third-order reflections may be individually resolvable, higher-order reflections occur so densely in time that the distinct specular content of each *echo* becomes practically irrelevant. Instead, from a certain transition time (which depends on the size of the room and on the surface corrugations), the reflections are smeared in time and heavily overlap to the point that they are well approximated as a so-called *diffuse* sound field. Therefore, many simulations only *spatialize* a relatively small number of the loudest, earliest-arriving reflections (e.g., up to second or third order) and then add random noise that dies off exponentially in time (uncorrelated at the two ears) to simulate later-arriving reflections that are dense in time and arriving from essentially random directions. Even with such simplifications, the computations necessary to generate *realistic* reverberation (particularly in a system that tries to account for movement of a listener) are a challenge; however, with today's computational power, such simulations are feasible and produce plausible results.

Figure 4.3 shows the room impulse response at the right ear for a source located at 45 azimuth, 0 elevation, and distance of 1 m. This impulse response was measured in a moderate-size classroom in which significant reverberant energy persists for as long as 450 ms. The initial few milliseconds of the response are shown in the inset. In the inset, the initial response is that caused by the sound

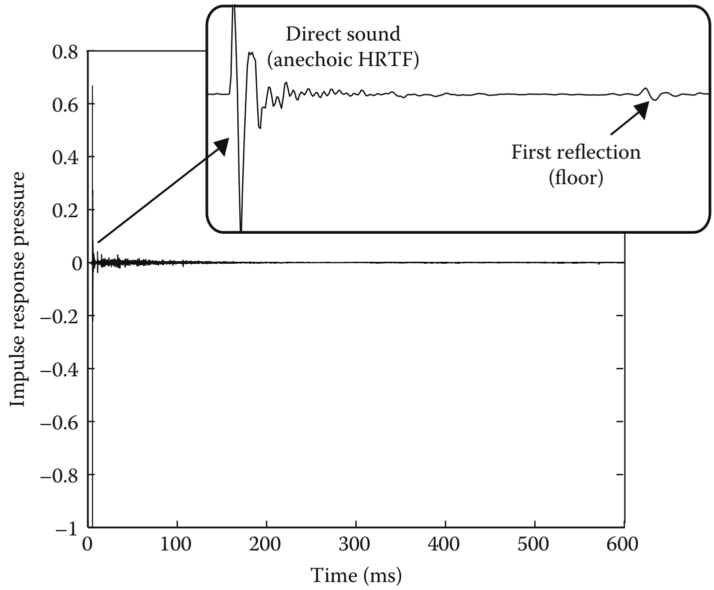

FIGURE 4.3 Impulse response at the right ear for a source at 45 azimuth, 0 elevation, and distance 1 m in a standard classroom. Inset shows first 10 ms of total impulse response.

wave that travels directly from the source to the ear. The first reflection is also evident at the end of the inset, at a much reduced amplitude. In the main figure, the decay of the reverberant energy can be seen with time.

The development of tractable reverberation algorithms for real-time systems has achieved some promising results but is still an ongoing area of research. Most algorithms are based on geometrical acoustics. They are known as hybrid specular and diffuse reflection models and combination of early (mainly specular) and late (mainly diffuse) components (e.g., see Vorländer, 2008, pp. 216–226).

Wave models are also in use in order to compute wave effects such as eigenmodes or diffraction (Aretz, Maier, & Vorländer, 2010; Botteldooren, 1995; Savioja, 2010). Usually this is implemented as a second level into the hybrid models so that the low-frequency range and the mid- and high-frequency range are treated by wave models and geometric models, respectively.

The rendering process using geometrical acoustics is based on a 3D room model consisting of polygons defining the room boundaries. These polygons represent the reflecting surfaces, and accordingly, they are tagged with acoustic absorption and scattering coefficients. For the simulation of reflections, the amplitude and the time of arrival must be calculated. For this, geometrical construction methods are used, which result in identifying the reflection paths between the reflection boundaries of sound propagation in the room.

As soon as the reflection components of the room impulse response are found, they must be convolved with the directional-dependent HRTF and with the sound source stimulus—the so-called *dry* signal.

One crucial part of real-time systems is a rapid calculation of the geometric reflection paths using the polygon model of the room. This task can be highly accelerated by using tree structures in the polygon database (Schröder & Lentz, 2006) or other search algorithms such as spatial hashing (e.g., Schröder, Ryba, & Vorländer, 2010) or frustum tracing (Chandak, Lauterbach, Taylor, Ren, & Manocha, 2008).

Another crucial part of the sound rendering process is real-time convolution of the binaural impulse response with the sound source. Room impulse responses are typically of some seconds in length. The corresponding binaural room impulse response is a finite impulse response (FIR) filter.

For achieving appropriate immersion, the latency of the audio output cannot exceed more than a few milliseconds. Discrete convolution in the time domain and synchronous audio output may be straightforward conceptually but is inefficient in the number of operations required to compute the output samples one by one. Fast Fourier transform (FFT)-based convolution is more efficient, computationally, but requires block processing; this, in turn, leads to block-size-dependent latency in the output. One approach to solving these problems is to use nonuniform segmented block convolution algorithms (Garcia, 2002; Wefers & Vorländer, 2012) or infinite impulse response (IIR) filters.

The binaural filters can be interpreted in the time domain as a temporal series of reflections. Alternatively, they can be interpreted in the frequency domain as a binaural transfer function. The extent to which listeners are sensitive to interaural (temporal) and spectral details in reverberant energy is also not well understood and requires additional research. Nonetheless, there is clear evidence that the inclusion of reverberation can have a dramatic impact on the subjective realism of a virtual auditory display and can aid in perception of source distance (e.g., see Brungart & D'Angelo, 1995).

4.4.2 Headphone Simulation

In order to simulate any source somewhere in space or in a room over headphones, one must simply play a stereo headphone signal that recreates at the eardrums the exact acoustic waveforms that would actually arise from a source at the desired location. This is generally accomplished by empirically measuring the HRTF that describes how an acoustic signal at a particular location in space is transformed as it travels to and impinges on the head and ears of a listener. Then, in order to simulate an arbitrary sound source at a particular location, the appropriate transfer functions are used to filter the desired (known) source signal. The resulting stereo signal is then corrected to compensate for transfer characteristics of the display system (for instance, to remove any spectral shaping of the headphones) and presented to the listener. This holds for a single source at a particular direction or for numerous sources or room reflections, provided they have been filtered with each particular directional HRTF.

4.4.2.1 Diotic Displays

The simplest headphone displays present identical signals to both ears (*diotic* signals). With a diotic display, all sources are perceived as inside the head (not *externalized*), at midline. This internal sense of location is known as *lateralization* not *localization* (Plenge, 1974). While a diotic display requires no spatial auditory processing, it also provides no spatial information to a listener. Such displays may be useful if the location of an auditory object is not known or if spatial auditory information is unimportant. However, diotic displays are the least realistic headphone display. In addition, as discussed in Section 4.3.6.9, benefits of spatial hearing can be extremely useful for detection and recognition of auditory information. For instance, when listeners are required to monitor multiple sound sources, spatialized auditory displays are clearly superior to diotic displays (Haas et al., 1997).

4.4.2.2 Dichotic Displays

While normal interaural cues vary with frequency in complex ways, simple frequency-independent ITDs and ILDs affect the perceived lateral position of a sound source (e.g., see Durlach & Colburn, 1978). Stereo signals that only contain a frequency-independent ITD and/or ILD are herein referred to as *dichotic* signals (although the term is sometimes used to refer to any stereo signal in which left and right ears are different).

Generation of a constant ITD or ILD is very simple over headphones since it only requires that the source signal be delayed or scaled (respectively) at one ear. Just as with diotic signals, dichotic signals result in sources that appear to be located on an imaginary line inside the head, connecting the two ears. Varying the ITD or ILD causes the lateral position of the perceived source to move toward the ear receiving the louder and/or earlier-arriving signal. For this reason, such sources are usually referred to as *lateralized* rather than *localized*.

Dichotic headphone displays are simple to implement but are only useful for indicating whether a sound source is located to the left or right of a listener. On the other hand, when multiple sources are lateralized at different locations (using different ITD and/or ILD values), some binaural unmasking can be obtained (see Section 4.3.6.9).

4.4.2.3 Spatialized Audio

Using signal-processing techniques, it is possible to generate stereo signals that contain most of the normal spatial cues available in the real world. In fact, if properly rendered, spatialized audio can be practically indistinguishable from free-field presentation (Langendijk & Bronkhorst, 2000). When coupled with a head-tracking device, spatialized audio provides a true virtual auditory interface. Using a spatialized auditory display, a variety of sound sources can be presented simultaneously at different directions and distances. One of the early criticisms of spatialized audio was that it was expensive to implement; however, as hardware and software solutions have proliferated, it has become feasible to include spatialized audio in most systems. Spatialized audio solutions can be fit into any budget, depending on the desired resolution and number of sound sources required. Most virtual reality (VR) systems are currently outfitted with headphones of sufficient quality to reproduce spatialized audio, making it relatively easy to incorporate spatialized audio in an immersive VE system.

4.4.2.4 Practical Limitations on Spatialized Audio

While in theory, HRTF simulation should yield stimuli that are perceptually indistinguishable from natural experience, a number of practical considerations limit the realism of stimuli simulated using HRTFs. Measurement of HRTFs is a difficult, time-consuming process. In addition, storage requirements for HRTFs can be prohibitive. As a result, HRTFs are typically measured only at a single distance, relatively far from the listener, and at a relatively sparse spatial sampling. Changes in source distance are simulated simply by scaling the overall signal intensity. Because the HRTFs are only measured for a finite number of source directions at this single source distance, HRTFs are interpolated to simulate locations for which HRTFs are not measured. While this approach is probably adequate for sources relatively far from the listener and when some inaccuracy can be tolerated, the resulting simulation cannot perfectly recreate spatial cues for all possible source locations (Wenzel & Foster, 1993). Individual differences in HRTFs are very important for some aspects of sound source localization (particularly for distinguishing front/back and up/down). However, most systems employ a standard set of HRTFs that are not matched to the individual listener. Using these *nonindividualized* HRTFs reduces the accuracy and externalization of auditory images but still results in useful performance increases (Begault & Wenzel, 1993). Researchers have explored a variety of HRTF compression schemes in which individual differences are encoded in a small number of parameters that can be quickly or automatically fit to an individual (e.g., see Kistler & Wightman, 1991; Middlebrooks & Green, 1992). Nonetheless, many *typical* systems cannot simulate source position along a cone of confusion because they do not use individualized HRTFs.

The most sophisticated spatialized audio systems use trackers to measure the movement of a listener and update the HRTFs in real time to produce appropriate dynamic spatial cues. The use of head tracking dramatically increases the accuracy of azimuthal localization (Moldrzyk, Ahnert, Feistel, Lentz, & Weinzierl, 2004; Sorkin, Kistler, & Elvers, 1989). However, time lag in such systems (from measuring listener movement, choosing the new HRTF, and filtering the ongoing source signal) can be greater than 30 ms. While the binaural system is sluggish, the resulting delay can be perceptible. Real-time systems are also too complex and costly for some applications. Instead, systems may compute signals off-line and either ignore or limit the movement of the listener; however, observers may hear sources at locations inside or tethered to the head (i.e., moving with the head) with such systems.

Many simulations do not include any echoes or reverberation in the generated signals. Although reverberation has little impact (or degrades) perception of source direction, it is important for distance perception. In addition, anechoic simulations sound subjectively artificial and less realistic than do simulations with reverberation.

4.4.3 Simulation Using Speakers

The total acoustic signal reaching each ear is simply the sum of the signals reaching that ear from each source in an environment. Using this property, it is possible to vary spatial auditory cues (e.g., ITD, ILD, and spectrum) by controlling the signals played from multiple speakers arrayed around a listener. In contrast with headphone simulations, the signals at the two ears cannot be independently manipulated; that is, changing the signal from any of the speakers changes the signals reaching both ears. As a result, it is difficult to precisely control the interaural differences and spectral cues of the binaural signal reaching the listener to mimic the signals that would occur for a real-world source. However, various methods for specifying the signals played from each loudspeaker exist to simulate spatial auditory cues using loudspeakers.

To reduce the variability of audio signals reaching the ears, careful attention should be given to speaker placement and room acoustics. If speaker systems are not properly placed and installed in a room, even the best sound systems will sound inferior. Improperly placed speakers can reduce speech comprehension, destroy the sense of immersion, and dramatically reduce bass response (Holman, 2000). This is especially true when dealing with small rooms. If the system is installed properly, there will be a uniform (flat) frequency response at the listening area.

One example of a four-speaker system (two front and two surround) is described in an International Telecommunications Union (ITU) recommendation and places speakers at ±30° in front of the listener and at ±110° behind the listener (ITU-R BS. 775-1). It is further recommended that the signals emanating from the two surround channels be decorrelated to increase the sense of spaciousness. Correlated mono signals may give a sense of lateralization rather than localization. If a subwoofer is used, it is usually placed in front of the room. Placing the subwoofer too close to a corner may increase bass response but may result in a muddier sound. The subwoofer should be moved to achieve the best response in the listening area. Unfortunately, speaker placement will vary depending on the dimensions and shape of the room, as well as the number of speakers employed. If the system is mobile, the sound system will have to be readjusted for every new location, unless the simulation incorporates its own enclosure. If the simulation will be housed in different sized rooms, the audio system (amplifiers and speakers) must have enough headroom (power) to accommodate both large enclosures as well as small. When possible, acoustical tile and diffusers should be employed where appropriate to reduce reverberation and echoes.

4.4.3.1 Nonspatial Display

Many systems use free-field speakers in which each speaker presents an identical signal. Such systems are analogous to diotic headphone systems; although simple to develop, these displays (like diotic headphone displays) provide no spatial information to the listener. Such systems can be used when spatial auditory information is unimportant and when segregation of simultaneous auditory signals is not critical. For instance, if the only objects of interest are within the visual field and interference between objects is not a concern, this kind of simplistic display may be adequate.

4.4.3.2 Stereo Display

The analog of the dichotic headphone display presents signals from two speakers simultaneously in order to control the perceived laterality of a *phantom* source. For instance, simply varying the level of otherwise identical signals played from a pair of speakers can alter the perceived laterality of a phantom source. Most commercial stereo recordings are based on variations of this approach.

Imagine a listener sitting equidistant from two loudspeakers positioned symmetrically in front of the listener. When the left speaker is played alone, the listener hears a source in the direction of the left speaker (and ITD and ILD cues are consistent with a source in that leftward direction). When the right speaker is played alone, the listener hears a source in the direction of the right speaker. When identical signals at identical levels are played from both speakers, each ear receives two direct signals, one from each of the symmetrically placed speakers. To the extent that the listener's head is left–right symmetric, the total direct sound in each ear is identical, and the resulting percept will be of a single source at a location that gives rise to zero ITD and zero ILD (e.g., in the listener's median plane). Varying the relative intensity of otherwise identical signals played from the two speakers causes the gross ITD and ILD cues to vary systematically, producing a phantom source whose location between the two speakers varies systematically with the relative speaker levels (e.g., see Bauer, 1961).

This simple *panning* technique produces a robust perception of a source at different lateral locations; however, it is nearly impossible to precisely control the exact location of the phantom image. In particular, the way in which the perceived direction changes with relative speaker level depends upon the location of the listener with respect to the two loudspeakers. As the listener moves outside a restricted area (the *sweet spot*), the simulation degrades rather dramatically. In addition, reverberation can distort the interaural cues, causing biases in the resulting simulation. Nonetheless, such systems provide some information about source laterality and can be very effective when one wishes to simulate sounds from angular positions falling between the loudspeaker positions.

4.4.3.3 Multichannel Loudspeaker Systems

Two-channel stereo can be extended to multichannel panning techniques, so-called vector base amplitude panning (VBAP). The technique can be used to place virtual source in 3D space if the loudspeaker arrays are surrounding the listener (Pulkki, 1997). Loudspeaker triplets are used with particular amplitudes in order to create a sound image in the corresponding triangle. Larger areas around the listener are covered by several of those triplets.

Another technique, actually one of the most popular techniques of spatial audio, is *Ambisonics* (Gerzon, 1976). The mathematical basis is the set of spherical harmonics (SH), a set of orthogonal functions in spherical coordinates. With these, sound fields can be decomposed into their directional components (SH coefficients); these exact coefficients are used to filter the speakers of the reproduction array. The speaker arrays can be freely designed in 3D space, but simple solutions correspond to Platonic bodies such as cubes, dodecahedrons, or icosahedrons. The SH coefficients can be derived from simulation or from recordings with spherical microphone arrays. The first-order approach as defined by Gerzon requires a setup of an omnidirectional and three figure-eight microphones. The spatial resolution and the corresponding details of directional sounds are limited by the low-order SH representation, which creates a kind of spatial smoothing. Higher-order Ambisonics (HOA) allow reproduction of more spatial detail; for this, higher-order microphone arrays must be used (Meyer & Elko, 2002). Wave field synthesis (WFS) technology is another theoretical approach to wave field reconstruction (Berkhout, 1988). In WFS, sound waves are decomposed into plane waves sampled on linear or circular microphone arrays, typically in 2D. If the discrete spatial sampling of the array is sufficiently high, any wave field can be reconstructed with a corresponding large number of loudspeakers in a dense distribution. This goes back to the Huygens principle that explains wave propagation by arrays of numerous point sources, each radiating elementary waves, which interfere to form wave fronts (e.g., see De Vries, 2012). To use WFS, the process of sound recording and mixing is different from usual techniques applied in audio engineering. The spatial decoding of virtual sources is integrated in a flexible way so that position, orientation, and movements of the listener are not restricted in dynamic scenes.

All of the multichannel techniques listed suffer from the fact that a large number of loudspeakers must be used with accordingly a large amount of signal processing, control, and amplifier units. For VR installations with surrounding displays, a practical problem often arises from the conflicting demands of trying to place loudspeaker arrays along with video screens with an undisturbed video image and free line of sight. Acoustically transparent video screens would solve the problem, but the image resolution of such screens is significantly less than for high-quality hard projection screens.

4.4.3.4 Cross-Talk Cancellation and Transaural Simulations

More complex signal-processing schemes can also be used to control spatial cues using a small setup of loudspeakers. In such approaches, the total signal reaching each ear is computed as the sum of the signals reaching that ear from each of the speakers employed. By considering the timing and content of each of these signals, one can try to reproduce the exact signal desired at each ear.

The earliest such approach attempted to recreate the sound field that a listener would have received in a particular setting from stereo recordings taken from spatially separated microphones. In the playback system, two speakers were positioned at the same relative locations as the original microphones. The goal of the playback system was to play signals from the two speakers such that the total signal at each ear was equal to the recorded signal from the nearer microphone. To the extent that the signal from the far speaker was acoustically canceled, the reproduction would be accurate. Relatively simple schemes involving approximations of the acoustic alterations of the signals as they impinged on the head were used to try to accomplish this *cross-talk cancellation*.

As signal-processing approaches have been refined and knowledge of the acoustic properties of HRTFs improves, more sophisticated algorithms have been developed. In particular, it is possible to calculate the contribution of each speaker to the total signal at each ear by considering the HRTF corresponding to the location of the speaker. The total signal at each ear is then the sum of the HRTF-filtered signals coming from each speaker. If one also knows the location and source of the signal that is to be simulated, one can write equations for the desired signals reaching each ear as HRTF-filtered versions of the desired source. Combining these equations yields two frequency-dependent, complex-valued equations that relate the signals played from each speaker to the desired signals at the ears. To the extent that one can find and implement solutions to these equations, it is possible (at least in theory) to recreate the desired binaural signal by appropriate choice of the signals played from each speaker. The problem with such approaches is that the simulation depends critically on the relative location of the speakers and the listener. In particular, if the listener moves his head outside of the sweet spot, the simulation degrades rapidly. Head tracking and dynamic cross-talk filtering is one solution to this problem, which also allows head movements explicitly so that the listener can benefit from a more natural behavior in the virtual scene (Lentz & Behler, 2004). Head trackers can be used in conjunction with multispeaker simulations in order to improve the simulation. However, this requires that computations be performed in real time and significantly increases the cost and complexity of the resulting system. It can be difficult to compute the required loudspeaker signals and the computations are not particularly stable numerically (Masiero & Vorländer, 2012). The technique called *stereo dipole* or *sound bar* is actually more stable. It is an optimized distributed source approach where the high frequencies are radiated from the frontal region while the loudspeakers for the low frequencies span a larger angle (Takeushi, Teschl, & Nelson, 2001). To the extent that reverberation in the listening space further distorts the signals reaching the ears, the derived solutions are less robust than those for headphones. In all cases of binaural reproduction using headphones or cross-talk cancellation, the quality is best if the HRTFs used in the equations are matched to the listener.

4.4.3.5 Lessons from the Entertainment Industry

The ability to generate an accurate spatial simulation using loudspeakers increases dramatically as the number of speakers used in the display increases. With an infinite number of speakers around the listener, one would simply play the desired signal from the speaker at the desired location of the source to achieve a *perfect* reproduction. Surround sound technologies, which are prevalent in the entertainment industry, are implemented via a variety of formats. Surround sound systems find their genesis in a three-channel system created for the movie *Fantasia* in 1939.