Sight  Sound  Motion
FROM THE CENGAGE SERIES IN COMMUNICATION ARTS

General Mass Communication
Belmas/Overbeck, Major Principles of Media Law, 2016 Edition
Biagi, Media/Impact: An Introduction to Mass Media, Twelfth Edition
Fellow, American Media History, Third Edition
Lester, Visual Communication: Images with Messages, Sixth Edition
Straubhaar/LaRose/Davenport, Media Now: Understanding Media, Culture, and Technology, Ninth Edition
Zelezny, Cases in Communications Law, Sixth Edition
Zelezny, Communications Law: Liberties, Restraints, and the Modern Media, Sixth Edition

Journalism
Bowles/Borden, Creative Editing, Sixth Edition
Davis/Davis, Cengage Advantage Books: Think Like an Editor: 50 Strategies for the Print and Digital World, Second Edition
Hilliard, Writing for Television, Radio, and New Media, Eleventh Edition

Public Relations and Advertising
Drewnianny/Jewler, Creative Strategy in Advertising, Eleventh Edition
Hendrix/Hayes, Public Relations Cases, Ninth Edition
Newsom/Turk/Kruckeberg, Cengage Advantage Books: This is PR: The Realities of Public Relations, Eleventh Edition

Radio, Television, and Film
Albarran, Management of Electronic and Digital Media, Sixth Edition
Alten, Audio Basics, First Edition
Alten, Audio in Media, Tenth Edition
Gross/Ward, Digital Moviemaking, Seventh Edition
Hausman/Messere/Benoit, Modern Radio and Audio Production: Programming and Performance, Tenth Edition
Hilliard, Writing for Television, Radio, and New Media, Eleventh Edition
Hilmes, Only Connect: A Cultural History of Broadcasting in the United States, Fourth Edition
Mamer, Film Production Technique: Creating the Accomplished Image, Sixth Edition
Lewis, Essential Cinema: An Introduction to Film Analysis, First Edition
Zettl, Sight Sound Motion: Applied Media Aesthetics, Eighth Edition
Zettl, Television Production Workbook, Twelfth Edition
Zettl, Video Basics, Seventh Edition
Zettl, Video Basics Workbook, Seventh Edition

Research and Theory
Sparks, Media Effects Research: A Basic Overview, Fifth Edition
Wimmer/Dominick, Mass Media Research: An Introduction, Tenth Edition
To all who see virtue
in optimizing the aesthetics
of visual communication
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Applied Media Aesthetics</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>The First Aesthetic Field: Light</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>Structuring the First Aesthetic Field: Lighting</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>The Extended First Field: Color</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>Structuring Color: Function and Composition</td>
<td>72</td>
</tr>
<tr>
<td>6</td>
<td>The Two-Dimensional Field: Area</td>
<td>86</td>
</tr>
<tr>
<td>7</td>
<td>The Two-Dimensional Field: Forces within the Screen</td>
<td>108</td>
</tr>
<tr>
<td>8</td>
<td>Structuring the Two-Dimensional Field: Interplay of Screen Forces</td>
<td>134</td>
</tr>
<tr>
<td>9</td>
<td>The Three-Dimensional Field: Depth and Volume</td>
<td>160</td>
</tr>
<tr>
<td>10</td>
<td>Structuring the Three-Dimensional Field: Screen Volume and Effects</td>
<td>188</td>
</tr>
<tr>
<td>11</td>
<td>Building Screen Space: Visualization</td>
<td>220</td>
</tr>
<tr>
<td>12</td>
<td>The Four-Dimensional Field: Time</td>
<td>246</td>
</tr>
<tr>
<td>13</td>
<td>The Four-Dimensional Field: Motion</td>
<td>270</td>
</tr>
<tr>
<td>14</td>
<td>Structuring the Four-Dimensional Field: Timing and Principal Motions</td>
<td>286</td>
</tr>
<tr>
<td>15</td>
<td>The Five-Dimensional Field: Sound</td>
<td>310</td>
</tr>
<tr>
<td>16</td>
<td>Structuring the Five-Dimensional Field: Sound Structures and Sound/Picture Combinations</td>
<td>334</td>
</tr>
<tr>
<td>17</td>
<td>Visual Narrative: The Syntax of Continuity Editing</td>
<td>368</td>
</tr>
<tr>
<td>18</td>
<td>Visual Narrative: The Syntax of Complexity Editing</td>
<td>394</td>
</tr>
</tbody>
</table>
Contents

About the Author  xxi

Preface  xxiii

Prologue  1

1  Applied Media Aesthetics  2

Definition: Media  4

Definition: Applied Media Aesthetics  4

Applied Aesthetics and Art  5

Art and Experience  5

Applied Aesthetics and Contextual Perception  5

Stabilizing the Environment  6

Selective Seeing and Selective Perception  6

The Power of Context  8

Associative Context  8

Aesthetic Context  9

The Medium as Structural Agent  11

Applied Media Aesthetics: Method  12
Contents

Fundamental Image Elements  12
  Analysis and Synthesis  13
  Content  13
Responsibility  14
  SUMMARY  15
  NOTES  16

2

The First Aesthetic Field: Light  18
  The Nature of Light  19
  Lighting Purposes and Functions  20
  The Nature of Shadows  21
    Attached and Cast Shadows  22
    Falloff  24
  Outer Orientation Functions: How We See an Event  26
    Spatial Orientation  26
    Tactile Orientation  26
    Time Orientation  28
  Inner Orientation Functions: How We Feel about an Event  30
    Establishing Mood and Atmosphere  30
    Above- and Below-Eye-Level Key-Light Position  31
    Predictive Lighting  32
    Light and Lighting Instruments as Dramatic Agents  33
    SUMMARY  34
    NOTES  35

3

Structuring the First Aesthetic Field: Lighting  36
  Standard Lighting Techniques  37
  Chiaroscuro Lighting  38
    Analysis of Chiaroscuro Lighting  38
    Functions of Chiaroscuro Lighting  40
Specific Chiaroscuro Lighting Types 42
  Rembrandt Lighting 42
  Cameo Lighting 42

Flat Lighting 44
  Functions of Flat Lighting 44

Silhouette Lighting 46

Media-Enhanced and Media-Generated Lighting 47

Single- and Multicamera Lighting 48
  Single-Camera Lighting 48
  Multicamera Lighting 49

The Aesthetic Edge: Unusual Lighting 49

  SUMMARY 50

  NOTES 51

4

The Extended First Field: Color 52

  What Is Color? 53

  How We Perceive Color 54
    Basic Physiological Factors 54
    Basic Aesthetic Perception Factors 55
    Grayscale 57
    Color Models 57
    Compatible Color 59

  How We Mix Color 59
    Additive Color Mixing 59
    Subtractive Color Mixing 62
    Mixed Mixing 63

  Relativity of Color 63
    Light Environment 64
    Surface Reflectance 64
    Color Temperature 65
    Surrounding Colors 65
    Color Juxtaposition 66
    Brightness and Color Constancy 66
Visual Narrative:
The Syntax of Complexity Editing

Analytical Montage 396
  Sequential Analytical Montage 396
  Sectional Analytical Montage 398

Idea-Associative Montage 400
  Comparison Idea-Associative Montage 401
  Collision Idea-Associative Montage 402

Metric Montage 405

Audio/Video Montage 406
  Sequential Analytical A/V Montage 406
  Sectional Analytical A/V Montage 407
  Comparison Idea-Associative A/V Montage 407
  Collision Idea-Associative A/V Montage 407

Media Aesthetics and the Human Condition 408
  SUMMARY 409
  NOTES 410

Glossary 412

Index 423
Herbert Zettl, PhD, taught for 40 years in the Broadcast and Electronic Communication Arts Department at San Francisco State University. His research emphases were, and still are, media aesthetics and video production. While at San Francisco State, he headed the Institute of International Media Communication (IIMC). The IIMC facilitated international visitors through the auspices of the US State Department's International Information Programs and the San Francisco International Diplomacy Council, sponsoring international visiting scholars. Dr. Zettl is one of the founders of the Annual Visual Communication Conference, a national conference for visual communication scholars. He received the California State Legislature Distinguished Teaching Award in 1966, and in 2004 he received the Distinguished Education Service Award of the Broadcast Education Association.

Prior to joining the San Francisco State University faculty, Dr. Zettl worked at several professional television stations, including KPIX, the CBS-owned and operated station in San Francisco, where he was a producer-director. He participated in numerous CBS and NBC network television productions, such as Edward R. Murrow's Person to Person and several network specials. He is a member of the prestigious Gold Circle of the National Academy of Television Arts and Sciences (NATAS), Northern California Chapter, for his outstanding contributions to the television profession. He is also a member of NATAS Broadcast Legends.

Dr. Zettl has been a visiting professor at Concordia University in Montreal, Canada; Heidelberg University in Heidelberg, Germany; the Institute for Television and Film (now the Hochschule für Fernsehen und Film) in Munich, Germany; and the University of South Africa, Pretoria. For one year he served as resident director in Germany for California State University students at Heidelberg and Tübingen Universities. For several years he consulted as an academic specialist with broadcast institutions in various countries, frequently under the auspices of the State Department's International Information Programs. He also acted as consultant to a number of universities and professional broadcast institutions in North and South America, Europe, the Middle East, Africa, Asia, and Southeast Asia.

In his seminar on experimental production, Dr. Zettl spearheaded various experimental television productions, such as dramas for simultaneous multiscreen presentation, binaural audio for aural depth perception, and inductive visual presentation techniques.
He has presented many papers on media aesthetics and video production for a variety of academic and professional media conventions both in the United States and abroad. He has also published numerous articles, many of which have been translated into foreign languages and published abroad. His other books on television production and aesthetics, all published by Cengage Learning, include: *Television Production Handbook*, 12th ed. (2015); *Television Production Workbook*, 12th ed. (2015); *Video Basics 7* (2013); and *Video Basics 7 Workbook* (2013). *Television Production Handbook, Sight Sound Motion*, and *Video Basics* have been translated into several foreign languages (including Spanish, Portuguese, Greek, Chinese, and Korean) and are used in key television production centers and universities around the world.

His interactive DVD-ROM, *Zettl’s VideoLab 4.0*, published by Cengage Learning in 2011, contains basic information about video production and interactive simulated production exercises in camera, lighting, audio, switching, and editing modules. His previous versions have netted several awards.
Preface

For the Student

The new video technology, which lets you produce high-quality standard two-dimensional (2D) and even three-dimensional (3D) images and sound with relatively inexpensive equipment, puts more pressure on you to match this technical quality with equally high aesthetic standards. This means that, when framing a shot or adding music to your video track, you must have the knowledge and skill to select those aesthetic elements that not only result in optimally effective communication but also reflect your artistic sensitivity and capabilities.

Today the many screen sizes and aspect ratios of large digital movie screens, high-definition home video screens, various computer monitors, and small mobile media displays require not only new framing and compositional principles but new sound considerations as well. When dealing with stereo 3D and virtual reality (VR), you are, in effect, confronted with new media. In this context applied media aesthetics has gained new prominence and urgency. If you find that this book is not exactly bedtime reading and that it sometimes seems too theoretical, you are encouraged to find examples of television shows, motion pictures, or Internet footage that illustrate the discussion and make the concepts more concrete. Your learning will be greatly enhanced by realizing that every aesthetic concept discussed in this text has a practical application.

You will also discover that these concepts will help you go way beyond an “I like it” or “I don’t like it” evaluation and provide you with solid criteria for critical analyses of video programs and films. In this endeavor you must learn to recognize and evaluate not only the established production standards but also the new ways in which aesthetic elements are used. For example, the jump cut, extreme high-contrast lighting, and color and sound distortion can be either purposeful aesthetic effects to intensify the message or gross production mistakes. By knowing the aesthetic tools you have on hand and how to use them in the context of the screen event, you will have little trouble deciding whether the effects were done intentionally or out of ignorance. If you have already been applying most of these principles in your work, you now have proof that you were on the right track all along.
Sight Sound Motion describes the four fundamental and contextual image elements—light and color, space, time/motion, and sound—and how they are used in electronic media. These fundamental image elements are discussed in the context of the five principal aesthetic fields: light and color, two-dimensional space, three-dimensional space, time/motion, and sound. This organization allows the examination of each individual aesthetic element while maintaining an overview of how these elements interact contextually.

FEATURES

Although the basic aesthetic principles of the five aesthetic fields do not necessarily change with the advancement of technology, their applications do. I have therefore expanded certain areas that have become especially important in such fields as digital video, 2D and 3D digital cinema, and virtual reality.

Color All illustrations are now in color except those that emphasize the aesthetics and power of black-and-white images. There are still instances in which color interferes with, rather than facilitates, empathy and the expression of intense emotions. This topic is explored in the context of the desaturation theory.

Screen size In today’s digital video world, you will have to be an expert miniature painter as well as one for huge outdoor advertising panels. This book contains valuable information about how to compose effective shots for the tiny mobile media display as well as the large screens in motion picture theaters.

Aspect ratio Most illustrations in this text are shown in the familiar 16 × 9 aspect ratio because it represents the wide-screen aspect ratios of high-definition television (HDTV) and, with some modification, standard motion picture screens. The 4 × 3 aspect ratio is still discussed not only because it is the ratio of all traditional movies and standard television but also because it approximates the ratio of most still camera viewfinders. The variable aspect ratios of smartphones are also discussed in this context.

Inductive sequencing The inductive approach to the visual narrative, by showing a series of close-ups rather than moving from a master shot to close-ups, has become a fundamental prerequisite for telling a story effectively on the small screen. In combination with surround sound, the inductive visual approach has also become an effective technique for creating high-energy sequences on the large movie screen.

Stereoscopic 3D and VR Although some proponents of 3D claim that there is no difference in the acquisition of 2D and 3D images, the aesthetics of the two production modes differ to a great extent. The principal features of 3D are explained throughout the book and examined more extensively in the chapters on the three-dimensional field. In this context the major aesthetic differences between 3D and virtual reality are highlighted.

Visual narrative The chapters on the five-dimensional field of sound precede the last two chapters on editing because sound structures figure prominently in the narratives and syntax of continuity and complexity editing.
PEDAGOGY

As in the previous edition, this eighth edition of *Sight Sound Motion* incorporates several pedagogical devices for optimal learning, such as chapter summaries, source references, relevant photos and diagrams, and an extensive glossary.

**Chapter summaries**  The chapter summaries recapitulate the main ideas in each chapter for quick review, but students should not take them as a shortcut to reading the text.

**References**  Although the numerous notes at the end of each chapter are not essential for understanding the text, they identify the significant research and can serve as a useful guide for further study of media aesthetics. The text includes frequent references to how a specific aesthetic principle in one field operates contextually in another or several other fields. Throughout the book the examples are chosen for how well they illustrate or support a specific media aesthetic concept rather than whether they are current or popular.

Instructors are encouraged to contribute their own material whenever it is deemed more appropriate and effective than the examples in the text. Note that each concept in *Sight Sound Motion* can be applied. Instructors are also encouraged to find as many applications as possible to help make the text optimally relevant to students.

**Glossary**  All the fundamental media aesthetic terms are defined in the glossary and appear in *bold italic* in the chapter text in the context in which they are defined. Like the vocabulary of a foreign language, knowledge of these terms is an essential prerequisite to communication about media aesthetics. Realize that some of the less familiar terms, such as *vector*, are used in this book not to test the reader’s patience but because they are more precise than the ones commonly used. Perusing the glossary before reading the text may facilitate understanding of the various concepts.

ANCILLARIES

The following ancillaries have been designed to aid in the teaching and understanding of *Sight Sound Motion*.

**MindTap**  MindTap Radio/TV/Film is now available with *Sight, Sound, Motion: Applied Media Aesthetics*, Eighth Edition. Fully integrated into a seamless experience, MindTap combines readings, multimedia, activities, and assessments into a single learning path—guiding students through the course, maximizing their study time, and helping them master course concepts. Instructors can personalize the learning path by customizing Cengage Learning resources and adding their own content via apps that integrate into the MindTap framework with any learning management system. To learn more ask your Cengage Learning representative to demo MindTap for you—or visit www.cengage.com/mindtap.

**Online Instructor’s Manual**  The fully revised and updated Online Instructor’s Manual contains suggestions for classroom demonstrations, exercises, and discussions and a battery of multiple-choice questions for each chapter. The manual is intended as a guide, not a dictum. The demonstrations can be done with a smartphone camera or a small camcorder together with a playback device—but they can also be staged somewhat more effectively in a multicamera studio setup. Ideally, the Online Instructor’s Manual should stimulate you to come up with maximally
effective ways to make the connection between media aesthetic principles and their applications. The Online Instructor’s Manual is available for download at:

login.cengage.com

If you require assistance, please contact your Cengage Learning representative.

**Zettl’s VideoLab 4.0**  This interactive DVD-ROM combines basic television production techniques with some fundamental principles of media aesthetics. It is truly interactive: the student can zoom in and out, turn on lighting instruments and move them around, mix sounds, and edit together certain shots and see the results immediately. An extensive quiz feature and instant access to the glossary reinforce learning. The DVD-ROM can be used as a convenient way to help students acquire or reinforce basic video and film production techniques and to illustrate aesthetic concepts that need to be shown in motion. The Online Instructor’s Manual refers to the relevant sections of Zettl’s VideoLab 4.0 for each chapter of *Sight Sound Motion*.

**Music examples**  To actually listen to all the major music examples in chapter 16, go to the MindTap for *Sight Sound Motion*. The examples are listed by the figure number referenced in the chapter. Each example appears in musical notation so that while you listen to it you can follow the notes or, if you don’t read music, the vector directions. These audio files are also available as part of the book’s online instructor resources.

**ACKNOWLEDGMENTS**

I am greatly indebted to the editors of Cengage Learning, all the people who work under the auspices of Ideas to Images, my colleagues and students at San Francisco State University and other learning institutions, and reviewers of the previous edition.

Specifically, I would like to thank the Cengage team of Kelli Strieby, product manager; Erin Bosco, associate content developer; and Rachel Schowalter, associate content developer, for their help and support throughout the writing and publishing process. Erin and Rachel deserve special credit for treating me as though I were their only author. What amazes me especially are their quick and helpful responses to my many questions and requests. For me this is a true index of their professionalism.

I have been working with Gary Palmatier of Ideas and Images for the past six editions, and each time I am impressed anew by the quality of his work. As art director, designer, and project manager, he has the rare gift of being highly creative yet a stickler for details. He also knows how to pick an effective team. Elizabeth von Radics, copy editor for the past six editions, has once again lent her talent to make the text easily readable and its more difficult concepts transparent—not an easy job by any means! Ed Aiona, photographer, provided several new and highly effective illustrations. I marvel at the patience and skills of proofreader Mike Mollett. They all merit my deep gratitude.

I am greatly indebted to my grandson, Christopher Child, who made it possible for me to experience the future right now. He introduced me to Arthur van Hoff, founder and chief technology officer of Jaunt, a virtual reality company in Palo Alto, California. With his staff, Arthur treated me to an impressive demonstration of Jaunt’s latest VR content and hardware developments. My thanks and good wishes for success go to all of them. I would also like to acknowledge Franz Troppenhagen, product manager of multimedia devices at Carl Zeiss, Germany,
for letting us use the photos of his impressive VR head-mounted display. Finally, my colleagues Dr. Hamid Khani and Dr. Vinay Shrivastava have been sharing with me on a regular basis their longtime teaching experience of the *Sight Sound Motion* course at San Francisco State University. Their comments have been most helpful in clarifying various sections of the book. Dr. Khani arranged for my use of the TV studios of the Broadcast and Electronic Communication Arts Department and assisted in recruiting student volunteers as models. I thank them for this courtesy.

I want to thank the reviewers of the previous edition for their valuable suggestions for improving the current one: JC Barone, Western Connecticut State University; Josh Ellis, Miami Dade College–North; Michael Ogden, Central Washington University; Dann Pierce, University of Portland; David Rembiesa, Mott Community College; and Bill Yousman, Eastern Connecticut State University. Hsianghsiong Liao, who translated the first and the two most recent editions of *Sight Sound Motion,* deserves my special thanks, as does Dr. JC Barone of Western Connecticut State University for his excellent work on this book's MindTap.


Finally, I want to again thank my wife, Erika, for her unflappable support during this writing marathon.

Herbert Zettl
This book gives you the tools to clarify, intensify, and interpret events for television, Internet, and cinema presentation. In effect, it teaches you how to apply the major media aesthetic elements to manipulate how people see screen images. Because media consumers are largely unaware of the power of media aesthetics, they must and do trust your professional judgment and especially your good intentions.

Irrespective of the scope of your communication—a brief news story, an advertisement, a documentary, or a major dramatic production—your overriding aim should be to help people attain a higher degree of emotional literacy, the ability to see the world with heightened awareness and joy. All of your aesthetic decisions must ultimately be made within an ethical context—a moral framework that holds supreme the dignity and the well-being of humankind.
Consciously or not, you make many aesthetic choices every day. When you decide what to wear, arrange stuff on your desk, or choose what flowers to put on the dinner table, or even when you judge the speed or distance of your car relative to other cars while driving, you are engaging in basic perceptual and aesthetic activities. Even the everyday expression “I know what I like” requires aesthetic judgment.

When you select a certain picture to put on your wall, choose a specific color for your car, or look through the viewfinder of a camera, you are probably more conscious of making an aesthetic decision. This kind of decision-making, as any other, requires that you know what choices are available and how to make optimal decisions with a minimum of wasted effort. Painting your bathroom first red, then pink, then orange only to discover that off-white is in fact the best color would be not only expensive and time-consuming but also cumbersome and frustrating.

As a responsible media communicator, you must go beyond everyday reflexes and approach creative problems with educated judgment. You must also develop a heightened sense of vision to recognize the universal needs and desires of human beings and learn how to give such vision significant form so that you can share it with all of us. Applied media aesthetics helps you in this formidable task. If not communicated effectively, even significant vision subsides into an insignificant dream. Despite the enormous changes that the digital revolution has brought about in video and film production hardware, software, and production methods, the basic media aesthetic principles still stand. In fact, because of the vastly increased choices in digital audio and video manipulation, media aesthetics has become an indispensable tool for structuring content.

To provide you with some overview of applied media aesthetics, we need to start by briefly explaining what we mean by media and media aesthetics. Then we expand our discussion to these six areas: applied aesthetics and art; applied aesthetics and contextual perception; the power of context; the medium as structural agent; applied media aesthetics: method; and fundamental image elements and personal responsibility.
DEFINITION: MEDIA

Throughout this book you will find media references to video, television, computer displays, film, and filming. To avoid any confusion about what is meant by these terms, we briefly define them here.

**Video**  
Video is the more inclusive term for all kinds of video productions, including television. Television is sometimes singled out, however, especially when discussed as broadcast television in connection with a specific transmission mode (such as live versus recorded), reception situation, or program genre. Because computer images are displayed on a video screen, we incorporate them in the video category.

**Film**  
The term film refers to traditional motion pictures that are screened and watched in movie theaters. Although the photochemical film medium is a thing of the past and has been replaced by digital electronics, film is still a more established and convenient term than digital cinema.

The term filming is often used to refer to a great variety of digital video capture, regardless of whether it is for a brief television news story or a blockbuster movie. “Making a film” can imply the use of a detailed script and careful preproduction, a shot-by-shot capture with mostly a single camera, and an extended postproduction period, but it can also refer to the video-recording of someone’s weekend trip to the beach.

Despite the technical convergence of video and film, they still differ substantially in how they are watched. Films are watched primarily in a darkened movie theater on a very large screen with high-quality surround-sound audio. Video productions, even the rather ambitious ones, are normally watched at home on a television or computer screen or even outdoors on a tiny smartphone display. The remaining and crucial difference between the two media is how the various media aesthetic principles are applied.

DEFINITION: APPLIED MEDIA AESTHETICS

Applied media aesthetics differs from the traditional concept of aesthetics in three major ways. First, we no longer limit aesthetics to the traditional philosophical concept that deals primarily with the understanding and appreciation of beauty and our ability to judge it with some consistency. Nor do we consider aesthetics only to mean the theory of art and art’s quest for truth. Applied media aesthetics considers art and life as mutually dependent and essentially interconnected.

The major functions of media aesthetics are based on the original meaning of the Greek verb *aisthanomai* ("I perceive") and the noun *aisthetike* ("sense perception"). Applied media aesthetics is not an abstract concept but a process in which we examine a number of media elements, such as lighting and sound, how they interact, and our perceptual reactions to them. Second, the media—in our case primarily video, film, and to a lesser extent web images—are no longer considered neutral means of simple message distribution but essential elements in the aesthetic communication system. Third, whereas traditional aesthetics is basically restricted to the analysis of existing works of art, applied media aesthetics serves not only the analyses of the various forms of videos and movies but also their synthesis—their creation.

In contrast to traditional aesthetic theories, almost all media aesthetic principles and concepts discussed in this book can be applied to a variety of media production tasks. A thorough understanding of media aesthetic principles will also help you adjust relatively easily to the new and always-changing production requirements of various digital media. Finally, the criteria of applied media
applied aesthetics let you employ formative evaluation, which means that you can evaluate the relative communication effectiveness of the aesthetic production factors step-by-step while your production is still in progress.

**APPLIED AESTHETICS AND ART**

*Applied aesthetics* emphasizes that art is not an isolated object hidden away in a museum and that aesthetic experiences are very much a part of everyday life. Whatever medium you choose for your expression and communication, art is a process that draws on life for its creation and, in turn, seems necessary, if not essential, for living life with quality and dignity. We need art to educate our emotions. Even if you are not in the process of creating great works of art, you are nevertheless constantly engaged in myriad aesthetic activities that require perceptual sensitivity and judgment. But if ordinary life experiences are included in the process of art, how are you to distinguish between aesthetic processes that we call “art” and those that are not art? Is every aspect of life, every perceptual experience we have, art? No. Ordinary daily experiences may be full of wonder, but they are not art—not yet, in any case. But they do have the potential of serving as raw material for the process of aesthetic communication that we call art.

**Art and Experience**

What then is the deciding element that elevates an ordinary life experience to the realm of art? The critical factor is you—the artist—or a group of artists, such as the members of a video or film production team, who perceive, order, clarify, intensify, and interpret a certain aspect of the human condition for themselves or, in the case of media communication, for a specific audience.

Philosopher Irwin Edman pioneered a new aesthetic concept almost a century ago that stresses the close connection between art and life. He wrote: “So far from having to do merely with statues, pictures, symphonies, art is the name for that whole process of intelligence by which life, understanding its own conditions, turns these into the most interesting or exquisite account.” This process presupposes that life is given “line and composition” and that the experience is clarified, intensified, and interpreted. “To effect such an intensification and clarification of experience,” Edman says, “is the province of art.” From this perspective, events that some may consider ugly or utilitarian have as much chance of becoming an aesthetic experience as a beautiful sunset. SEE 1.1

This process of clarification, intensification, and interpretation is also the province of applied media aesthetics. Whenever you look through the viewfinder of a camera to compose a shot, arrange some visual elements on a computer screen, or edit a film or video sequence, you are engaged in the creative act of clarifying, intensifying, and interpreting some event for a particular audience.

**APPLIED AESTHETICS AND CONTEXTUAL PERCEPTION**

We perceive our world not in terms of absolutes but rather as changing contextual relationships. When we look at an event, we are constantly engaged in judging one aspect of it against another aspect or another event. A car is going fast because another one is going slowly or because it moves past a relatively stationary object. An object is big because another one is smaller. The beam from the same flashlight looks pitifully dim in the midday sun but bright and powerful in a dark room.

When you drive a car, your perceptual activities work overtime. You are constantly evaluating the position of your car relative to the surroundings as well as the changes in the surroundings relative to your car. No wonder you feel tired after even a short drive through the city during rush hour. When you sit perfectly
1.1 Art and Life

Within the contextualistic framework, we can draw aesthetic experience from all aspects of life. By giving “line and composition” to even a relatively ordinary scene, like the renovation of a college dormitory, an artist can help us perceive its inherent beauty. Still and stare at a stationary object, such as a table, your eyes nevertheless move constantly to scan the object. You then fuse the many, slightly different views together into a single image of the table, much as a well-edited sequence of various camera angles becomes a cohesive unit.

How then can we ever make sense of our multiple views of a changing world with its onslaught of sensations? Our mental operating system encourages a considerable perceptual laziness that shields us from input overload. We all develop habitual ways of seeing and hearing that make us focus on and notice only a small portion of what is actually there. We screen out most of the sensations that reach our eyes and ears, and we stabilize and simplify as much as possible what we do perceive.5

Stabilizing the Environment

Our perceptual mechanisms are designed to simplify and stabilize our surroundings as much as possible so that they become manageable. We tend to cluster certain event details into patterns and simple configurations, perceive the size of an object as constant regardless of how far away we are from it, and see the same color regardless of the actual color variations when part of the object is in the shade. Another of our automatic, “hardwired” perceptual stabilizers is the figure/ground principle, whereby we order our surroundings into foreground figures that lie in front of, or move against, a more stable background.6

Selective Seeing and Selective Perception

Most of us tend to notice especially those events, or event details, that we want to see or are used to seeing. In our habitual ways of seeing, we generally select information that agrees with how we want to see the world. This type of selective seeing—frequently but not too accurately called selective perception—is like selective exposure to information. Once we have made up our minds about something, we seem to expose ourselves mostly to messages that are in agreement with our existing views and attitudes, ignoring those messages that would upset our deeply held beliefs.7 We also choose to look at things we like to see and are especially interested in, and we ignore those that mean little to us. See 1.2

Although such cue reductions can clarify and intensify an event for us, they can also create problems. For example, we often see and hear only those details of an experience that fit our prejudicial image of what the event should be and ignore the ones that interfere with that image. The don’t-confuse-me-with-evidence joke

**Note:** The text is a summary of the content that might be found in the image.
1.2 Selective Seeing

We tend to see events or event details that fit our perceptual expectations or that interest us highly. Each of us sees an event from his or her own point of view and according to a specific experiential context.
aptly mirrors this attitude. We then justify our questionable selection process by pointing out that the event details selected were, indeed, part of the actual occurrence. For example, if you have come to believe (perhaps through advertising or a recommendation) that the Shoreline Café has a nice atmosphere and serves excellent food, a friendly waiter may be enough evidence to verify your positive image, even if the restaurant’s food is actually quite awful. By looking only at what we want to see rather than at all there is to see, we inevitably gain a somewhat distorted view of the world.

Selective perception, on the other hand, is much more automatic; in most cases, we have no control over it. For example, if you are talking to a friend in a streetcar, you are probably not aware of most of the other sounds surrounding you, unless they start interfering with your conversation or are especially penetrating, such as a police siren or a car crash. When you see somebody wearing a white shirt, you will perceive the same white regardless of whether the person is standing in bright sunlight or in the shade. Your book pages will not look bluish when you read under a fluorescent light instead of the normal incandescent indoor lighting. Although a video camera would make such distinctions quite readily, you would have trouble seeing them, especially if you were not looking for them. Your selective perception shields you from seeing too many varieties of shades and colors so that you can keep your environment relatively stable.

THE POWER OF CONTEXT

Many of our perceptions are guided if not dictated by the event context. When context is imposed by the event itself, such as a snowstorm in May, you have little control over it. Such a context is sometimes called a “bottom-up” context. If you now decide to pitch your tent rather than abandon your backpacking trip, you are establishing a new event context—the setting up of the tent. This new context is called “top-down” because it is based on the intentionality of your actions, the pitching of the tent.

Sometimes we interpret an event by a virtual context that we form through our experience and knowledge of how the world works or ought to work, and even through our prejudices. At other times we react to contextual cues more viscerally, on a gut level, without much thought about it. Because we engage our cognitive faculties in the first situation, we call this the associative context. The second context is based more on an immediate, nonrational emotional reaction and is therefore called the aesthetic context.

**Associative Context**

One of the more important top-down contexts is the associative context. It consciously establishes and applies a code that dictates, at least to some extent, how you should feel about and interpret what you see. Here is a simple example of an associative context. Assume that you are to write down quickly the names of major US television networks:

ABC, NBC, CBS, CNN, Fox

Now we change the context to helping a child learn to write numbers from 11 to 15.

11, 12, 13, 14, 15

Take another look at the network names and the numbers. You may have noticed that the B in CBS and the 13 in the number series are very similar. In fact, they are identical. Obviously, the associative context has had a powerful influence
1.3 Associative Context

In the context of the horizontal row, the symbol at the center of this intersection is read as the letter B. In the context of the vertical row, the identical symbol is read as the number 13.

1.4 Eggs for Sale

If convenient, would you respond to this sign and buy some eggs? Justify your action.

1.5 Cheap Flying Lessons

Would you respond to this advertisement and take some flying lessons from the Affordable Flights Company? Justify your decision.

on the radically different perceptions of an identical sensation. The power of the context is so strong that you will probably find it difficult to see a 13 in the network context and a B in the numbers. Going against the established context is almost as hard as nodding your head affirmatively while uttering “no” or shaking your head sideways while saying “yes.” SEE 1.3

Another example of associative context shows how we may react to the immediate world we have constructed around us and how this world is definitely culture-bound. SEE 1.4 AND 1.5

What is your initial reaction to the two advertisements? Whereas you might respond positively to the eggs-for-sale sign and even buy some eggs if convenient, you would probably not be eager to sign up for your first flying lesson with the Affordable Flights Company. Why? Because our experience tells us that awkward hand lettering may be appropriate in the context of a small, family-run, charmingly inefficient operation that occasionally sells surplus eggs; but in the context of aviation, the sloppy hand-lettered sign is not a good indicator of reliability, efficiency, and safety. You are now comparing, however unintentionally or even subconsciously, what you see with your previous experiences and prejudices.

Aesthetic Context

When confronted with an aesthetic context, our perceptual processes are so immediate and forceful that we respond to certain stimuli in predictable ways even when we know that we are being perceptually manipulated. The many well-known optical illusions are good examples. SEE 1.6
1.6 Optical Illusion

Although we may know that the center circles in this Ebbinghaus figure are identical, we still perceive the center circle in (a) as smaller than the one in (b). The large surrounding circles in (a) make the center circle look relatively small, and the small surrounding circles in (b) make the center circle appear relatively large.

Even if you try vigorously to resist the idea of aesthetic manipulation, you cannot help but perceive the center circle in figure 1.6a as smaller than the one in figure 1.6b although in reality they are exactly the same size. The contextual circles make you perceive the central circles as being different sizes whether you like it or not. When surrounded by small circles, the central circle appears larger than it does when surrounded by larger circles.

Sufficient consistency exists in human perceptual processes so that we can predict with reasonable accuracy how people will respond to specific aesthetic stimuli and contextual patterns regardless of where they grew up. To test this, the next time you invite a friend to visit, move some of your pictures a little so that they hang slightly crooked, then watch your friend. Most likely, he or she will adjust the pictures so that they hang straight again. Your friend’s action is a predictable response to a powerful aesthetic stimulus: the disturbance of strong horizontals and verticals, of standing upright on level ground. You apply the same principle when you cant the camera to make a scene look more dynamic. **SEE 1.7**

As you know, certain lighting, colors, and especially types of music can have an immediate emotional effect on you. They all sidestep our rational faculties and therefore play a big role in establishing an aesthetic context.

1.7 Tilted Horizon

We automatically perceive a tilted horizon line as a relatively unstable event. This car seems to travel precariously fast around the turn.
But if we seek only information that reinforces our personal projection of reality and are so readily manipulated by context, how can we ever attain a relatively unbiased view of the world? The fine arts have tried for centuries to break this vicious circle. Although we may still be tied to our automatic perceptual processes and stabilizing cue reductions, all art leads, at least to some extent, to counter this automatization, to see events from various points of view and shift from glance to insight. While we may perceive a shirt as uniformly white, in a painting the artist may not only see but exaggerate the various colors reflected off the white shirt—all this so that we too can share the beauty of this world.

Significant video productions and films, regardless of genre, can and should do the same. Depending on where you put a camera or microphone, and what field of view or camera angle you select, your viewers have no choice but to share your point of view. You can prod them to see an event from different perspectives and advance them from “looking at” to “looking into.” In essence, you can help viewers educate their way of seeing, if not their perceptions. SEE 1.8

Before you can expect to help viewers become more sensitive to their surroundings and unlearn, at least to some degree, their habitual ways of seeing, you will have to acquire a degree of aesthetic literacy that allows you to perceive the complexities, subtleties, and paradoxes of life and to clarify, intensify, and interpret them effectively for an audience.12

THE MEDIUM AS STRUCTURAL AGENT

Even when your primary function in talking to someone is to communicate certain information, your behavior exerts considerable influence on how a specific message is received. It certainly makes a difference to the message recipient whether you smile or frown when extending the familiar how-do-you-do greeting. The smile will show that you are, indeed, glad to see the other person or that your message is a pleasant one; a scowl would signal the opposite. You, as the communication medium, have now become part of the structuring of the message. Well-known communication scholar Marshall McLuhan proclaimed a half century ago that “the medium is the message.” With this insightful overstatement, he meant that the medium, such as television or film, occupies an important position not only in distributing the message but also in shaping it.

Despite overwhelming evidence of how important the medium is in shaping the message, many prominent communication researchers remain more interested in analyzing the content of the literal message than in the combined effect of the message and the medium as a structural agent. In their effort to keep anything from contaminating their examination of mass-communicated content, they consider the various media as merely neutral channels through which the all-important messages are squeezed. Their analysis would reveal only your how-do-you-do greeting but ignore your smile or scowl. Gerhard Maletzke was one of the first significant mass communication scholars in Europe to advocate that it may not be only cultural or aesthetic preference that influences the shaping of the message but especially the Zwang des Mediums—the force of the medium. This concept was convincingly reinforced almost four decades later by Lev Manovich for new media, specifically various computer interfaces. Although this concept is obvious to the people who actually do the productions, it is, unfortunately,
Chapter 1

1.9 Early Communication Model
This model suggests that the communication process goes from idea to message and from message to recipient. It ignores the medium as a factor in the communication process.

Still neglected by many media scholars. This apparent lack of medium awareness stems from the very beginnings of systematic mass communication studies, where the influence of the medium on the message was almost totally ignored.16

If you have ever tried to make oil paints or clay do what you wanted them to do, you will readily admit that the medium is not neutral by any means. In fact, it has a decisive influence on the final outcome of your creative efforts. Even if you intend to communicate the same message, you will have to go about it in different ways depending on whether, for example, you design the message for wide-screen digital cinema, standard video, or a small mobile media display.

The encoding (production) as well as the decoding (reception) of the message are, to a considerable extent, a function of the technical and aesthetic potentials and requirements of the medium. Exactly how media (video, film, the computer screen, and especially the tiny smartphone display) shape or must shape the message for a specific viewer response is the subject of applied media aesthetics.

APPLIED MEDIA AESTHETICS: METHOD

The method of presenting applied media aesthetics is loosely based on Leonardo da Vinci’s Notebooks, in which he describes the “Ten Attributes of Sight Which All Find Expression in Painting.” Rather than deductively analyze a specific painting, da Vinci describes inductively the perceptual attributes that all paintings have to deal with: darkness and brightness, substance and color, form and place, and so forth.17 More specifically, applied media aesthetics is modeled after the theories and the practices of Russian painter and teacher Wassily Kandinsky. For Kandinsky abstraction did not mean reducing a realistic scene down to its essential formal elements.18 Rather, it meant an inductive process of building a scene by combining the “graphic elements”—the fundamental building blocks of painting, such as points, lines, planes, color, texture, and so forth—in a certain way.

Following this approach, he was not limited by what was there in the world around him; instead he could extend his vision to what he felt ought to be there—the construction of a new world.

As you can see, the final outcome of the deductive and inductive abstraction processes is the same, but the deductive world was reduced to its basic aesthetic elements, the inductive one built by them.

FUNDAMENTAL IMAGE ELEMENTS

In a similar inductive way, I have identified and isolated four fundamental and contextual image elements of video and film: light and color, space, time/motion, and sound.20 This book examines the aesthetic characteristics and potentials of these image elements and how we can structure and apply them within their respective aesthetic fields (light and color, two-dimensional space, three-dimensional space, time/motion, and sound). This analysis is an essential prerequisite to understanding their contextual and expressive functions.
Once you know the aesthetic characteristics and potentials of these fundamental image elements, you can study how they operate in the context of a larger aesthetic field and combine them knowledgeably into patterns that clarify, intensify, and effectively communicate a significant experience. A thorough grasp of the four image elements will help you establish an aesthetic vocabulary and language unique to the medium of your choice—a language that will enable you to speak with optimum clarity, impact, and personal style.

Analysis and Synthesis

As an analysis tool, the use of the image elements differs considerably from the traditional methods of media analysis, such as semiotics and rhetorical media criticism. Rather than analyze video and film as mostly narrative “texts” to discover how their signs function and ultimately create higher meaning, media aesthetics investigates how their fundamental image elements—light and color, space, time/motion, and sound—function within specific contexts and how they contribute to how we feel. As pointed out previously, the great advantage of applied media aesthetics over other media analysis techniques is that all of its theories can be directly applied not only to media analysis but also, if not especially, to media synthesis, or the creation of media events—the production process.

Content

You may wonder at this point what happened to the story content in all this discussion of fundamental aesthetic elements. Is not content—some form of human experience—the most fundamental of all aesthetic elements? Do we not first need an event, or some basic story idea, before we can shape it to fit the various medium and audience requirements? The answer to both of these questions is, of course, yes; but it is valuable to realize that a good idea by itself does not necessarily make for effective media communication. You must learn how to develop and mold an idea so that it fits the medium’s technical as well as aesthetic production and reception requirements. This molding process, called encoding, presupposes a thorough knowledge of such production tools as cameras, lenses, lighting, audio, and so forth as well as applied aesthetics, such as selective focus, the proper framing of a shot, the use of color, the selection of music, and the sequencing of various parts of a scene.

1.10 Deductive Abstraction

In the deductive approach to abstraction, we move from photographic realism to the essential qualities of the event.

1.11 Inductive Abstraction

In the inductive approach to abstraction, we study the formal elements of painting, or of video and film, and then arrange those elements to express the essential quality of an event. In this case, we combine lines, circles, and areas to build up (inductively) the essence of a cityscape.
This so-called formalistic approach to applied media aesthetics is similar to the study of production techniques. In both cases, we learn the tools and techniques before putting them to work in different contexts for a variety of communication purposes. Concern about significant literal content is not unimportant; it is merely premature. The study of vocabulary and the parts of speech does not preclude a respect for literature, but it is an essential prerequisite for writing the great American novel.

Once you have a strong grasp of applied media aesthetics, you can select those elements and techniques that are most appropriate and maximally effective for shaping specific ideas. More importantly, you will gain the opportunity to combine aesthetic elements in nontraditional ways so that your viewers can perceive the world with fresh eyes and ears and from a new and unique perspective. Conversely, the requirements and the potentials of applied media aesthetics could also generate new ideas—content that might otherwise have remained dormant. Your familiarity with the formal elements of applied media aesthetics and their respective fields will enable you to exercise your creativity to its fullest.

RESPONSIBILITY

As you now know, the basic purpose of applied media aesthetics is to clarify, intensify, and interpret events for a specific audience. Although such processes are designed to help the audience see the world from a new perspective and experience it in heightened ways, they also imply a direct and calculated manipulation of the audience’s perceptions. Even when producing a simple commercial, you are purposely exploiting the feelings, the emotions, and ultimately the behaviors of your viewers. Worse, although the recipients of your aesthetically clarified and intensified messages may realize that they are being manipulated, they are usually not quite sure how. For example, alert viewers will usually recognize blatantly biased editing, but they may remain largely unsuspecting when manipulated through subtle means such as color, lens distortions, lighting effects, or contextual background sounds.

An anesthetized patient on the operating table and the aesthetically illiterate video or film viewer have much in common. Both have little control over what is happening to them, and both must trust the skills, judgment, and, above all, good intentions of someone else. Thus the surgeon and the media producer bear a heavy responsibility. One penetrates human beings with a scalpel whereas the other uses highly charged, keenly calculated aesthetic energy. This is why you, as a media communicator, must make all of your decisions within the context of established ethics—within a basically moral frame of reference.

As a mass communicator who daily influences millions of unsuspecting people, or as a video artist with an audience of a few friends, acceptance of such responsibility is a major job prerequisite. Skill alone is not enough. First and foremost you must bring to the job a genuine concern and respect for your audience. And you must be prepared to bear responsibility for your actions.

As consumers of media communication, we cannot escape similar responsibilities. If we want to guard against irresponsible persuasion and take an active part in making media communication more beneficial to our fellow human beings, even as consumers we must learn as much as we can about the methods of media aesthetics.

Once we learn how lighting or sound can influence our perceptions and emotions, we are less susceptible to blind persuasion. We will be able to identify aesthetic techniques and the reasons for their use, enabling us to analyze the message for its true communication value, judge the mediated event’s relative bias,
and ultimately preserve our freedom of choice. Such media literacy will help us experience with heightened awareness and joy the mediated world on the screen as well as the real world in which we live.

When applied media aesthetics has become the common province of both the communication producer and the consumer, the imprudent use of media will become less of a problem. Both will find it easier to trust the other and to treat each other with the respect and dignity worthy of our global community.

**Summary**

Applied media aesthetics differs from traditional aesthetics in three major ways: rather than being concerned primarily with beauty and the philosophy of art, applied aesthetics deals with a number of aesthetic phenomena, including light and color, space, time/motion, and sound, and our perceptual reactions to them.

The media (video, film, and computers) themselves play an important part in shaping the message. Whereas traditional aesthetics is used primarily for analysis, media aesthetics can be applied to both analysis and synthesis—production.

In the framework of applied media aesthetics, every aspect of life has the potential to become art and to serve as raw material for aesthetic processes, so long as it is clarified, intensified, and interpreted for an audience by the artist.

Common to all perceptions is our innate urge to stabilize our environment and the practice of selective seeing and perception.

To cope with the onslaught of changing stimuli and to make our environment more manageable, our mental operating system establishes perceptual filters and has us perceive stable patterns rather than unrelated event detail.

We tend to select information that agrees with how we want to see the world and to screen out other data that might interfere with our constructs. Such habitual cue reductions tend to make us perceptually lazy and can even lead to prejudiced perceptions.

We perceive an event relative to the context in which it occurs. The bottom-up context is a given over which we have little control. As media people we have no choice but to work within it. The top-down context is intentional and in our control throughout the production process. In media aesthetics we stress the associative context, which calls up a cognitive framework in which we judge what we see by our experiences and prejudices. It is definitely culture-bound. The aesthetic context, on the other hand, is independent of a cultural frame of reference. We seem to perceive certain contextual stimuli in much the same way, irrespective of cultural upbringing or experience.

Applied media aesthetics places great importance on the influence of the medium on the message. The medium itself acts as an integral structural agent.

The method of presenting applied media aesthetics is an inductive one: rather than analyze existing video program fare and films, we isolate the four fundamental image elements of video and film, examine their aesthetic characteristics and potentials, and structure them in their respective aesthetic fields. The image elements are light and color, space, time/motion, and sound. The five principal aesthetic fields are light and color, two-dimensional space, three-dimensional space, time/motion, and sound. We thus do not take the traditional literal content (ideas to be encoded) as an essential pre- or co-requisite to the discussion of the fundamental image elements. Rather we consider the study of the image elements to be the essential prerequisite to the proper shaping of ideas and events into messages.
Because the process of clarification, intensification, and interpretation of events is based on the selection and the specific use of aesthetic elements, the recipient’s perceptions are indirectly and, more often, directly manipulated. Such aesthetic manipulation must always occur and be evaluated within a framework of basic ethics. To facilitate effective communication, the consumers as well as the producers of media communication have the responsibility to learn as much as possible about applied media aesthetics and its communicative power.


2. The word anesthetic suggests that we are bereft of all aesthetics, that our perceptions are dulled or totally shut off so that we no longer receive any stimuli, even physical ones.


4. Edman, Arts and the Man, p. 12.


7. The idea of selective exposure is broadly based on the theory of cognitive dissonance, advanced by Leon Festinger in his A Theory of Cognitive Dissonance (Evanston, IL: Row, Peterson, 1957). Basically, the theory states that we try to reduce dissonance by seeking out comments and other information that support—are consonant with—the decisions we have made.

8. Malcolm Gladwell describes in detail how a specific top-down context (dangerous neighborhood) can lead to the most tragic events (an innocent person’s being shot). See the chapter “Seven Seconds in the Bronx: The Delicate Art of Mind Reading” in his Blink (New York: Little, Brown, 2005), pp. 189–97.

9. Herbert Zettl, “Contextual Media Aesthetics as the Basis for a Media-Literacy Model,” Journal of Communication 48, no. 1 (1998): 86–89. You may also encounter the term contextualism to describe the associative and aesthetic contexts, but contextualism can also refer to a specific branch of philosophy. Basically, as a philosophical term, contextualism means that we should evaluate art within its historical epoch and according to what the artist felt while creating it. All events, or “incidents of life,” are relative and must be understood within their cultural contexts. Very much in the sense of a television docudrama, such incidents of life are interconnected and alive and spontaneous in their present, regardless of when they happened. See Stephen C. Pepper, Aesthetic Quality: A Contextualistic Theory of Beauty (New York: Charles Scribner’s Sons, 1938). Also see Stephen C. Pepper, The Basis of Criticism in the Arts (Cambridge, MA: Harvard University Press, 1943); Stephen C. Pepper, World Hypotheses (Berkeley: University of California Press, 1942, 1970); and Lewis Edwin Hahn, A Contextualistic Theory of Perception, University of California Publications in Philosophy, vol. 22 (Berkeley: University of California Press, 1939).

A more modern representative of contextualistic aesthetics is Hans-Georg Gadamer. Although he calls the basis for his aesthetic theory hermeneutical epistemology, he nevertheless represents the contextualistic point of view. See his Truth and Method (New York: Seabury Press, 1975). His basic credo is that understanding (Verstehen) can occur only within the context of everyday living and that we interpret art not outside of our actual experiential context but very much within it.
In this book I use **contextualism** to mean that all events we perceive are greatly influenced by their context. It also stresses the interconnection of the five principal aesthetic fields of applied media aesthetics: light and color, two-dimensional space, three-dimensional space, time/motion, and sound. Finally, it helps organize the discussion of the great variety of aesthetic elements in each field and their influence and dependence on one another.


12. Being literate, or the term **literacy** in this context, does not mean the ability to read and write but rather having achieved proficiency and polish in some area of knowledge. **Media literacy** refers to a basic knowledge of how, for example, video structures pictures and sound for specific purposes. See Paul Messaris, *Visual Literacy: Image, Mind, and Reality* (Boulder, CO: Westview Press, 1994).


14. Compare the convincing argument that it is the information systems in general and the media specifically that shape media content rather than the other way around. Some of the classic arguments are published in Joshua Meyrowitz, *No Sense of Place* (New York: Oxford University Press, 1985), pp. 13–16.


18. Wassily Kandinsky, *Point and Line to Plane*, trans. by Howard Dearstyne and Hilla Rebay (New York: Dover, 1979). This work was originally published as *Punkt und Linie zu Fläche* in 1926 as the ninth in a series of 14 Bauhaus books edited by Walter Gropius and László Moholy-Nagy.


Light is essential to life. It is necessary for most things to grow. It is the key element of visual perception, and it orients us in space and time. It also affects our emotions. Light is the agent that makes things visible. When we look at our surroundings, we receive a multitude and variety of light reflections. Each reflection has a certain degree of light intensity and complexity. The intensity variations appear to us as light or dark areas—as light and shadow—and the complexity as color.

Of course, we perceive light reflections as actual things. Most likely, we do not say, “I see the light variations that are reflected off these different surfaces.” Rather we say, “This is an automobile.” Often we conceive light to be the property of the objects themselves. We speak of light and dark hair, a red ball, a green frog, and a bright sky.

Video and film, as well as computer images, are pure light shows. In contrast to the theater, for example, where light is used simply to make things visible on-stage and to set a mood, the final images on the movie screen and on electronic screens consist of light. The materia of the theater—the stuff that makes theater—is people and objects in the real space and time of the stage. The materia of television and film, however, is light. The control of light is therefore paramount to the aesthetics of television and film. Lighting, then, is the deliberate manipulation of light and shadows for a specific communication purpose.

Before you try to manipulate light and shadows and use them creatively, you need to familiarize yourself with the nature of light, lighting purposes and functions, the nature of shadows, and the outer and inner orientation functions of lighting.

THE NATURE OF LIGHT

Light is a form of radiant energy. It consists of separate bits of energy—energy particles—that behave commonly as electromagnetic waves. It makes up a part of the total electromagnetic spectrum, which includes such other magnetic energy waves as radio waves, X-rays, satellite transmissions, and the waves in your microwave oven that heat up your coffee.
So-called white sunlight consists of a combination of light waves that are visible as various colors. When white sunlight is bent by a prism, it separates into a spectrum of clearly discernible hues: red, orange, yellow, green, blue, and violet. **SEE 2.1**

Because we can see the colors, that is, the various electromagnetic waves, light is usually defined as “visible radiant energy.” Actually, light is invisible: we can see it only at its source or when it is reflected. **SEE 2.2** For example, a beam of light that shoots across a clean room or studio remains invisible to our eyes and to the camera unless the light hits a reflecting agent, such as dust, smoke, an object, or a person.

If there were not a reflecting atmosphere, the sky would appear always dark, and you could see the stars even during the day. In deep space the astronauts see a black sky even in sunlight. If our surroundings did not reflect light, we would live in total darkness, much as if there were no light at all.

**LIGHTING PURPOSES AND FUNCTIONS**

Lighting is the deliberate control of light and shadows. The basic purpose of lighting is to manipulate and articulate our perception of the environment. It can also establish an aesthetic context for our experiences, a framework that tells us how we should feel about a certain event. Lighting helps us, or makes us, see and feel in a specific way.

Through lighting we can articulate our outer space/time environment and our inner environment—our emotions. Lighting reveals what objects look like, where they are located, and what surface textures they have. It also influences how we feel about a person or an event. Very much like music, lighting can bypass our usual cognitive perceptual screens—our rational faculty with its critical judgment—and affect us directly and immediately. Because lighting helps articulate our outer and inner environments, it has outer and inner orientation functions. Both functions depend to a great extent on the proper control of shadows. Let’s take a closer look at shadows before discussing the specific orientation functions of lighting.
**THE NATURE OF SHADOWS**

Ordinarily, we are not aware of shadows; we take them for granted. We readily accept the harsh and distinct shadows on a sunny day, the soft shadows on an overcast day, and the virtual absence of shadows under fluorescent lights. Only occasionally do we become more conscious of shadows. For example, we seek the shade when the sun gets too hot during an outdoor picnic, we adjust the reading lamp so that the shadow of our head or hand does not fall directly on the page, or we might chuckle when our shadow shows us an especially distorted image of ourselves.

When you are engaged in clarifying and intensifying an event through lighting, however, you become very aware of shadows and learn to use them for specific orientation tasks. It is not the basic illumination that defines the texture of people and things—it is the shadows. You will find that in critical lighting situations, you often need more lighting instruments for controlling the shadows than for making things visible.

Let’s look at an example. **SEE 2.3A** Both objects look like simple white discs; both are lighted with highly diffused light (by using floodlights), rendering them practically shadowless and revealing little more than their basic contour. As soon as you use a more directional light source, such as a Fresnel spotlight, and place it somewhat to the side of the object, you have no trouble distinguishing between the two. **SEE 2.3B** Because the directional light produces dense shadows, you can now see that the left object is a white ball and not a disc. But the object on the right remains evenly lighted, without any shadows. It looks like, and indeed is, a disc.

---

**2.3 Shadows Define Space**

a When objects are lighted “flat” with a highly diffused light source, such as a scoop or softlight, we see nothing more than two flat discs.

b With a more directional light source, such as a Fresnel or ellipsoidal spotlight, placed somewhat to the side of the object, we see that the object on the left is a sphere and the one on the right a flat disc.
2.4 Shadows Define Shape and Location

The attached shadows give us additional information about the true shape of the object (the hollow cone). The cast shadow tells us where the object is relative to its surroundings.

Now look at the next figure. **See 2.4** Without seeing any shadows (figure 2.4a), we perceive only the basic contour of the object—an inverted triangle with a curved top—but the true spatial nature of the object and its location relative to its environment remain ambiguous. As soon as we attach a shadow (2.4b), pretending that the main light source is coming from the right, we perceive the object as rounded and three-dimensional. An additional shadow on the top of the object (2.4c) reveals that the object is a cone and that it is hollow. The cone’s shadow that is cast on another surface tells us where the cone is in relation to the horizontal surface (a table) directly underneath it (2.4d); according to this shadow, the cone obviously floats above the table. The shadow in figure 2.4e is connected to the tip of the cone; we now see the cone as touching the table. Thus the initial spatial ambiguity has been drastically reduced by the various shadows.

**Attached and Cast Shadows**

If you take another look at figure 2.4, you will probably notice that some shadows are attached to the cone and others are relatively independent of it. An equally astute observer, Leonardo da Vinci, called these two types of shadows **attached** and **cast**.

**Attached shadows** An attached shadow is inevitably fixed to its object. No amount of wiggling or turning will remove the shadow from the object, assuming you keep it under the same lighting conditions. The attached shadow helps reveal the basic form of an object, but it can also fool you into perceiving what you normally expect to see. **See 2.5** Figure 2.5a shows an ornament that protrudes; figure 2.5b shows one that is indented. The major clues for such perceptions are the attached shadows. Now turn your book upside down and take another look.

**2.5 Attached Shadow Reversal**

a In this ornament the attached shadow is at the top of the circles. Because we naturally expect light to be coming from above, we see this ornament first to indent, then to protrude, then to indent and protrude again.

b This is the same ornament as in (a) except it is turned upside down. Now you will probably perceive the exact opposite: a protrusion first, then an indentation, another protrusion, then another indentation. By turning the book upside down, the ornaments will reverse once more.
Figure 2.5a now shows an indented ornament, and figure 2.5b shows a protruding one. Why? Because through lifelong experience, we assume that light comes from above rather than from below, so we expect the attached shadows to appear in the upper part of an indentation and the lower part of a protrusion. This perceptual habit is so strong that we will readily accept a change in the actual appearance of the object rather than the assumed direction of illumination.

If the proper perception of protrusions and indentations is crucial, you must light the object steeply from above to place strong attached shadows where we expect them to be. This type of space articulation is especially important for painters and graphic artists, who must suggest protrusions and indentations on a flat surface by painting in prominent attached shadows. Attached shadows help us primarily with interpreting an object’s basic shape and texture.

**Cast shadows** Whereas the attached shadow is always part of the actual object and is virtually glued to the side that is turned away from the light, the cast shadow always falls on something, in this case the road. **SEE 2.6** Other examples of cast shadows are the shadow of a power pole that stretches across the street, the shadow of a tree that falls on the grass beneath it and provides a cool spot for a picnic, or an airplane’s shadow moving across the landscape. Most cast shadows show a distorted shape of an object. **SEE 2.7** As you can see in figure 2.7, we can take a photograph of cast shadows without showing the people who cause them.

Another example of the independence of cast shadows is making shadow pictures on a brightly lighted wall. You can admire the cast shadows of your creations without ever looking at your hand. Although the shadows are projected onto the wall, they are not part of the wall and disappear as soon as you drop your hand.

Cast shadows help us locate an object relative to its surroundings. **SEE 2.8** The cast shadow indicates whether the object rests on the table or is suspended above it. Notice how the cast shadow becomes independent of the object and gets fuzzier at the edges as the object moves farther away from the table. **SEE 2.9**

Contrary to the attached shadow, which is inevitably fixed to its object, a cast shadow may be connected to the object that causes it or be totally free of it. **SEE 2.10** A cast shadow that is still connected to its object is called object-connected; one that is seen independent of its object is called object-disconnected. Note that object-connected cast shadows are not the same as attached shadows. In contrast to attached shadows, cast shadows become disconnected from the object as soon as you move it away from the surface on which it is resting. Figure 2.9 shows a good example: when the ball is lifted from the table, the cast shadow becomes object-disconnected; the attached shadow, however, remains on the ball.
Cast shadows are sometimes used to suggest a certain location, such as a jail cell, which is not actually shown.

We can use cast shadows to create or emphasize a dramatic event.

Although we are normally unaware of or unconcerned about cast shadows, we continually make spatial judgments by perceiving their general shape, intensity, and direction and are readily influenced by their dramatic implications and impact. Cast shadows can help break up large, monotonous surfaces and give an area more visual variety and interest. They can suggest a specific locale, add drama to an event, and even help tell time. See 2.11 and 2.12. The discussion of outer and inner orientation functions later in this chapter includes more specific information about the two types of shadows.

Not all cast shadows are desirable. The infamous microphone boom or camera shadow on someone’s face or on the living room wall is an ever-present menace during large studio productions. So are the multiple cast shadows in a scene that simulates illumination by sunlight or a single table lamp. When you see the lone lost and thirsty hero cast three long shadows onto the hot desert sand while desperately trying to make contact with his cell phone, you needn’t worry about him: he will have plenty of time to get a glass of water while the studio crew is resetting the lights to simulate the illumination coming from a single sun.

The distinction between attached and cast shadows is important not only in critical video and film lighting but also in creating scenes with a graphics generator or in working out computer-generated designs. Although ultimately linked logically, attached and cast shadows require separate and careful attention during the design phase. For example, you need to have attached shadows change positions on the object when the object moves relative to the (usually virtual) light source or when the light source moves relative to the object.

Figures 2.8 to 2.10 show that a cast shadow must also get larger and less dense when the object moves away from the surface on which the shadow is cast (or smaller and denser as the object moves toward such a surface). The constantly changing cast shadows are an important indicator of position change when you move through a virtual-reality environment. Just make sure that when simulating a light source, your three-dimensional (3D) software has both types of shadows move accordingly.

Falloff

We use the term falloff to mean two different yet related light/shadow relationships: the brightness contrast between the light and shadow sides of an object, and the rate of change from light to shadow.

Contrast If the brightness contrast between the lighted side of an object and the attached shadow is high, the falloff is fast. This means that the illuminated
Fast Falloff
Spotlights, which have a highly directional beam, produce fast falloff. Note that the light side and the dark attached-shadow side differ greatly in brightness. This results in high-contrast lighting.

Slow Falloff
A highly diffused floodlight produces slow falloff. There is little brightness contrast between the illuminated side and the shadow side. The attached shadow has become transparent.

Elimination of Falloff
When both sides are equally bright, there is no falloff: there is no longer a discernible shadow side, and the picture looks flat.

Change
Calling falloff “fast” or “slow” makes more sense when applied to the rate of change between light and dark. Look, for example, at the tops of the steps in figure 2.16: they are exposed to the sun and are very bright, but they suddenly turn into dense attached shadows at the risers. Such an abrupt change from light to shadow represents extremely fast falloff. Conversely, you automatically interpret such fast falloff on a surface as an edge.

Now imagine yourself moving, like Spider-Man, across the rounded surface of a domed building. You will move from bright sunlight to a hint of a shadow until you reach a dense area of attached shadow at the far side of the rounded building. Because such a change from light to dense shadow is much more gradual than on the steps, the falloff on a curved surface is slow. If the falloff on a curved surface is exceptionally fast, we have a tendency to perceive the rounded surface as an edge. To emphasize the rounded surface, you need to slow the falloff.

Controlling falloff
You can control falloff by using highly directional or diffused light for the basic illumination and by manipulating the amount of fill light.

Fast Falloff: Edge
The lighting on these steps shows fast falloff. The change from light to dark is sudden, signifying a sharp edge or corner.

Slow Falloff: Curved Surface
The light on the surface falls off gradually into its attached shadow. The surface of this building is obviously curved.
The directional beam of a spotlight (or the sun) causes sharp contrast between the illuminated area and the dense attached shadow. The resulting falloff is fast. Floodlights, on the other hand, produce slow falloff. The highly diffused and more omnidirectional spread of a floodlight not only illuminates the side of the object that is oriented toward the light but also “floods” the shadow side, rendering the attached shadow more or less transparent.

If the directional light source produces falloff that is too fast, with attached shadows so dense that you can no longer discern any detail (as in figure 2.13), you need to slow down the falloff to some degree. Slowing falloff means rendering the attached shadows somewhat transparent and reducing the contrast between light and shadow areas (see figure 2.14). You control such “contrast falloff” through various amounts of fill light. The more fill light you use, the more transparent the attached shadows become and the less contrast there is between light and dark.

Instead of using the customary fill-light instrument to slow down falloff, you can use simple reflectors. They are often used in outdoor shooting to make the sunlight do double duty—to work as both a key light and a fill light. When shooting outdoors many video camera operators prefer a foggy day to a brightly sunlit one. The fog acts as a giant diffuser of sunlight, producing soft, slow-falloff lighting with highly transparent shadows.

You may have noticed that it is not the illuminated side that tells us whether the light is “hard” (produced by spotlights) or “soft” (produced by floodlights) but rather the speed of the falloff and the density of the shadows. When looking at a close-up of the illuminated side of a face or an object, you cannot really tell whether the light used was a spotlight or a floodlight. When seeing fast falloff and dense shadows, however, you will immediately know that a hard (spot) light was used. When the falloff is slow and the shadows are transparent, a soft (diffused) light was used.

**OUTER ORIENTATION FUNCTIONS: HOW WE SEE AN EVENT**

Lighting orients us in space. It shows us what an object looks like: whether it is round or flat and has rough or smooth surfaces and round or sharp edges. It can also show us where the object is in relation to other things. It lets us know whether it is day or night, morning or noon, summer or winter. The use of lighting to articulate the outer environment is known as outer orientation. Thus we can identify three principal functions: spatial, tactile, and time.

**Spatial Orientation**

Lighting reveals the basic shape of an object and where it is located relative to its environment. The principal light source—the key light—and the attached shadows carry the major burden of fulfilling the basic shape function. The cast shadow indicates where the object is: whether it sits on a table or floats above it and whether it is close to the wall or away from it. Under certain circumstances, a cast shadow can give you a rough idea of what the object that caused the shadow looks like.

**Tactile Orientation**

Lighting for tactile orientation is very closely related to lighting for spatial orientation. Actually, texture is a spatial phenomenon because a texture, when sufficiently enlarged, resembles the peaks and valleys, ridges and crevasses of a rugged mountain range. The only difference is that lighting for space is done primarily to orient us better visually, whereas lighting for texture is supposed to appeal to our haptic sense—our sense of touch. As in lighting for spatial orientation, control of falloff is paramount.
To demonstrate the importance of falloff in texture, let us assume that the wrinkles and folds in a backdrop or curtain represent an enlarged surface texture. If you point a spotlight at the wrinkled side of a backdrop and direct the beam so that it hits the backdrop from the side, you will produce prominent, fast-falloff attached shadows; the wrinkles will be greatly emphasized. SEE 2.18 When illuminating the identical area directly from the front with a soft floodlight, however, the falloff is slowed down so drastically that the attached shadows become all but invisible to the camera. The backdrop now lacks texture and therefore looks taut. SEE 2.19

The same principle applies to lighting a face. You can use falloff control either to emphasize the texture of a face (wrinkles or beard stubble) or to de-emphasize it and make the skin look taut and smooth. Thus, if you have to light the face of a rugged adventurer, you go for fast falloff. In our society we seem to associate a man’s experience and masculinity, if not virility, with a moderate amount of wrinkles. SEE 2.20 This is not so with women, however. We seem to believe that women’s faces should look reasonably smooth regardless of age and experience. To emphasize the smoothness of a woman’s face, you obviously light for slow or no falloff. But notice that despite the strong fill light that eliminates the falloff, the back light prevents the picture from looking entirely flat. SEE 2.21 How slow the falloff should be (which translates into how much fill light you use) depends on the specific message you want to convey. For example, if you need to demonstrate the effectiveness of a new skin cream, you light the model’s face with extremely slow or no falloff. But to intensify a woman’s exhaustion after surviving a bad fall while rock climbing, you light for relatively fast falloff.

As adults we sometimes forget that our tactile, or haptic, sense is a very important (if not the most direct) means of perceiving our environment, of experiencing the nature of the objects around us. As infants we tend to learn as much, if not more, about our environment by touching as by looking, smelling, or listening. Only gradually, and after many warnings from our parents not to touch this or that, do we finally manage to drive the tactile sense underground. But the many do-not-touch signs in stores and especially in museums suggest that apparently we would still like to touch objects to get to know them better and to enrich our experience.

Johannes Itten, who taught the famous Basic Course at the Bauhaus in Weimar, Germany, from 1919 to 1923, put great emphasis on the study of textures. One of the texture exercises in the Basic Course involved making long boards on which a variety of materials with different consistency (steel wool, wood, cotton) were glued. The students would then run their fingers over these textures with their eyes closed. Itten found that through such systematic exercises, the students’ sense of touch improved to an amazing degree. In turn the students learned to appreciate texture as an important design element as well as an orientation factor to the materials used. Itten stated that through such texture exercises the students developed a real “design fever.”*
Control of light and shadows also helps viewers determine the time and even the seasons. In its most elementary application, lighting can show whether it is day or night. More specific lighting can indicate the approximate hour of the day or at least whether it is early morning, high noon, or evening. With certain subtle color changes, you can also suggest whether it is winter or summer.

Day and night In general, daytime lighting is bright and nighttime lighting is less so. Because we are used to seeing everything around us during the day but not at night, we keep the sky or background illuminated to indicate daytime. See 2.22 We leave it dark or only partially illuminated for nighttime. See 2.23 A daytime scene needs a great amount of all-around light with everything brightly illuminated, including the background. Nighttime lighting needs more specific and more selective fast-falloff illumination. Note that nighttime lighting does not mean “no light” but rather highly selective light with a minimum of spill. The light must also come from an obvious source, such as the moon or a streetlamp.

You can use the same lighting principle for indicating day or night indoors. For a daytime interior, light so that the background is bright and the rest of the interior is bathed in slow-falloff illumination. See 2.24 For nighttime the background

**Time Orientation**

<table>
<thead>
<tr>
<th>Image</th>
<th>Caption</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Daytime Scene" /></td>
<td><strong>2.20 Fast Falloff: Facial Texture Emphasized</strong> Highly directional hard spotlights hitting the face from a steep angle create fast falloff. The facial texture—the wrinkles, ridges, and hollows—is accentuated.</td>
</tr>
<tr>
<td><img src="image2.png" alt="Nighttime Scene" /></td>
<td><strong>2.21 Slow Falloff: Facial Texture Reduced</strong> When you want the skin to look smooth and wrinkle-free, you need to reduce, rather than emphasize, facial texture.</td>
</tr>
<tr>
<td><img src="image3.png" alt="Outdoor Illumination: Day" /></td>
<td><strong>2.22 Outdoor Illumination: Day</strong> In this daylight scene, the sky and the walls of the building are light. All recessed areas (doors and windows) are in the shade and therefore dark.</td>
</tr>
<tr>
<td><img src="image4.png" alt="Outdoor Illumination: Night" /></td>
<td><strong>2.23 Outdoor Illumination: Night</strong> In this night shot, the sky is dark as are all areas that are not illuminated by the street lamp. Because the lights are on inside the building, the windows appear light.</td>
</tr>
</tbody>
</table>
is predominantly dark, and the lighting in the rest of the interior becomes more selective. **SEE 2.25** In daytime lighting, the lamp is turned off and the window is light. In nighttime lighting, the lamp is on but the window is dark.

**Clock time** The common indicator of clock time is the length of cast shadows. As shown in figures 2.7 and 2.26, the early-morning or late-afternoon sun causes long cast shadows. **SEE 2.26** At high noon shadows are very short. Outdoors such shadows are usually cast along the ground. This is of little help, however, because the camera rarely shoots wide enough for us to see the shadows on the ground. You must therefore produce cast shadows somewhere in the background, such as on the wall of a building, so that the camera can see them even in fairly tight shots.

The same shadow requirement holds for shooting indoors. Rather than have the cast shadows fall on the studio floor, you should devote a fair amount of background lighting to producing cast shadows, or even slices of light, that cut across the background at the desired angle. Though it is illogical, we readily seem to substitute a slice of light for a cast shadow. **SEE 2.27** If clock time is critical to the plot, you may want to reinforce the lighting with other, less subtle cues, such as...
somesboby saying, “Oh, it’s five o’clock already” or by briefly intercutting a close-up of a clock. In any case, the lighting must correspond with such additional cues.

Also ensure that all the background cast shadows are consistent with the attached shadows. If the window of the room is the primary light source, with illumination coming from the left, attached and cast shadows should be on the opposite side. Because the indoor lighting setup usually requires several instruments, most of which come from somewhat different directions and angles, you can easily end up with a variety of cast shadows falling in different directions. For example, the hand that turned on the single light bulb in the cheap motel room should not cast multiple shadows on the wall. You must keep all distracting cast shadows out of camera range and show only one prominent cast shadow that falls opposite the primary light source.

**Seasons**

Generally, the winter sun is weaker and colder than the summer sun, so the light representing the winter sun should be slightly more bluish. The winter sun also strikes the earth’s surface from a fairly low angle, even at noon; this makes cast shadows in winter longer and not quite so dense. A slightly diffused light beam helps create softer winter shadows. Because snow reflects a moderate amount of diffused light, the falloff is somewhat slower than in a similar outdoor scene without snow.

---

**INNER ORIENTATION FUNCTIONS: HOW WE FEEL ABOUT AN EVENT**

So far we have explored outer orientation functions of lighting: light and shadows manipulated to articulate the outer environment. You now know the aesthetic principles of lighting to show what an object looks like, what texture it has, and where it is located in space and time. Such articulation is a major factor in structuring **screen space**. But you can also use light to articulate the inner environment—that of our feelings and emotions—a lighting function we call inner orientation.

The specific inner orientation functions of lighting are establishing mood and atmosphere, above- and below-eye-level key-light position, predictive lighting, and the use of light and lighting instruments as dramatic agents.

**Establishing Mood and Atmosphere**

Much like music, lighting can have an intuitive effect, influencing our emotions directly. Some lighting makes us feel happy, some sad, and some uncomfortable or even frightened. The two major aesthetic lighting techniques for establishing mood and atmosphere are high-key and low-key lighting and above- and below-eye-level key-light position.

**High-key lighting**

This kind of lighting has nothing to do with the position or intensity of the principal illuminating source, called the key light. Nor does it mean that the key light shines down on the scene from above. **High-key lighting** means that the scene has an abundance of bright, usually slow-falloff illumination and a light background. Television news sets and interview areas, game shows, and many situation comedies have high-key illumination. High-key lighting reflects normalcy and an upbeat feeling. **SEE 2.28**

**Low-key lighting**

The fast-falloff illumination of **low-key lighting** is highly selective, usually leaving the background as well as part of the scene predominantly dark. This type of lighting has less overall light and fewer light sources than high-key lighting. Scenes that deal with caves, medieval dungeons, submarine
interiors, or nighttime exteriors or interiors normally call for low-key lighting. Whereas high-key lighting is “up,” low-key lighting is “down.” There is hardly a crime or mystery show that does not use some form of low-key lighting. **SEE 2.29**

### Above- and Below-Eye-Level Key-Light Position

These positions are, indeed, meant to tell where the key light is located. As stated earlier, we usually expect the principal light source to come from above the object, producing attached shadows underneath protrusions. This normal lighting is achieved by having the key light illuminate a person from **above eye level**. Note that **eye level** refers to the eyes of the subject or the middle portion of the object lighted. As soon as the principal light source strikes the face from **below eye level** (sometimes called reverse modeling), the attached shadows reverse vertically and are exactly opposite their expected positions. Because we are so used to seeing the light come from above eye level, such a shadow reversal affixes to the outer disorientation an inner disorientation, which translates into surprise, suspicion, or fear. This lighting technique, sometimes called horror lighting, is as blatant as it is effective. **SEE 2.30 AND 2.31**
Chapter 2

Whom would you trust more as a news anchor, a computer salesman, or an attorney—the person on the left or the one on the right? Most of us would probably opt for the person on the left (figure 2.30). Why? Because when seeing the lighting effect in figure 2.31, we probably would not take the time to trace the cause of the person's strange appearance to a below-eye-level key-light position and the subsequent vertical reversal of the attached shadows; we would simply label the person as weird, dangerous, or at best untrustworthy. This seemingly minor position change of the key light has a decisive influence on our perception of the person's credibility if not character. Worse, we have a tendency to extend such unexamined aesthetic manipulations of appearance to the person's entire behavior and psychological makeup.

Predictive Lighting

The predictive lighting technique helps portend, however subtly, a coming event. Light that changes from high-key to low-key, from general to specific, from above eye level to below eye level or, more obviously, a flashing light can signal how the event will go. See 2.32 The lighting changes from normal slow-falloff lighting to fast-falloff lighting, with the key light moving quickly from camera-left to camera-right, past the person, to partial background lighting. Such drastic lighting changes give the viewer a strong clue to some unpleasant future event, even if the lighted character may be unaware of it. As in all good drama, we now know something the character does not, and we either empathize with her predicament or anticipate her deserved doom.

Similar predictions of trouble are possible when you change the lighting of a party from high-key to low-key while maintaining the seemingly happy and innocent mood of the festivities, with all other aesthetic devices remaining the same. By reversing this procedure and changing from “down” to “up” lighting, you can predict the happy ending (the famous ray of light), even if the other aesthetic elements are still signaling disaster.

You can also use moving light sources in predictive lighting. You may have seen some version of a scene in which the night watchman discovers that something is not quite right in Building 47, which houses top-secret documents. We see him getting up, looking left and right, and turning on his flashlight. We see the flashlight beam creeping nervously down the long, dark hallway until it

We can learn a great deal about lighting from classic black-and-white motion pictures. Lacking color, the lighting of black-and-white films had to be especially expressive. For example, Federico Fellini (1920–1993) used light as a dramatic agent with great virtuosity in his timeless films La Dolce Vita and 8½.

In La Dolce Vita, he shows a television remote unit covering the events of an alleged miracle. Young children claim to have seen Holy Mary descend from heaven and heard her speak to them. There is a great amount of confusion. Fellini cuts in close-ups of lighting instruments being turned on, shining their cold, controlled beams over the highly emotional, ecstatic crowd. To counterpoint even more the discrepancy between the emotionally charged crowd, which represents the uncritical world of blind faith, and the analytical and soulless modern age, as symbolized by the lighting instruments and cameras, he shows a tight close-up of a huge Fresnel lens bursting in the first seconds of a chilling downpour.

In 8½ Fellini uses many lighting instruments arranged in large circles, illuminating the representatives of humanity who, following the director's orders, march willingly like circus clowns within the lights' periphery. The lighting instruments and the light, which occasionally shines directly into the camera, are a strong reminder that when properly "enlightened" we may discover that we are all part of a big cosmic joke that some superior power occasionally plays on us. Unless we embrace humanity and "join the show," we remain alienated from the circus of life and face an empty existence.
2.32 Predictive Lighting

The lighting change in this picture series increases the tension of the event and suggests a somewhat ominous outcome for the woman.

finally reveals—you guessed it—the broken lock, the open file cabinets, and papers strewn on the floor.

As with any application of contextual aesthetics, predictive lighting rarely operates alone; it usually works in conjunction with appropriate sounds, suspenseful music, and the like. In fact, oncoming changes of events are more commonly introduced by predictive sounds than by predictive lighting. Flashes of white light, accompanied by a generous amount of whoosh sound effects, are some of the less subtle examples of predictive lighting and sound.

Light and Lighting Instruments as Dramatic Agents

You can use the light source itself as an effective dramatic agent—an element that operates as an aesthetic intensifier in a scene. By showing the actual light source—the sun, a spotlight, or a flashlight—you can intensify the scene, assuming that it is set up properly for such intensification. Well-known, and often well-worn,
examples are the close-up of the flashing red and blue lights on top of a police car, the strobe lights of a rock concert, or the searchlight from a prison tower that, by shining into the camera, not only searches for the escapees on the prison grounds but also ruthlessly invades your personal space and privacy as a viewer. The dim overhead lights in a garage or the on/off blinking motel sign are other examples of using lighting as a dramatic agent. Movies about extraterrestrials depend heavily on light as a dramatic agent. The creatures from space are usually introduced and dismissed as mysterious light beams, regardless of their eventual metamorphosis or whether they turn out to be friend or foe.

**SUMMARY**

Light orients us in space and time and influences how we feel about a thing, a person, or an event. The areas that need special attention when dealing with light are the nature of light, lighting purposes and functions, the nature of shadows, and the outer and inner orientation functions of lighting.

Light is a form of radiant energy that commonly behaves as electromagnetic waves. Perceptually, light is invisible except at its source and when it is reflected by an object.

Lighting is the deliberate control of light and shadows. Through lighting we can articulate our outer space/time environment and our inner environment—our emotions. These outer and inner orientation functions of lighting (how we see and feel) are primarily dependent on the proper control of shadows.

There are two types of shadows, attached and cast. The attached shadow is inevitably fixed to its object on the opposite side of the principal light source. The cast shadow is independent and may be connected to its object or disconnected from it. The attached shadow reveals the basic form of the object; the cast shadow tells where the object is in relation to its surroundings.

Falloff is the brightness contrast between the light and shadow sides of an object. The term also refers to the relative rate of change from light to dark (shadow). Slow falloff means that the illuminated side and the shadow side have very little brightness contrast or that the change from light to shadow area is gradual. Fast falloff means that the illuminated side and the shadow side have a great brightness contrast and that the change from light to shadow is abrupt.

The outer orientation functions of lighting include spatial, tactile, and time. The spatial orientation functions are to reveal the basic shape of the object and its location relative to its environment. The tactile orientation function means that fast-falloff lighting is employed to reveal and emphasize the object's surface texture and that slow- or no-falloff lighting is used to reduce or eliminate texture. Time orientation is achieved primarily by controlling the relative brightness of the background and the length and angle of cast shadows. A light background suggests daylight; a dark background, nighttime. Long cast shadows suggest early morning or late afternoon; short shadows suggest high noon. The winter sun is weaker and slightly more bluish than the summer sun.

The inner orientation functions of lighting include establishing mood and atmosphere, above- and below-eye-level key-light position, predictive lighting, and the use of light and lighting instruments as dramatic agents. Mood and atmosphere are affected by low- and high-key lighting as well as by above-eye-level and below-eye-level lighting. High-key lighting has an abundance of light, and the distribution of the light is nonspecific; the falloff is slow. With low-key lighting, the overall light level is low and the lighting illuminates specific areas; the falloff is fast. An above-eye-level key-light position places attached shadows
in a normal position; a below-eye-level key-light position reverses them vertically into an abnormal position. We usually interpret below-eye-level lighting as frightening or dangerous.

Predictive lighting refers to a lighting change that portends an upcoming event. One type of predictive lighting is a moving light source, such as a night watchman’s flashlight, that reveals something.

Light and lighting instruments can be used as dramatic agents: the light is used directly as an aesthetic intensifier. The flashing lights on a police car or during a rock concert are examples of such intensifiers.

NOTES

In video, film, and computer displays, as in painting and still photography, we must project the 3D world onto a two-dimensional surface. Fortunately, the camera and its single-lens optical system transact such a projection automatically. It is also fortunate that we are willing to accept such a projection as a true representation of all three dimensions: height, width, and depth. SEE 9.1

Although in video and film the camera and the event are often in motion, we can nevertheless draw on the firmly established and proven aesthetic rules of painting to enhance the process of projecting a three-dimensional world onto a two-dimensional plane. Note, however, that these techniques do not show us a true stereoscopic projection of the world. Stereoscopy, which is an old technique of displaying the 3D world on a 2D plane, uses two-lens photography or computer programs that simulate our actual two-eyed perception process. Stereoscopic projection is making a moderate comeback in film and has spawned a flurry of activity in video and digital animation.

In this chapter we examine four specific three-dimensional field areas: the z-axis, monocular graphic depth factors, depth characteristics of lenses, and 3D binocular depth. We also discuss the stereo 3D illusion and virtual reality.

THE Z-AXIS

As you probably remember from geometry class, the x and y coordinates precisely locate a point in a two-dimensional plane such as the frame of the picture opposite this page or a video screen. You can describe the width of the screen as the x-axis and the height of the screen as the y-axis. A point within the screen can be assigned an x-value, indicating its relative position along screen width, and a y-value, indicating its position relative to screen height. SEE 9.2

In the three-dimensional model, the z-axis is added, which describes depth. The z-axis value describes a point located away from the frontal plane—in our case, how far an object seems to be from the camera. You learned about the z-axis in chapter 7 in the context of index and motion vectors that extend from the camera to the horizon and vice versa. SEE 9.3

Amazingly enough, the illusory third dimension—depth—proves to be the most flexible screen dimension in film and especially in video. Whereas the
9.1 Projection of 3D Space onto a 2D Plane
In conventional video and film, as in painting and still photography, we project the three-dimensional event onto a two-dimensional surface.

9.2 X and Y Coordinates
The x and y coordinates locate a point precisely within an area, such as the screen. A point within the screen can be assigned an x-value, indicating where it is located on the x-axis (screen width), and a y-value, indicating its position on the y-axis (screen height).

9.3 Three-Dimensional Model
To locate a point precisely within a described volume, the z-axis—describing depth—becomes an essential dimension. The z-value describes how far a point is located away from the frontal plane (the screen).

9.4 Z-axis Dimension
Although the z-axis—the depth dimension—is illusory in television and film, it is aesthetically the most flexible screen dimension.

Notice that without 3D stereovision or hologram projection (as is the case with all single-lens film, video, and computer displays), we perceive the z-axis as originating from the screen and going backward, from the camera lens to the horizon. The closest object seems to lie on the screen surface; it does not extend toward the viewer.
The Three-Dimensional Field: Depth and Volume

In stereovision or a hologram, the z-axis extends not only to the horizon but also to the viewer: objects appear to extend out from the screen toward the viewer. We call the normal 2D z-axis, which extends from camera or screen to horizon, the \( z_h \)-axis (z-horizon axis); the one that extends from screen to viewer is the \( z_v \)-axis (z-viewer axis). As you will see later in this chapter, the double z-axes are an important issue in the stereo 3D discussion. **SEE 9.6**

**MONOCULAR GRAPHIC DEPTH FACTORS**

In this context monocular depth factors refer to how we articulate the single-lens z-axis, which extends from camera to horizon, to create picture depth. But how, exactly, can we create the illusion of monocular depth on the two-dimensional plane of the screen? **SEE 9.7** Examine the figure and try to identify the many factors that contribute to the illusion of depth on the 2D picture space.
You probably noticed that some objects are partially hidden by other objects. The large arch overlaps parts of the building, which in turn overlap other parts. Whatever is overlapped must be behind the object doing the overlapping. 

SEE 9.8 The farther away the doorways are, the smaller they appear and the higher they are in the picture field. The small size of the woman under the third arch also gives a clue to how far away she is from the first arch in the foreground and how large these arches are. 

SEE 9.9 Parallel lines, such as the cobblestone lines in the center, appear to converge in the distance. 

SEE 9.10 The cobblestones in the back look less distinct and more crowded than the ones in front; and the car and the house seen through the last arch are so indistinct that they almost blend together. 

SEE 9.11 Finally, as you learned in chapter 2, the attached and cast shadows give the objects and the entire scene volume—that is, the presence of a third dimension. 

SEE 9.12 This section discusses five such graphic depth factors: overlapping planes, relative size, height in plane, linear perspective, and aerial perspective.

**Overlapping Planes**

The most direct graphic depth cue is an overlapping plane. When you see one object partially covered by another (technically called occlusion), you know that the one doing the covering must be in front of the one that is partially covered. 

SEE 9.13 AND 9.14
**Relative Size**

If you know how big an object is or can guess its size by contextual clues (such as other known objects), you can tell approximately how far the object is from the camera by the *relative size* of the screen image. The larger a subject or an object appears relative to the screen borders, the closer it seems to us. Appropriately, we call this framing a *close-up*. **SEE 9.15**

The smaller a subject or an object appears relative to the screen borders, the farther away it seems. This seems to be a hardwired response. **SEE 9.16**

**9.13 Overlapping-Planes Principle**

Object A is partially covering object B, which is partially covering object C. Although all three figures obviously lie on the same plane (this page), A seems to be in front of B, which seems to lie in front of C but behind A.

**9.14 Depth through Overlapping Planes**

Any object that is partially blocked from our view by another object must lie behind that object. Even with other depth cues missing, we perceive a third dimension by readily assigning partially overlapping objects a "behind" or "in front of" position.

**Medieval painters** relied heavily on overlapping planes to indicate depth. In this detail from *The Meeting at the Golden Gate*, one of the many excellent frescoes in the Arena Chapel in Padua, Italy, by Florentine painter and architect **Giotto di Bondone** (ca. 1267–1337), we can see how effectively overlapping planes were used to indicate depth. The only depth confusion arises from the merging halos of Joachim and Anna—caused by the fading of the contour of Joachim’s halo at the point of overlap.

**9.15 Relative Size: Close-Up**

The larger the object or subject appears within the screen—that is, the more area it takes up relative to the screen borders—the closer it seems to us. Appropriately, we call this framing a close-up.

**9.16 Relative Size: Long Shot**

The smaller the object or subject appears within the screen, the farther away it seems. We call this framing a long shot.
9.17 Interpreting Object Size as Distance: Far
The man seems farther away from us than the woman because his screen image is considerably smaller than hers.

9.18 Interpreting Object Size as Distance: Close
The man seems much closer to the woman now because his screen image is almost as large as hers.

9.19 Height in Plane: Camera Parallel to Ground
In the absence of contradictory distance cues and with the camera shooting parallel to the ground, the people seem farther away the higher up they move toward the horizon in the picture plane. As they line up along the x-axis, they do not move up in the picture plane any longer.

9.20 Height in Plane: Camera from Below
When the camera does not shoot parallel to the ground, height-in-plane distance cues are no longer valid.

Height in Plain
Assuming that no contradictory distance cues are evident and that the camera is shooting parallel to the ground, you will perceive people and objects as being more and more distant the higher they move up in the picture field. This distance cue operates only until they have reached the horizon line. This is known as height in plane. See 9.19
Because of the mobility of the camera, however, which causes the horizon line to shift constantly within a shot or from shot to shot, the height-in-plane distance cue is not always reliable. See 9.20
One of the potential problems with stereo 3D is that changing the virtual distance between object and viewer does not necessarily change the object size. This spatial paradox is discussed in detail in chapter 10.

If you know that two objects are similar or identical in size, you perceive the smaller screen image as being farther away and the larger screen image as being closer. See 9.17 In the absence of contradicting contextual clues, we automatically interpret the smaller screen image of the man as being relatively far away from the woman, rather than being unusually small. The more comparable the head sizes, the closer the subjects seem to stand to each other along the z-axis. See 9.18
ground. Especially when shooting up or down a large object, the height-in-plane cue is no longer a valid depth indicator.

**Linear Perspective**

This is among the more powerful and convincing graphic depth factors. In a *linear perspective*, all objects look progressively smaller the farther away they are, and parallel lines converge in the distance, with the vertical and horizontal lines becoming more crowded as they move away from the observer (camera). See 9.21

All parallel lines converge and stop or disappear at the vanishing point, which always lies at eye level or camera level on the horizon line. See 9.22

To find the horizon line and eye level, simply stand erect and look straight forward or point the camera parallel to the ground. Assuming that your index

---

**9.21 Linear Perspective**

In this architect’s drawing of an Italian palazzo, all the prominent horizontal lines (graphic vectors) converge at one point. We call this perceptual phenomenon linear perspective.

*Courtesy of the DeBellis Collection, San Francisco State University*

**9.22 Vanishing Point**

The point at which all parallel lines converge and discontinue (vanish) is aptly called the vanishing point. The vanishing point always lies at eye (or camera) level on the horizon line.

---

If we look down on an object, the eye (camera) level is above the object. Therefore the horizon line, and with it the vanishing point, lies above the object. We see the object—in this case, the bridge—from above.

If we are below the object and look up, the eye (camera) level, the horizon line, and the vanishing point lie below the object. We see the object—the bridge—from below.

If we see two sides of a building, with one corner closest to us, we perceive two sets of converging lines going in opposite directions. We have, therefore, two vanishing points. But note that the two vanishing points lie on the horizon line; after all, we look at both sides of the building from the same eye level. This is called two-point perspective.
vector runs parallel to the ground, the horizon line moves up or down with your eyes (camera) regardless of whether you are kneeling on the ground, standing on a ladder, or pointing the camera out a helicopter window. **SEE 9.23**

Now take another look at figure 9.21. Can you tell from which height the artist looked at the palazzo? Was he looking at it from the street level? Sitting in a chair? Standing up? Perhaps from the balcony or window of an unseen building opposite the clock tower?

If you chose the window or balcony, you estimated the artist’s correct position. As you can clearly see, the parallel lines converge at a vanishing point that lies above the palazzo near the roofline of the clock tower building (see figure 9.22). The artist must therefore have looked at the building from that position.

Also note that the arches and windows of the building seem to lie closer together the farther away they are from the observer (see figure 9.21). As described later in this chapter, this crowding effect is especially noticeable when using a wide-angle lens. **SEE 9.24** Many painters have used this crowding effect, or texture, to simulate depth. You can apply this principle just as effectively with computer-generated graphics. As you can see in the next figure, we interpret the progressively narrower distances between horizontal lines as screen depth rather than shrinking line space. **SEE 9.25**

### 9.23 Horizon Line
The horizon line is an imaginary line parallel to the ground at eye level. More technically, it is the plane at right angles to the direction of gravity that emanates from the eye of the observer at a given place. If you want to find the eye level and the actual horizon line, simply stand erect and look straight forward. The eye level and the horizon line are in the distance where you do not have to look up or down.

### 9.24 Crowding Effect through Texture
Notice how the sunflowers appear more and more crowded the farther away they are from the camera. This crowding effect is an important depth cue.

### 9.25 Depth through Crowding
In this computer-generated image, the illusion of depth is generated by a crowding effect of horizontal lines. The lines are simply drawn progressively closer the higher up they move in the picture plane.

**Attempts at using** linear perspective to create the illusion of depth on a two-dimensional surface were made by many painters long before the Renaissance, but it was not until the first half of the fifteenth century that Italian artists established scientific laws of linear perspective, such as the horizon line and the vanishing point. With this woodcut one of the masters of the Renaissance, German painter **Albrecht Dürer**, illustrated some of the techniques used by the artist to ensure correct foreshortening (linear perspective).

Taken from *The Complete Woodcuts of Albrecht Dürer*, edited by Dr. Willi Kurth, republished in 1963 by Dover.
**Forced perspective**  Because we tend to interpret image size and convergence of lines with relative distance, we can generate the impression of distance by having parallel lines converge “faster”—more readily—and make distant objects appear smaller than we would ordinarily perceive. Such an artificial forcing of linear perspective is called, appropriately enough, *forced perspective*. One of the more striking applications of such a forced perspective is the grand staircase in one of Hong Kong’s luxury hotels. The wide staircase seems to curve up to the mezzanine in a long, impressive sweep. But when you actually climb the stairs, you will notice that they gradually narrow to less than half of their original width about a third of the way up. As you will discover in this chapter, we can achieve the same effect with the proper choice of lenses.

**Aerial Perspective**

A certain amount of moisture and dust is always in the atmosphere. We therefore see objects that are close to us somewhat more sharply than those farther away, a phenomenon known as *aerial perspective*.  

SEE 9.26 In fog this difference in sharpness and image density between foreground and background is especially pronounced. Colors also lose their density and become less saturated the farther away they are from the observer (camera). Outdoors, distant colors take on a slightly blue tint.2

When creating scenery, for example, you should paint the background objects slightly bluer and less sharp than the foreground objects. This greatly enhances the illusion of depth.

Generally, warm hues seem to advance and cold hues seem to recede. Highly saturated colors seem closer than less saturated colors. Assuming a fairly dark background, the brighter colors (higher position on the brightness scale) seem closer, and the less bright colors (lower position on the brightness scale) seem farther away. We can say that the more the brightness of an object assumes the

**9.26 Aerial Perspective**

Notice how the foreground objects in this picture are relatively sharp and dense and that the background objects become progressively less clear and less textured the farther away they are.
So far as aerial perspective is concerned, objects with warm, highly saturated colors that are the opposite of the background brightness seem closer to the viewer than objects with cold, less saturated colors that are similar to the background brightness. **SEE 9.27**

**DEPTH CHARACTERISTICS OF LENSES**

The optical characteristics of lenses can greatly enhance or hinder the illusion of a third dimension on the video or movie screen. Moreover, your choice of lens is important in achieving the certain “feel” of a screen event—whether buildings or objects look squeezed or stretched or whether the z-axis looks compressed or elongated. Synthetic computer images generally simulate the depth characteristics of lenses in their manipulation of the third dimension.

Before discussing the psychological impact of such space manipulation, we review the basic depth characteristics of wide-angle and narrow-angle lenses.

Note that our use of the terms wide-angle (or short-focal-length) and narrow-angle (or long-focal-length or telephoto) lenses includes the wide- and narrow-angle positions of the zoom lens. To put the zoom lens in the extreme wide-angle position, you zoom all the way out. To put it in the extreme narrow-angle, or telephoto, position, you zoom all the way in. The so-called normal lens position is in the middle of the zoom range. For emphasis we concentrate here on the extreme wide-angle and narrow-angle lens positions.

The focal length of lenses influences four principal graphic depth factors: overlapping planes, relative size, linear perspective, and aerial perspective.

**Overlapping Planes: Wide-Angle Lens**

Although the wide-angle lens does not rely on overlapping planes as much as the narrow-angle lens does, it cannot avoid showing them. Because the objects along the z-axis look more stretched out with the wide-angle lens, it renders overlapping planes less essential as a depth indicator. **SEE 9.28**

**Overlapping Planes: Narrow-Angle Lens**

The narrow-angle lens does just the opposite of the wide-angle lens, making objects appear closer together along the z-axis than they actually are. Because the narrow-angle lens enlarges the background objects where things look crowded, foreground and background objects look similar in size. Consequently, they ap-
Overlapping Planes: Wide-Angle Lens
Overlapping planes are reduced in prominence, but are not eliminated, with the wide-angle lens.

Overlapping Planes: Narrow-Angle Lens
With a narrow-angle lens, overlapping planes are a major depth cue.

Relative Size: Wide-Angle Lens
The wide-angle lens greatly exaggerates relative size. Objects that lie close to the camera photograph as relatively large, whereas similar objects positioned on the z-axis only a short distance behind the close object show up in a dramatically reduced image size. See 9.30 The image size of the foreground tugboat is relatively large, and the one just a short distance behind it is relatively small. This great difference in relative size is lessened only at the far end of the z-axis. Because image size is an important distance cue, we interpret this difference as meaning that the background object is farther behind the foreground object than it really is. Thus the wide-angle lens stretches the virtual z-axis.

Relative Size: Narrow-Angle Lens
When the same scene is photographed with a narrow-angle lens, the two boats appear closer together than they really are. Objects positioned along the z-axis look squeezed, and the z-axis itself appears shorter than you would ordinarily see.

Because of the similarity in size of foreground and background objects, overlapping planes become a major depth cue for separating one object from another and, ultimately, separating foreground, middleground, and background objects. See 9.29

Relative Size: Narrow-Angle Lens
The narrow-angle lens enlarges the background image so drastically that the tugboats as well as the background hills seem much closer together than in figure 9.30 although their actual positions along the z-axis have not changed.
Chapter 9

Distance is identical to that in figure 9.30. Now you know why: The narrow-angle lens enlarges the background, making the second object appear close in size to that of the foreground object. We translate this similarity in size as relative proximity. The narrow-angle lens shows objects placed along the z-axis squeezed; the z-axis therefore appears shortened.

This compression effect of the long (or narrow-angle) lens is very apparent when you shoot the same row of columns with both a wide-angle and a narrow-angle lens. Using the wide-angle lens, the columns quickly diminish in size the farther away they are from the camera; they seem comfortably stretched out. See 9.32 But when you shoot the same scene using a narrow-angle lens, the image size of the background columns is almost the same as that of the foreground columns. They now seem closer together than they really are; they no longer feel graceful but instead look massive and crowded. See 9.33

Linear Perspective: Wide-Angle Lens

The wide-angle lens accelerates the convergence of parallel lines; that is, they seem to converge more quickly than when seen normally, thereby giving the illusion of stretching an object or a building. The z-axis space appears elongated. See 9.34

Linear Perspective: Narrow-Angle Lens

In contrast, the narrow-angle lens inhibits the normal convergence of parallel lines and thus reduces the illusion of depth through linear perspective. It also squeezes space and makes the doors appear narrower and closer together than they actually are. See 9.35

By now you should have no problem distinguishing between the wide-angle and narrow-angle shots of a piano keyboard. See 9.36 and 9.37 The wide-angle lens makes the graphic vectors of the keyboard converge much more drastically than when shot with a narrow-angle lens. The “wide-angle” keyboard looks longer, with the keys farther away from the camera looking distinctly smaller. The “narrow-angle” keyboard, on the other hand, does not seem to converge much toward a vanishing point. In fact, the keys farthest from the camera look almost as big as the ones closest to it. This makes the keyboard look short and squeezed. Were you to watch somebody playing quick runs up and down the keyboard, the wide-angle lens would exaggerate such dexterity; the narrow-angle lens would reduce such motion and, with it, the pianist’s virtuosity.
The Three-Dimensional Field: Depth and Volume

9.34 Linear Perspective: Wide-Angle Lens
The wide-angle lens makes parallel lines converge much “faster” (more drastically) than when seen normally.

9.35 Linear Perspective: Narrow-Angle Lens
The narrow-angle lens “retards” our normal expectations of parallel lines converging. The horizontal lines do not converge as readily as with a normal or wide-angle lens.

9.36 Piano Keys: Wide-Angle Lens
When shot with a wide-angle lens, the piano keys reduce drastically in size the farther they are from the camera.

9.37 Piano Keys: Narrow-Angle Lens
When shot with a narrow-angle lens, the same keys look squeezed. The keys at the far side of the z-axis look almost as big as those that are close to the camera.

Working with Aerial Perspective
You can achieve aerial perspective by manipulating the depth of field—the area along the z-axis that appears in focus—and by making use of selective focus, that is, focusing on only a specific area along the z-axis.

When objects are placed at different distances from the camera along the z-axis, some of them will appear in focus and some will be out of focus. The portion of the z-axis in which the objects appear in focus—depth of field—can be shallow or great. In a shallow depth of field, only a relatively small portion of the z-axis shows objects in focus. In a great depth of field, a large portion of the z-axis shows objects in focus. The depth of field depends on the focal length of the lens, the lens aperture (iris opening), and the distance from the camera to the object.

Assuming that you shoot with a wide-angle lens under normal light levels and do not move the camera extremely close to the target object, the depth of field will be great. A narrow-angle lens gives a shallow depth of field. Generally, wide shots have a great depth of field; close-ups have a shallow one. Take another look at figures 9.36 and 9.37. The wide-angle lens shows the whole keyboard in
focus. When shot with the narrow-angle lens, the keys closest to the camera are out of focus because the camera was focused on the middle part of the keyboard.

**Aerial perspective: wide-angle lens** Because the wide-angle lens generates a great depth of field, it de-emphasizes aerial perspective. In a great depth of field, most of the articulated z-axis appears in focus (see figure 9.36). This means you cannot easily focus on only one spot along the z-axis while keeping the rest of the z-axis out of focus. A great depth of field is obviously advantageous when covering news, where you normally have little time to achieve optimal focus. Although a misnomer, a great depth of field is also called deep focus.

**Aerial perspective: narrow-angle lens** The narrow-angle lens has a shallow depth of field and thus emphasizes aerial perspective. Once you focus on an object using a narrow-angle lens, the areas immediately in front and in back of the object are out of focus. Even a slight position change of camera or object along the z-axis will necessitate refocusing (see figure 9.37).

Although it is difficult to keep a moving object or camera in focus in a shallow depth of field, the advantage of this critical focal plane is that you can use selective focus to emphasize events. You have probably noticed that shooting in a shallow depth of field has become stylish in video and film production. For example, you may see two out-of-focus people walking along the z-axis toward the camera until their images become focused in the depth of field. Such aerial-perspective maneuvers are often accompanied by a similar audio manipulation: when the people are out of focus, you can barely make out what they are saying, but once they are in focus, their dialogue becomes loud and clear. In an over-the-shoulder shot, you may initially see in focus the shoulder and head of the camera-near person, but the camera-far person, who is facing the camera, is out of focus. The camera will then switch the focus to the camera-far person, with the camera-near person being out of focus.

**Selective focus** The technique of selective focus allows you to choose the precise portion (plane) of the z-axis that you want to be in focus, with the areas immediately in front of or behind the focused object being out of focus. Contrary to a natural aerial perspective that occurs on a foggy day—where only the foreground object is “in focus,” that is, more clearly visible than the background objects—the optically induced aerial perspective using selective focus allows you to move the focused plane from the foreground to the middleground or background or the other way around.

In the next figure, we start out with the focus on the foreground person, with the people in the middleground and background out of focus. **SEE 9.38** Or you can feature the middleground and leave the foreground and background out of focus. **SEE 9.39** You can also focus on the background, while the foreground and middleground remain out of focus. **SEE 9.40**

**Rack focus** The rack focus effect involves changing the focus from one location on the z-axis to another. If, for example, you want to shift the emphasis from a smartphone to the person holding it without changing the shot (through a dolly, zoom, cut, or dissolve), you can first focus on the smartphone with the person out of focus and then “rack through” the focus range of the lens until the person’s face comes into focus, throwing the smartphone out of focus. **SEE 9.41 AND 9.42** Obviously, you need a relatively shallow depth of field to accomplish such a rack focus effect, which means you must use a narrow-angle lens.

If you had a great depth of field (wide-angle lens with a small lens aperture), you could just about rack through the entire focus range without noticeably affecting the focus. A rack focus effect is therefore not possible in this case. With
9.38 Selective Focus: Person in Front
A narrow-angle lens is used to create a shallow depth of field that allows selective focus. Note how the focus is on the person closest to the camera, with the people behind out of focus.

9.39 Selective Focus: Person in Middle
The middleground person is in focus, with the foreground and background persons out of focus.

9.40 Selective Focus: Person in Back
The person farthest from the camera is in focus, with the foreground and middleground persons out of focus.

9.41 Rack Focus Effect: Object Emphasized
In this shot the focus is on the smartphone. The shallow depth of field renders out of focus the person holding the smartphone.

9.42 Rack Focus Effect: Person Emphasized
Emphasis has shifted from one z-axis location (the smartphone) to another (the person). Because the depth of field is shallow, we can shift focus from the smartphone to the person by changing (racking through) the camera’s focus.
a narrow-angle lens, on the other hand, the depth of field becomes so shallow that even a slight racking of focus shifts the focal plane from one point along the z-axis to another. This means that a little adjustment of the focus control shifts the focus from one object to the other, even if they are only a short distance from each other along the z-axis. The table above summarizes how lenses influence our perception of depth.

You can also achieve a type of aerial perspective by using “fog filters” that render portions of the picture out of focus while keeping other portions sharp. Although the filter does not actually distinguish among different z-axis locations but rather among picture areas that are in and out of focus, we still perceptually interpret this as changes in the picture depth.

As useful as such aerial perspective techniques are in all types of two-dimensional photography, including film and video, they can cause big trouble when working with stereoscopic 3D. This problem is detailed in chapter 10.

### 3D BINOCULAR DEPTH

Binocular means that we are looking at the world with both eyes. Stop reading for a moment and look around you, first with one eye, then with both eyes. Do you see a big difference? Probably not. This is because several of the monocular depth cues also apply to binocular stereo vision, and we are normally unaware of the two additional depth cues provided by binocular vision—convergence and accommodation—or even the space between two objects on the z-axis.

**Convergence**

Because our eyes are normally 2½ inches (6.35 centimeters) apart, called interocular distance, each eye sees an object from a slightly different point of view. All you have to do is stretch out one of your arms with the thumb turned up, close one eye, and then have your thumb cover part of a more distant object, such as a tree or power pole. When you now switch to the other eye, your thumb jumps to the side. The closer you move your thumb toward your eyes, the larger the jump will be. This well-known experiment proves that the image of your thumb is projected on slightly different spots on the retinas of your left and right eyes. This binocular disparity—the difference between the right- and left-eye images—is also called parallax.

Whenever you fixate on an object, such as your thumb or a distant tree, the optical axes of your eyes—or, in media aesthetic terms, the index vectors of the eyes—cross. This is the point of convergence (POC) at which the images of both eyes fall on corresponding points of the retina, relieving your brain from dealing with binocular disparity, or parallax.
This point of convergence obviously changes with the distance at which the object of fixation is located. When the object of fixation is fairly close, our eyes turn in more than when we fixate on an object that lies farther away. The angle of how cross-eyed we become when the object of fixation lies closer or farther away tells our brain, in concert with all the other monocular cues, approximately how far away the object is.\(^5\) See 9.44

**Accommodation**

Much like we have to adjust the camera lens when focusing on a near or far object, the lenses of our eyes need adjustment to focus on objects at various distances. Fortunately, like the cameras, our eyes have a built-in auto-focus. When focusing on a far object, the lenses of our eyes get relatively thin; when focusing on a near object, they get relatively thick. As useful as this auto-focus feature is for navigating our actual environment, it can cause trouble when watching stereo 3D images on-screen.

**THE STEREO 3D ILLUSION**

Couldn’t we now reverse this process and fake various stages of binocular disparity by photographing an object from two slightly different viewpoints and showing one to the left eye and the other to the right eye? Would our brain fall for such a crude trick and interpret this artificial parallax of the two 2D pictures as depth? The answer to both questions is yes. In fact, this principle is the key to creating stereo 3D images, and it has not changed since the stereoscopes of the early 1830s (see figure 9.50).

In the following discussion of the technical and aesthetic details of stereo 3D, you should realize that even the most sophisticated hardware and software could not produce a 3D image; it is, finally, the brain that is tricked into the 3D illusion. As obvious as this sounds, you should always remain cognizant of the fact that a 3D illusion is not a natural visual process but an artificially constructed perception that is as far removed from reality as is a painting. Whereas you can actually point at traditional (single-lens) depth cues in a 2D photo or video (see figure 9.7), the 3D stereoscopic depth exists only in your mind. There is no way you can point at an image that floats between you and the projection screen. The stereoscopic effect is truly virtual.

**Stereo 3D Image Capture**

Unless you want to achieve the entire stereo 3D effect through digital manipulation in postproduction, you need two separate recording devices—one for the left eye and another for the right eye. The most basic way to capture stereo images is with two video cameras that are mounted side by side or with a twin-lens housing that directs the output of each lens into its own dedicated camera unit and recorder. See 9.45

Most high-end stereo video cameras have controls that allow two adjustments, unique to stereography: lens convergence and interocular distance.

**Manipulating lens convergence**

Much like our eyes, the “toe-in” angle of the two lenses in a 3D camera can be adjusted either physically or through an optical system to

**9.44 Point of Convergence**

When focusing on an object fairly close to us, our eyes turn inward to make their index vectors cross at the point of fixation. In 3D lingo, we call this the point of convergence.

**9.45 Stereo 3D Camcorder**

This 3D camcorder houses two lenses and two separate camcorders. Each lens sends its pictures to its dedicated camcorder.
effect various points of convergence. Unlike our eyes, however, such a convergence manipulation of lenses is not done to assist with focusing on close or far objects but rather to manipulate the apparent location of an event on the $z_h$-axis (screen-to-horizon) and the $z_v$-axis (screen-to-viewer). At this point you know enough about stereo 3D to adopt some of the stereographer’s lingo. For example, the traditional projection screen (computer, video, or film) has become the window. There is some justification to this semantic folly because in stereo 3D you can virtually see through the screen and have an event play on it, in back of it ($z_h$-axis), or in the space in front of it ($z_v$-axis).6

The POC marks the location of the window and, with it, what appears in back of it, on it, and in front of it. If you place an event in back of (beyond) the POC, it will appear behind the window (on the $z_h$-axis, or positive parallax). If you place the event at the POC, it will be perceived at the window (on the screen). If you place the event in front of the POC, it will play in front of the window between the screen and the viewer (on the $z_v$-axis, or negative parallax). For example, if you want to have person A appear in front of the window and person B behind it, you need to place the POC between the two people. **SEE 9.46** If you want both of them behind the window, you need to have the POC in front of them. **SEE 9.47**

What happens to the point of convergence when both lenses are parallel so that their index vectors do not cross? The POC lies in infinity, which means that all events play in front of the window along the $z_v$-axis. As exciting as it may seem to have person A float in front of you while talking to person B, who is behind the window, or having objects extend or shoot toward you through the window, experienced stereographers try keep most events behind the stereo window.

**Manipulating interocular distance, or stereo base** Contrary to our eyes, whose interocular distance is fixed at about 2½ inches (or 6.35 centimeters) and varies only slightly among the adult population, 3D cameras enable you to vary
the distance between the camera eyes—the two lenses. The distance at which you set the lenses is called the **stereo base**. The stereo base has a profound influence on how pronounced the stereo effect will be.

If it is set at the normal interocular distance of 2½ inches, persons A and B appear to be about as far apart as they actually are. **SEE 9.48** Widening the stereo base will stretch the $z_v$-axis and, with it, the distance between persons A and B on the $z$-axis. **SEE 9.49**

Experienced stereographers warn against setting the stereo base too wide, however. An exaggerated stereo base (with the lenses too far apart) can create a variety of negative perceptual consequences, among them the flattening of spherical objects, such as a ball, and the miniaturization of objects projected on the $z_v$-axis. Most importantly, however, the large parallax will excessively stress our visual system when trying to fuse the extreme right-eye and left-eye disparity into a single stereo image. Setting an optimal stereo base must also take into account the size of the screen on which the 3D images are projected and the distance from the viewer to the screen. **SEE 9.50**

**Stereo 3D Image Viewing**

To reproduce the all-important parallax for binocular seeing, the left eye must be kept from peeking at what the right eye is seeing and vice versa. This problem was solved somewhat with the invention of the stereoscope almost two centuries ago. This type of stereo viewer lets you look through one port that is dedicated to the left eye and other port that is dedicated to the right eye. The two ports are separated by a long wooden or metal panel that prevents you from crossing over from one visual channel to the other. Because such “modesty panels” would have to be unusually long when watching a 3D film or TV program, a series of other methods have been developed to keep each eye channel from being polluted by the other. The three most
Stereoscope

In the traditional stereoscope, the viewer directs the left eye to the left-eye picture of a scene, and the right eye to the right-eye picture. The center panel from viewer to screen prevents one eye from seeing the other's point of view.

Popular current separation methods for watching 3D images are the anaglyph, polarization, and shutter systems. All three require the use of stereo 3D glasses.

**Anaglyph System**  In the anaglyph process, the left-eye image of a scene is marked with a specific color (usually red), and the right-eye image is marked with the complementary color (cyan). Both images are slightly offset to simulate the different points of view of each eye. When looking at this double red/cyan image with stereoscopic glasses, the left red lens lets you see the red part of the double image with your left eye, but not with the right one, which is blocked by the cyan-filtered image. The right cyan lens lets you see the cyan-colored image with your right eye but blocks most of the red image.

You probably noticed that we are dealing here with basic subtractive color mixing: the red filter lets no light of the complementary color pass through and renders the entire cyan part of the stereoscopic image black. The complementary cyan filter does the same thing to the red light, rendering the red part of the double image almost invisible to the left eye. Whatever colors are used in the stereoscopic glasses, they must be complementary to effect subtractive color mixing and thereby prevent both images from simultaneously entering either eye. Careful color correction is required for all techniques so that what you see through your rose-colored glasses represents the normally colored world.

**Polarized Light System**  When using polarization, the light waves for the left-eye image are lined up horizontally and, for the right eye, vertically. The stereo 3D glasses have a horizontal filter on the left side and a vertical filter on the right, so they effectively block each other’s images and let your brain do the stereo mixing. The advantage of this system is that the colors of a scene are not affected by the red/cyan marking.

**Shutter, or active-glasses, System**  In the shutter, or active-glasses, system, the images for the left and right eyes are delivered alternately in rapid succession. To perceive a binocular stereo effect, you must wear glasses that are synchronized with
the alternate delivery. When the left-eye image is flashed on-screen, the glasses let you see only the left channel, and then they switch to let you see the right channel. This switching is fast enough to fake a simultaneous display of the left and right stereo channels.

Much like the polarized light system, the shutter system has a higher color fidelity than the anaglyph system. The drawbacks are that the glasses are significantly more complicated and expensive and that such faking simultaneity through fast switching puts an additional strain on our perceptual mechanism.

The various systems under development that do not require stereo 3D glasses are based on slightly offset fields in interlaced video scanning or on pixels that achieve the binocular disparity through a slight pixel shift. Some 3D cameras have small foldout monitors that allow glasses-free stereo viewing.

**Media Aesthetic Considerations**

You have probably noticed that the industry’s claim of binocular stereo 3D’s being close to how we actually see the world is simply wrong. When looking at where you are right now, objects do not extend toward you; they are simply nearer to you than others. To get a closer look at a nearby object, your visual system accommodates focus. When looking at something farther away from you, you do not judge the distance by the space between objects but by all the monocular cues discussed in this chapter. But when viewing a stereo 3D image, we become aware of space itself, or—more so—space is transformed into volume. This metamorphosis of space, as well as some perceptual and psychological liabilities of 3D viewing, are taken up in chapter 10.

**VIRTUAL REALITY**

A special way of simulating the three-dimensional world around us is called *virtual reality*, or VR for short. VR simulates your physical presence in a specific environment. A VR event cannot be experienced by an audience but rather only by the person who is wearing the VR mechanism. Although several people might simultaneously watch the same event as you are, you may be looking at one thing and each of the others at something else.

There are two main features that distinguish virtual reality from a 3D movie image. The first is that you are no longer looking at a window that marks the point of convergence, dividing the z-axis into the \( z_h \)- and \( z_v \)-axes; rather you *are* the screen from which the z-axis extends to the horizon. The second feature is that you can choose to some extent what in the VR environment you want to look at. Your virtual world no longer plays on a fixed screen in front of you but, like in real life, all around you.

Picture yourself standing at a busy intersection, waiting with a few other people for the traffic light to change. You look up at the dark clouds and wonder whether you should have brought an umbrella. Behind you a small girl asks her mother why they have to wait to cross the street. The mother points to the pedestrian signal and tells her daughter never to walk when a symbol or some writing is in red but that she can cross when it changes to white or green. Just then the signal changes and the cars to your left and right come to a stop. As you look down at the rather high curb, you see the little girl rush into the street. You hear and then see a car in the curb lane running the red light, barely missing the girl. The mother shouts, jumps into the street, and carries her terrified daughter back to the sidewalk while the other people are crossing the street. Before you cross, you turn around to take another look at the mother hugging her daughter, both with tears in their eyes. The mother waves at you, indicating that they are all right.
Of course you were not at this fictitious street corner, but you could have experienced this frightening scene by wearing VR goggles and an audio headset. Assuming that the VR camera was set up properly at this intersection, you can see all of these event details by turning your head or swiveling around—all while sitting in a comfortable chair. What would happen if you chose not to look back at the mother and daughter but rather to your right to get the license number of the speeding car? Assuming that the VR camera has recorded a 360-degree view, you could see the back of the speeding car, but you could not zoom in to get its license number. The VR display allows a 360-degree horizontal and a 180-degree vertical point of view but not a change of viewer-induced field of view, such as a zoom-in on the license plate of the fleeing car.

The following is a brief description of the major VR equipment and functions required to bring such a scenario to life.

**Head-Mounted Display**

Basically, the head-mounted display (HMD) consists of a small box that has a left-eye port with a lens and a right-eye port with a lens, through which you can view a stereo image displayed on a flat screen that is slipped inside the box. Sound familiar? Yes, it uses the same basic principle as the stereoscope in figure 9.50. The VR display looks much like oversized ski goggles and can be worn like them. Instead of the static dual-image of the stereoscope, you can now slip in your smartphone, whose highly sophisticated application delivers the VR images and sound.

Unlike the old stereoscope that displayed the same image regardless of whether you move your head, however, the VR unit has motion sensors built-in that detect your head and body movements and adjust your point of view accordingly. When you look up, you see what is above you, such as the overcast sky in our street-crossing scenario. When looking down you see the floor or, in our example, the curb. When looking left or right, or when turning around to look behind you, the HMD shows what you would actually see from your position. The problem in the early days of VR was that the head unit was quite heavy and tethered to a thick cable. Also, the images lagged considerably behind the body movements. The current HMD is extremely responsive, easy to wear, and totally independent of any backup equipment.¹⁰ **SEE 9.51**
VR Camera

If you can see all around you from a specific point in a scene, wouldn't you need a camera that can see all around? Yes, you would, and yes, such a camera exists. You could possibly use a single camera and shoot one view after another and then in postproduction stitch together the individual shots into a 360-degree panorama. But it is much simpler to use a camera that can record 360-degree horizontal and 180-degree vertical views from a single position.

The camera in figure 9.52, developed by the Jaunt company, has 16 wide-angle lenses arranged around a small drum, which fits easily into a briefcase; the camera's views are recorded on 16 Secure Digital (SD) cards located directly behind the lenses. Note that some lenses are mounted on the top and bottom of the drum for the above and below points of view. Because all the lenses are fixed, it is easier to stitch the individual views together into a seamless scene than if you were to use a single camera. Still, this stitching requires extensive postproduction so that the panorama looks continuous in shape and especially colors.  

9.52 Virtual-Reality Camera

This VR camera has 16 wide-angle lenses whose input is recorded on 16 SD memory cards. Note the lenses for above and below views.

VR Audio

Going back to the VR intersection scene, you would certainly expect to hear the traffic go by before the stoplights change; the voices of the mother and the girl behind you, talking about the pedestrian signal; the sounds of the speeding car in the curb lane running the red light; the mother's warning shouts from behind; and the crying of the girl and the mother as they pass you and the other people crossing the street. The problem is not the initial pickup of all of these sounds during the production of the scene, or even generating them in postproduction, but that they must be precisely synchronized with the event detail of the scene you choose to look at. For example, you should hear the mother talking to her daughter more clearly when you turn around to see them than if you were looking at the cars passing by. Various techniques have been employed for sound capture, but all systems are based on binaural sound, which requires headphones for the VR user.
As you know, through stereo audio we can quite readily define approximate locations of the sound source along the x-axis—the width of our visual field—but not the z-axis, which defines how sounds come from the front and back of us and how close or faraway the sound source appears. Additionally, the sound source might be above, with an airplane flying overhead, or below, such as the child talking to her mother. Whatever the sound pickup technique might be, VR sound must match what the VR visual images show, regardless of which way you are virtually looking. Such a complex synchronization of pictures and sound requires the precise matching of sound with its source in virtual space, regardless of whether it is at rest or moving. We discuss VR sound further in chapter 15.

**VR Aesthetics**

If virtual-reality images are based on the stereo principle, are they not pretty much the same as those of a 3D movie? Although there is some overlap, the answer is a definite no. The VR experience differs from 3D movies in technical and especially experiential aesthetic terms.

**Technical difference** The most obvious difference between experiencing virtual reality and a 3D movie is that your field of view in VR is spherical whereas in movies it is restricted to the screen. Note that the term *field of view* is used here not to mean the steps between close-ups and long shots but to describe the extent of your virtual environment. By turning your head sideways, you see what is to the left and right of you; by tilting your head, you see what is above and below; and by turning around, you see what is behind you. All of these movements are picked up by tiny gyroscopes that are built in to the video display and instantly transmitted to the VR display. Regardless of which part of the virtual environment you choose, it is always seamlessly stitched to the rest of the panorama. The other difference is that you are wearing your screen instead of watching it from a fixed position.

The audio, which is synchronized to the video, is delivered either by a large aggregate of speakers or, most often, to stereo headphones. The headphones are not only easier to use but, as we just pointed out, also deliver a more realistic soundfield in relation to the visual environment.

**Aesthetic phenomena** As in real life, you are the center of your virtual environment and what you see is always in relation to you. Up and down, left and right, above and below, and in front or behind always refer to your position. By always being the center of the surrounding event, you are literally immersed in it. This is why VR proponents speak of an immersive experience when engaging in a virtual-reality event.

A potential problem is that the computer responding to your physical cues (nodding, turning your head, spinning around) might not be fast enough to call up the corresponding views. If the visuals lag your body movements, or if you are standing or sitting still but see the virtual world spinning around you, your brain gets confused by the disparate signals between body position and what you see. This type of vertigo, technically called vection, can be so strong that some VR proponents recommend sitting in a swivel chair rather than standing up when entering the VR world. You will read about a similar perceptual reversal in the discussion of figure and ground in chapter 13.

The inherent problem with virtual reality is that it is, so far, indeed virtual because its environment is inevitably fixed. Regardless of how fast you spin in a chair or how slowly you turn your head to scrutinize your surroundings, they remain the same in subsequent viewings. Although you can control the part of an event you want to see, the event itself does not change. But once an event is
picked up and transmitted live to your goggles, you will be able to not only im-
merse yourself in the virtual event but also participate in it. All this depends, of
course, on where the VR cameras and microphones are placed and how the sig-
nals are processed and transmitted. If, at some future time, you are able to move
the cameras and microphones interactively from your viewing position, you will
come quite close to the real experience of being there. But until then, it might be
easier just to go there.

**SUMMARY**

In video, film, and computer displays, the three-dimensional world must be
projected onto the two-dimensional surface of the screen. Although the third
dimension (the z-axis) is illusionary, it proves to be aesthetically the most flexible
screen dimension.

Four concepts describe the aesthetics of the three-dimensional field: the
z-axis, monocular graphic depth factors, depth characteristics of lenses, and 3D
binocular depth.

Whereas the screen width (x-axis) and height (y-axis) have definite spatial
limits, screen depth (z-axis) is virtually limitless. The camera is therefore much less
restricted in its view and movement along the z-axis than it is either horizontally
or vertically. We perceive the z-axis as originating from the camera or screen and
extending toward the horizon.

The normal 2D z-axis, which extends from screen to horizon, is the \( z_h \)-axis.
In stereovision and hologram displays, the z-axis extends toward the viewer as
well. This is the \( z_v \)-axis.

Graphic depth factors include overlapping planes (objects that partially
overlap each other); relative size (an object positioned close to the camera projects
a larger screen image than one of similar size that lies farther down on the z-axis);
height in plane (assuming that the camera shoots parallel to the ground, we
perceive an object that is higher on the screen as farther away from the camera);
linear perspective (horizontal parallel lines converge toward a vanishing point
at the eye-level horizon line; equally spaced objects appear to lie closer together
the farther they are from the camera); and aerial perspective (the foreground
in focus with the background out of focus). More generally, aerial perspective
means selected focus on a spot along the z-axis in a shallow depth of field with
the surrounding areas out of focus.

The depth characteristics of lenses are significant in the manipulation of the
third dimension of a screen image. Wide-angle lenses exaggerate relative size and
linear perspective; they de-emphasize overlapping planes and aerial perspective.
Narrow-angle (telephoto) lenses exaggerate overlapping planes and aerial perspec-
tive, de-emphasizing relative size and linear perspective.

Selective focus and rack focus effects are powerful means of articulating the
z-axis and drawing attention to a specific plane along the z-axis. They are possible
only in a shallow depth of field and therefore with narrow-angle (telephoto) lenses.

A stereoscopic 3D effect differs from the standard 3D representation on a
screen in how we perceive the z-axis of picture space. In the standard 3D simula-
tion, the graphic depth factors define the z-axis from the screen to the horizon
(\( z_h \)-axis); in a stereoscopic projection, the articulated z-axis extends through the
screen space toward the viewer (\( z_v \)-axis). It is generated by a two-lens recording
system, whereby the lenses are spaced apart similar to our eyes; we get a left-eye
point of view and a right-eye point of view of the same scene.
Chapter 9

This shift can also be accomplished by a pixel shift with computer software. To show both slightly offset points of view on a single screen, one is colored red and the other is cyan. When the viewer wears stereo 3D glasses with a red lens for the left eye and a cyan lens for the right eye, each eye sees its dedicated point of view. The optical system in the viewer’s brain melds the two images into a single three-dimensional projection, which extends through the screen toward the viewer.

Virtual reality simulates your presence in a specific environment. It can be experienced only by wearing a headset, called a head-mounted display (HMD), that provides a certain 3D visual and sound environment. Depending on where you turn your head or body, you will see a specific part of the visual panorama surrounding you. The HMD goggles consist of two lenses that enlarge stereo images displayed by a smartphone.

The VR camera has a cluster of lenses arranged for 360-degree horizontal and 180-degree vertical views, which can be recorded on corresponding SD cards. The sound is picked up by a variety of omnidirectional microphones positioned for 3D binaural sound.

Because you are wearing rather than looking at the screen and are always at the center of the entire panorama surrounding you, the VR experience is genuinely immersive.

NOTES

1. The x-, y-, and z-axes are used here in the traditional sense of the Cartesian model that quantifies Euclidian space.

2. Although we call both phenomena “aerial perspective,” there is a difference between aerial perspective and detail perspective. Detail perspective refers to the gradual diminishing of detail in the distance, whereas aerial perspective refers to the more bluish tint the farther away the object is from the observer.

Leonardo da Vinci vividly described what happens in aerial perspective:

There is another kind of perspective which I call aerial, because by the difference in the atmosphere one is able to distinguish the various distances of different buildings when their bases appear to end on a single line, for this would be the appearance presented by a group of buildings on the far side of a wall, all of which as seen above the top of the wall look to be the same size; and if in painting you wish to make one seem farther away than another you must make the atmosphere somewhat heavy. You know that in an atmosphere of uniform density the most distant things seen through it, such as the mountains, in consequence of the great quantity of atmosphere which is between your eye and them, will appear blue, almost of the same colour as the atmosphere when the sun is in the east. Therefore you should make the building which is nearest above the wall of its natural colour, and that which is more distant make less defined and bluer; and one which you wish should seem as far away again make of double the depth of blue, and one you desire should seem five times as far away make five times as blue. And as a consequence of this rule it will come about that the buildings which above a given line appear to be of the same size will be plainly distinguished as to which are the more distant and which larger than the others.


4. Actually, the narrow-angle lens simply enlarges the end of the z-axis, where even with a wide-angle lens objects look crowded and space is squeezed. This crowding effect is entirely in line with increased density of texture at the far end of the z-axis.


6. Window can mean the computer, television, and movie screens but also the allowable parameters for proper stereo capture and viewing. See Bernard Mendiburu, *3D Moviemaking: Stereoscopic Digital Cinema from Script to Screen* (Burlington, MA: Focal Press, 2009), pp. 79–84.

7. Even though interocular distance refers specifically to the distance between our eyes, the term is also used to describe the stereo base—the distance between 3D lenses. Some stereographers object to this misnomer and call it “interaxial distance.” Because of its popularity, we use the term interocular distance to mean how far apart our eyes are but also for the stereo base.


9. The first commercial use of stereoscopic 3D is credited to Sir Charles Wheatstone, who introduced binocular vision and his stereoscope to the Royal Scottish Society of Arts in 1853. During the second half of the nineteenth century, stereoscopy rapidly gained in popularity. Producing stereo photos and viewing them with the stereoscope became a huge commercial success and a popular pastime. For more information about stereoscopy photography, see Ray Zone, *Stereoscopic Cinema and the Origins of 3-D Film 1838–1952* (Lexington: University Press of Kentucky, 2008); and his *3-D Filmmakers: Conversations with Creators of Stereoscopic Motion Pictures* (Lanham, MD: Scarecrow Press, 2005), part of the Scarecrow Filmmakers series.

10. To have Palmer Luckey, the inventor of the new VR head-mounted display, explain how it works, visit http://www.smithsonianmag.com/videos/category/innovation/smithsonian-ingenuity-awards-2014-palmer-lu/?no-ist.
In this chapter you will encounter terms that are quite familiar to musicians and people who deal with music theory but that may be new to you. You may also be somewhat reluctant to make sense of the many examples that use musical notation, much like our tendency to skip examples that contain mathematical formulas. But even if you do not read music, you can understand the examples by simply connecting the notes so that they form vectors. When translated into graphic, index, and motion vectors, musical notation can describe a line, a direction, a movement, and horizontal and vertical structures. Take a moment to look ahead to the figures indicated in the following paragraph.

If the horizontal vectors have an uphill slant, the tune is rising; if they go downhill, the tune gets lower (see figure 16.25). If the vectors go up and down, the melody similarly goes up and down. If the vertical vectors show many notes stacked on top of each other, the chords are dense; if they consist of only two or three notes, the chords are less dense (see figures 16.18 and 16.19). A single note can be subdivided into several shorter notes, which are played faster (see figure 16.3). The basic beat is indicated by the numbers preceding the first note, such as 4/4 or 3/4. The 4/4 has four beats to each unit, called a bar; the 3/4 has three beats to each bar (see figure 16.20).

In this chapter you need to pay particular attention to elements of sound, basic sound structures, picture/sound combinations, and picture/sound matching criteria. When talking about sound, you will have to consider not only its physical characteristics, such as the frequency or loudness of a sound, but also its psychological characteristics, such as how high or low or how loud or soft we perceive a tone to be. As with color, however, neat physical formulas for sound use do not always translate into equally neat perceptions. Our definitions are therefore often based on perceptual rather than strict physical phenomena and processes. Such perceptions are truly contextual.

Now that you have some idea of how to deal with musical notation, I suggest that you listen to the examples by going to the MindTap for *Sight Sound Motion* or by asking your instructor to download the music resources and make them available to you. Try to listen to the music while following the notation vectors.

That’s all there is to it, at least so far as this chapter is concerned. Even if you do not intend to deal this closely with sound structures, you will nevertheless learn patterns that transfer quite readily to the organization of visual elements.
ELEMENTS OF SOUND

When you strike a single piano key, blow into a trumpet, or draw the bow across a violin string, you can hear five distinct attributes, or elements, of sound: pitch, timbre, duration, loudness (dynamics), and attack/decay.

Pitch

Pitch indicates the relative highness and lowness of a sound measured against an agreed-upon scale. The pitch of a tone is perceived and measured by its frequency—its vibrations per second. A high-pitched tone vibrates with a higher frequency than a low-pitched tone. The generally accepted pitch standard is the A above middle C, called A₄ because it is four octaves above A₀, the lowest note of a standard piano. A₄ vibrates 440 times per second, which is expressed as 440 Hz (hertz, the international unit of frequency). SEE 16.1

Timbre

Timbre (rhymes with "amber") describes the tone quality or tone color. The timbre of a tone tells you whether it is produced by a trumpet or a violin. Technically, the timbre of a sound is created by the type and relative number of overtones. Overtones are a number of frequencies with which a sound-producing source vibrates in addition to its fundamental frequency—the one that we hear as a specific pitch.

All overtones have a higher pitch than the basic tone, although you do not hear the individual overtones as higher pitches in combination with the basic tone. Rather, you perceive the overtones more like a superimposition, as a richer, more complex tone than one with fewer or no overtones. The figure on the facing page illustrates the formation of overtones. SEE 16.2

The point to remember is that a rich, full sound has many overtones; a thin or hollow sound has few overtones. Some instruments, such as the flute, produce sounds with fewer overtones than the violin or cello. The human voice has more overtones than most instruments. When computers were first used to synthesize the human voice, it was the lack of sufficient overtones that gave the voice its robotic monotone “computer-speak” quality.

16.1 Pitch

You can recognize the pitch of a tone by its relative position on the staff. Note that a higher octave of a tone is exactly twice its frequency; the lower octave is half its frequency.
16.2 Timbre

a When you pluck a violin string, it moves back and forth between fixed points (the upper end of the fingerboard and the bridge) in a wavelike pattern. The crest of the wave moves in one direction, and the valley comes back from the opposite direction. Practically, the whole string appears to be moving up and down. The number of such up-and-down movements of the string per second determines its basic frequency (hertz) and pitch. This fundamental frequency is the tone’s first harmonic. Because the vibration of the string as drawn in this illustration has no other vibrations superimposed on it, it has no overtones.

b The violin string also vibrates simultaneously in separate sections. In this illustration each half of the string vibrates twice as fast as the total string, which is an exact octave of the basic frequency. This octave is the second harmonic and represents the first overtone.

c The more sections of the string that vibrate in multiples of the basic frequency, the higher and more numerous the harmonics become. Although you may not be able to hear each overtone separately, you perceive the sum of these overtones as a richer sound.

The frequency of the fundamental tone—and those frequencies that are simple multiples of the fundamental tone, such as double, three times, or six times the frequency of the fundamental—are called harmonics. Thus the fundamental tone (the one that determines the pitch) is the first harmonic. The second harmonic may be a frequency twice as high as that of the fundamental. This second harmonic represents the first overtone. Because it has a frequency twice that of the fundamental tone, it is an octave higher. The second and all subsequent harmonics have a progressively higher frequency than the fundamental tone. Let us assume that the fundamental tone (first harmonic) vibrates at a frequency of 440 Hz. You would hear an A4—the normal A you hear when an orchestra tunes up. If the first overtone (second harmonic) vibrates at 880 Hz, which is twice the frequency of the fundamental, you would hear a tone that is an octave higher than the fundamental. If the second overtone (third harmonic) vibrates at three times the frequency of the fundamental, you would have a 1,320 Hz tone (which is a very high E added to the overtones).
16.3 Duration
In musical notation the duration of a tone is precisely determined by a specific symbol, such as whole notes, half notes, quarter notes, eighth notes, and so forth. Just how long a whole note is supposed to be when measured by a stopwatch depends entirely on the basic tempo of the piece or section.

**Duration**

Duration refers to how long you hear a sound lasting. **SEE 16.3** You may perceive very short bursts of a tone or a long, continuous one. Most musical instruments permit some control of sound duration, like the bow that one pulls across the strings or the piano pedals that one depresses and releases. In musical notation there are specific symbols that determine the relative duration of various sounds. Even then different musicians and conductors like to decide on their own how long a note should be.

**Loudness (Dynamics)**

The loudness of a tone is its apparent strength as we perceive it. You can play a tone of a certain pitch and timbre either loudly or softly. Essentially, when you have to hold your ears while listening to music, the sounds are loud; when you have to strain to hear them, they are soft. Loud sounds have a high magnitude; soft sounds have a low one. The variations of perceived strength are the dynamics of the sound. **SEE 16.4**

\[
\begin{align*}
\text{f} &= \text{forte (loud)} \\
\text{ff} &= \text{fortissimo (very loud)} \\
\text{fff} &= \text{forte fortissimo (extremely loud)} \\
\text{p} &= \text{piano (soft)} \\
\text{pp} &= \text{pianissimo (very soft)} \\
\text{ppp} &= \text{piano pianissimo (extremely soft)}
\end{align*}
\]

16.4 Loudness (Dynamics)

In musical notation the loudness of a tone is indicated by symbols. The electronic volume control on an audio console ultimately affords, of course, more loudness variations.
**Attack/Decay**

The attack or decay of a sound is part of its dynamics and duration. *Attack* refers to how fast a sound reaches a certain loudness level. After the attack the sound dips a little in volume and then maintains the desired volume level. This steady maximum loudness of a tone is called sustain level. From the point the tone starts to get softer until we can no longer perceive it is called its *decay*. The whole process, from initial attack to final decay, is called the sound envelope or, simply, *envelope*. The envelope represents the total sound duration and not just its sustain level. **SEE 16.5**

To make things even more complicated, there is some confusion among audio experts about just what “decay” means. For some, especially the people who are engaged in synthesized sound, decay refers to the little dip after the attack just before the volume reaches its sustain level and not the fading of the sound volume. What is normally meant by decay, they call “release.” For our purposes we will settle on the more common use of *decay*—the fading of the volume until the sound is no longer perceivable.

Both attack and decay can be fast or slow. In vector terminology a fast attack means that the sound vector achieves its maximum magnitude quickly. **SEE 16.6** In a slow attack, it takes the sound vector some time to get up to its maximum strength (loudness). **SEE 16.7**

**16.5 Envelope**

The basic sound envelope consists of attack, initial decay, sustain level, and final release. The time it takes for the release is the final decay of the sound. The length of the release is the decay time.

**16.6 Fast Attack**

In a fast attack, the sound gets loud quickly; the rise time is short. It is a high-magnitude sound vector.

**16.7 Slow Attack**

In a slow attack, the sound takes a while to reach the desired loudness; the rise time is slow. It is a low-magnitude sound vector.
Chapter 16

The sustain level is how long the vector maintains a relatively steady magnitude. When its magnitude starts to decrease, the sound has entered its decay phase. A fast decay means that the sound vector drops from its high magnitude to zero relatively quickly. **SEE 16.8** When it takes some time for the vector to lose its strength, the sound has a slow decay. **SEE 16.9**

In musical language crescendo stands for a relatively gradual change from soft to loud, and diminuendo is a gradual change from loud to soft. They differ from attack and decay mainly in terms of speed. Even a slow attack is much faster than a crescendo.

Some instruments have by their very nature a fast or slow decay. A drum has a relatively fast decay (thump), and a church bell has a relatively slow one (ring). Much like timbre, the stages of the sound envelope, such as the attack phase, have some influence on how we perceive a specific tone. A fast attack gives the tone a more aggressive edge than does a slow attack.

Not all instruments give you the same control over the sound envelope (duration and dynamics from attack to final decay). As a violin player, for example, you would have maximum control over the attack time, the sustain level, and the decay time. Depending on how you use the bow, you can produce a fast or slow attack, a short or long sustain level, and a fast or slow decay. When playing the piano, however, you have control over the dynamics of the tone (its loudness or softness) but not over its attack and sustain level. The only control you have is over the decay period. When you press the sustain pedal, the sound takes more time to fade; the pedal stretches the decay. The dampening pedal shortens the decay; the tone dies more quickly. When playing the drums, you are at the mercy of the instrument for the entire envelope. You can strike it only softly or hard, to produce a soft or loud bang, but you have no control over attack, sustain level, or decay. The cymbals, on the other hand, give you an almost instant attack but a very slow decay, unless you stop their vibrations with your hand.

The decay of sound is also influenced by the acoustics of the room. In a fairly "live" acoustical environment, the decay is slower (more reverberations) than in a rather "dead" environment (fewer reverberations). When you are synthesizing sounds and sound combinations, the control of all the major elements of the sound envelope obviously plays an important role.

When you combine picture and sound vectors, you can match the attack and decay variables of both vector types. For example, a visual motion vector can reflect a variety of attack and decay modes, depending on how fast it reaches or loses a specific maximum magnitude. If an object, such as a dragster, accelerates to a specific speed quickly, you have a motion vector with a fast attack; if it accelerates more gradually, like a big truck with a heavy load, you have a slow-attack...
motion vector. **See 16.10** If the object motion decelerates quickly, its magnitude drops equally quickly, and you have a fast-decay motion vector. If it decelerates more gradually, it produces a slow-decay motion vector. **See 16.11** The advantage of translating sound into vectors is explored more fully later in this chapter.

**BASIC SOUND STRUCTURES**

This discussion of sound structures is not meant to be exhaustive but to aid you in dealing effectively with the sound portion of a video or film production. It will help you especially in developing, or deciding on, a sound track that not only supports the video portion but also combines with it synergistically to become an organic, maximally effective whole. To achieve this you should try to translate musical elements and structures into various types of sound vectors whenever possible. As mentioned at the beginning of this chapter, seeing music in terms of sound vectors means you can, for example, deal successfully with the basic notation of a musical piece even if you do not read music. More importantly, vectors allow you to compare sound with pictures and find relationships between sound and pictures that would otherwise not have been so evident. Vectors convey more readily than any other aesthetic factor the contextual nature of media aesthetics.

Let us now take a closer look at the following basic sound structures: melody, harmony, homophony, and polyphony.

**Melody**

A *melody* is a series of musical tones arranged in succession. Like a sequential series of shots, a melody is a sequential series of specific sounds that has a logic in...
Chapter 16

16.12 Melody

Melody moves as a horizontal sound vector in a linear fashion. Each tone leads to another until they become an entity—a tune. Its progression and that forms a tune. Melody can be represented by a horizontal vector. SEE 16.12

Like life itself, a melody is constantly progressing and is complete only when it has ended. You experience the various steps of progression, but you can only remember the total melody—the tune. You should note that the logic of melodic progression differs among various cultures. For example, the traditional melodies of Middle Eastern or Far Eastern songs are much more subtle in their progression than are Western melodies. And yet the Eastern cultures readily, if not eagerly, embrace Western music and play it extremely well.

All melodies are based on specific scales. In Western music there are major and minor scales, chromatic and diatonic scales, and whole-tone and twelve-tone scales. They differ basically in the number of steps and the interval between the steps within an octave (when the tone repeats itself higher or lower on the scale). SEE 16.13–16.17

16.13 Chromatic Scale

In the chromatic scale, an octave is divided into 12 equal steps, with the twelfth tone being the octave (double the frequency of the fundamental). It consists entirely of half-tone intervals.

16.14 Major Diatonic Scale

In the diatonic scale, an octave is divided into eight steps, with the eighth tone being the octave (double the frequency of the tonic—the first tone). The steps are not equal, with half steps between the third and fourth tones and the seventh and eighth tones. The diatonic scale is what we (in Western music) normally use in melodic and harmonic structures. Generally, a major scale expresses a positive, normal, practical mood.
16.15 Minor Diatonic Scale

In the minor diatonic scale, the eight steps are not equal. There are half steps between the second and third tones and the fifth and sixth tones. In the ascending minor scale (illustrated here), there is a one-and-a-half-step interval between the sixth and seventh tones and a half-step interval between the seventh and eighth (octave) tones. The minor scale reflects a sad, mysterious, haunting, less definite mood.

16.16 Whole-Tone Scale

The whole-tone scale uses full steps between all seven intervals.

16.17 Twelve-Tone Scale

Some composers simply regard the 12 tones of the chromatic scale as equal and arrange these notes in some form regardless of the specific intervals. Because there is a total of 12 notes in a scale before the notes begin to repeat themselves as an octave, the individual row can contain 12 notes. The structuring of such sets of 12 notes has become known as the twelve-tone scale.

Harmony

Whereas a melody consists of horizontally successive notes, **harmony** is a vertical combination of simultaneously played notes. Melody is linear and sequential; it forms a horizontal vector. Harmony is simultaneous and forms a vertical vector. **SEE 16.18**

Contrary to the melody, which you can perceive only incompletely in its development, you can hear harmony in its totality all at once. Melody leads somewhere; harmony is already there. It has always arrived. The harmonic combination of two notes is normally called an interval, and three or more notes compose a **chord**.

The magnitude of the vertical sound vector depends on several factors, principally the relative density, or **sound**
Chapter 16

16.19 Harmonic Density
Chords can vary considerably in their complexity—their relative chord density, or sound texture. To play this chord on the piano, you would have to press all the white and black keys between the highest and lowest notes of the chord.

Source: “Pentatonic” by P. Peter Sacco © 2014 Cengage Learning

In traditional music, harmonic structures must adhere to strict rules. The simplest harmonic structure consists of two tones that are played simultaneously. Chords consist of three or more simultaneous tones. Two simultaneous tones equal an interval.

The basic unit of the traditional harmonic structure is the triad, a combination of three tones.

Major triad

Minor triad

Vertical harmonic structures are usually built from the same scale in which the melody operates (that is, the chords have the tonality of the melody), but the melody and the chords may also operate in different keys or outside of any predetermined scale (as in twelve-tone structures, for example).

C major

G minor

Texture, of the chord (the number of notes in the chord and how close together they are), its perceived tension (consonant, pleasant sounding, or dissonant, consisting of tones that do not blend together well), and its tension relative to the melody. Generally, the higher the chord density and tension, the higher the vector magnitude. See 16.19

Homophony

Homophony means literally “alike sounding” (Greek homo means “the same”; phonos means “sound”). In music homophony refers to the structure in which a single predominating melody is supported by corresponding chords. The chords act like pillars (vertical vectors) that hold up the melody bridge (horizontal vector). See 16.20 In a homophonic structure, the horizontal vector of the melodic line is independent. You can re-create its logic by simply whistling any tune. See 16.21 But the accompanying chords are dependent on the melody; they cannot stand alone—and playing them without the melody would make little sense. The chords by themselves do not lead to satisfactory musical closure. See 16.22

Polyphony

In musical terminology polyphony refers to two or more melodic lines that, when played together, form a harmonic whole. Unlike homophonic structures, where a single dominating melody is accompanied by supporting chords, polyphonic structures are composed of multiple, coequal voices (melodies, or horizontal vectors). No single voice is relegated to a subordinate role throughout the piece; each runs its own course, sometimes dominating the other voices and sometimes temporarily assuming a supporting role. When played separately, each voice forms a self-sufficient entity. See 16.23 When played together the various horizontally independent voices form a vertical, harmonic structure. Vertical vectors are formed incidentally through planned juxtaposition and interaction of the horizontal vectors of the various voices. The voices must make sense not only horizontally (melodic development) but also vertically (harmonic development). See 16.24

Counterpoint Most polyphonic music is written in counterpoint—a specific polyphonic technique in which the individual notes and melodic lines are set against each other. It emphasizes an encounter among the various voices, a vector-against-vector affair. We use counterpoint to achieve a certain structural tension—a high-energy field. By contrasting in a calculated way the various horizontal vectors
16.20 Homophony
In a typical homophonic structure, the leading melody is supported by a parallel chord accompaniment. When we translate into vectors the written music of this waltz in G major by Franz Schubert, we can see how the horizontal vector of the melody is supported by the vertical vectors of the chords.

16.21 Melody
In a homophonic structure, the melodic horizontal vectors are independent. The melody can exist by itself.

16.22 Chords
In a homophonic structure, the accompanying chords are dependent on the melody. They are dependent vertical vectors; that is, they need the melody to form a gestalt.

16.23 Polyphony
In a polyphonic structure, each voice (horizontal vector) is basically independent; that is, each voice has its own logical melodic development.
of the independent voices, we also create vertical vectors of varying complexity and magnitude.

Although counterpoint in music is limited to specific melodic and harmonic juxtapositions, for our purposes we can broaden the concept to include techniques for creating tension that involve other aesthetic elements, such as the direction of the voices (melodies) and their pitch, timbre, dynamics, and rhythm. Some of these techniques are similar to those employed in the visual aesthetic fields.

Contrapuntal tension is achieved primarily by having the direction of the voices—the melodic lines—go against each other. In the language of media aesthetics, the voices act as converging and diverging vectors. **SEE 16.25**

---

**16.24 Horizontal and Vertical Structure**

Although in polyphonic music the individual voices are basically independent, they nevertheless connect vertically to form a tight harmonic structure.

---

**16.25 Tension through Direction**

To achieve the desired contrapuntal tension, we can contrast the direction of the voices—the horizontal vectors. While one melody line (voice) is going down, the other one is going up and vice versa. We have, in effect, converging and diverging vectors (and some continuing ones that run parallel to each other).

From *Inventio IV* by J. S. Bach
You can also contrast the pitch of the various voices. While one voice operates in a relatively high range (treble), the other voice develops in a lower range (bass).  **SEE 16.26**

You can achieve a contrast between voices by using timbre. One voice may be played by the violin, the other by the flute. **SEE 16.27**

Contrasting dynamics create various tensions even in a single melody (playing parts of it loudly and others softly), but they are especially effective when they occur between voices. For example, part of the upper voice is loud while the lower voice is soft; then the lower voice is played louder than the upper one. This use of contrapuntal dynamics draws attention to different parts (usually the theme) very much as a close-up does. If you think that this is like a figure/ground


16.27 Tension through Timbre
We can create contrapuntal tension by giving each voice a distinct timbre. This piece relies primarily on timbre for contrast. One voice is played by the flute, the other by the violin.

arrangement, you are on the right track. Through dynamics one voice can switch between being the figure for a few bars and then serving as background for the other voice. **SEE 16.28**

A rhythmic contrast is similar in effect to a dynamic one. While one voice is progressing quickly, the contrapuntal voice is slow—or the other way around. You can give one voice a sharp staccato beat and set it against another that proceeds in a more continuous legato line. **SEE 16.29**

16.28 Tension through Dynamics
We can create contrapuntal tension by contrasting the dynamics of the various voices. Here one voice (the theme) is rather loud; the other (the counterpoint) is much softer and less obtrusive. In effect there is a juxtaposition of sound vectors of varying magnitudes.

From *March in D Major* by J. S. Bach
16.29 Tension through Rhythm

One of the favorite techniques of creating tension in a polyphonic structure is to juxtapose different rhythms among the various voices. While one voice is progressing quickly, the contrapuntal voice is slow; while one has a sharp staccato beat, the other proceeds as a continuous legato line.

From The Well-Tempered Clavier by J. S. Bach

Contrapuntal structures  Widely used contrapuntal structures are imitation, the canon, and the fugue. All of these musical forms use the polyphonic principle in which each voice develops independently of the others yet plays against the others in highly calculated ways.

One of the most common elements of contrapuntal structure is imitation, wherein a short theme or subject is stated in one voice and then repeated verbatim or in a slightly changed form in the other voice or voices while the first voice continues on its way, providing the counterpoint to the imitated theme. See 16.30

Most contrapuntal structures use some form of imitation.

The canon, or round, is the purest and most obvious form of imitation. Not only is the theme repeated verbatim by the other voices but also the entire melody. The harmonic (vertical) structure is created by a phasic shift of the identical melodies; that is, each one starts a little later than the voice immediately above. See 16.31 Familiar canons such as “Three Blind Mice” and “Row, Row, Row Your Boat” work on the total-imitation principle.

Simple repetitions of the same subject without varying counterpoints can be boring, however, which is why the masters of polyphony, such as Johann Sebastian Bach, vary not only the counterpoint from voice to voice but also occasionally the theme itself.3
16.30 Imitation
One of the key elements of a polyphonic structure is imitation: a theme or short phrase is repeated verbatim one by one in the other voices. In this example the theme is stated in voice 2 and then imitated in a different pitch in voice 1.

From The Well-Tempered Clavier by J. S. Bach

16.31 Canon
In a canon (or round), the complete tune is repeated by phasing. One voice starts the tune, and the others carry the same tune at staggered intervals. Note how the same subject is stated in one voice and then repeated verbatim in the others. The harmonic (vertical) structure is achieved through a phasic shift of the identical melodies.
One of the most intricate contrapuntal structures is the fugue (which comes from the Latin fugere, “to run away, to flee”; and fuga, which means “flight”). In a fugue a theme or subject is chased and flees from voice to voice throughout the composition. The theme is imitated and expanded in each of the voices, relating vertically at each point to form a complex yet unified whole.

The theme in a fugue is normally introduced all by itself in one specific voice (such as the middle voice), then in another (the top voice), while the middle one proceeds on its own, providing the necessary counterpoint. The theme finally appears in the third voice (the bottom one), with the other two voices continuing on their ways, providing the necessary counterpoint for each other and for the third voice. **SEE 16.32**
When the theme has been introduced once by all the voices, we have the first exposition. The theme is then imitated, varied, and expanded throughout the voices in a free-flowing way. This part, where the composer is showing off, is called the episode. When the theme is clearly introduced again, we have another exposition; and when it is imitated again throughout the other voices, we have another episode. Normally, several expositions and episodes occur in a single fugue.

In the complex fabric of a contrapuntal vector field, the vertical vectors (the harmonic chords as formulated when we read the independent melodic voices vertically at any given point) act as important structural agents. They hold the horizontal voice lines together and give the seemingly independent voices their necessary structural dependence. Essentially, the vertical vectors represent space/time modulators, explicating the spatial (harmonic) as well as the temporal (melodic) relationships and interdependence of the individual voices. The vertical vectors tell us where the individual voices have been, where they are going, and how they fit together.

**STRUCTURED IMPROVISATIONS AND NOTATION OF INDETERMINATE MUSIC**

The phrase *structured improvisations* seems like an oxymoron because how can you structure something that is supposed to be improvised—delivered without previous preparation? Artists have struggled for centuries with the problem of exercising freedom of expression and breaking the rules while maintaining the degree of control that all art requires. In music any conventional score is like a fully scripted play: the theme, development, and final outcome are carefully prescribed before the piece is ever played. The only variable at your disposal is how you play the piece, not what you play. In this respect both the score of a musical piece and the fully scripted play are equally deterministic; they have no open future. Even in improvised jazz, the rhythmic conventions and the timing of solos leave very little wiggle room for the musicians. This is where improvisational structure comes in.

The key to this improvisational structure is a new notation, which provides some visual suggestions of how a piece might be played, but it prescribes neither the instruments nor any of the conventional melodic and harmonic sound.
combinations. There are instances, however, when the musicians are invited to initially interpret the pictographic score on their own, but they must perform it subsequently in exactly the same way, much in the spirit of a piece with standard notation. Although composer Mark Applebaum calls this “notation of indeterminate music,” the performance has become very determinate.4

PICTURE/SOUND COMBINATIONS

Even if you are successful in structuring the picture field and the sound field independently, you cannot expect to arrive at a meaningful audiovisual structure simply by adding the two together. You must combine the video and audio vector fields so that they form a synergistic structure. Such a video/audio combination requires that you hear the screen event while visualizing and sequencing it and see it while working with the sound. You should try to conceive and develop the video and audio vector fields together as much as possible.

But exactly how should you combine the pictures and sound so that they form such a synergistic unit, a maximally effective picture/sound gestalt? No easy recipe exists, and each case has its own specific requirements. Nevertheless, homophonic structures and polyphonic structures can act as general guidelines. Additional guidance is derived from the picture/sound matching criteria discussed in the following section.

Homophonic Structures

Much like the homophonic structure in music, where a single dominant melody is supported and undergirded by accompanying chords, you can support the video portion step-by-step with appropriate sound or the audio track with appropriate pictures. To accompany the lone cowboy riding through the meadow with the familiar clickety-clack banjo tunes is an example of a well-worn homophonic video/audio structure. A slightly more elaborate example of homophonic video/audio structure is the scene in which a car loses its brakes and is careening down a steep mountain road. The accompanying audio track consists of literal sounds such as squealing tires and nonliteral nervous music. This picture/sound combination represents a typical homophonic structure. The visuals (car without brakes) dominate the scene and tell the principal story. Literal and nonliteral sounds are precisely in step with the visual event, properly intensifying it from moment to moment. SEE 16.34

Many music videos illustrate their sound tracks with accompanying pictures in a homophonic fashion. The video shows the singer, band members, dancers, or pictures that support or illustrate the lyrics of the song.
Chapter 16

Polyphonic Structures

In a polyphonic picture/sound structure, pictures and sound seem to develop independently as “melodic” lines yet combine vertically into an intensified audiovisual experience. For example, some music videos show pictures that do not in any way parallel the lyrics of the song. The pictures seem to tell their own story (often of the lead singer’s psychological frustrations) and are relatively independent of the meaning of the lyrics. While the song proclaims tender love, the pictures may show the singer’s first unsuccessful auditions. The vertical structure is achieved through strong parallel rhythms of the pictures (tertiary motion rhythm) and sound.

Four of the more notable polyphonic audiovisual techniques are phasing, transitions, multiple texts, and multiple screens.

Phasing

Phasing is a technique in which the video and audio portions are not tightly synchronized—they are somewhat out of phase. Either the picture precedes the sound event or vice versa, or picture and sound are thematically out of phase, at least for a while. In phasing, the sound is asynchronous to the picture.

Flashbacks are a good vehicle for the phasing technique. Imagine, for example, a mountain climber surprised by a snowstorm on an especially difficult part of the climb. We see him trying to reach a ledge where he can find temporary shelter. We hear the howling storm, the sounds of his crampons, and his labored breathing (homophonic literal sound). Suddenly, these natural sounds switch to a conversation the climber had with his friend before undertaking the difficult climb. The friend warns about the fickle weather and the dangers of avalanches. Then the sounds switch just as suddenly to the laughter of the climber’s children. Finally, we switch back to the literal sounds of the climb. While the video tells the story progressively (the climber’s efforts to stay alive), the audio is out of phase: it shifts occasionally to the past (friend’s warning and children’s laughter). Yes, the climber made it back alive, but he had to abandon his solo attempt to summit.

Phasing lends itself especially well to space/time transitions. Here is an example:

<table>
<thead>
<tr>
<th>Video</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tight 2-shot: Larry and Barbara in front of the library.</td>
<td>Larry</td>
</tr>
<tr>
<td>ECU of Barbara.</td>
<td>Don’t you want to go to a movie tonight?</td>
</tr>
<tr>
<td>Zoom back to reveal entrance to the movie theater.</td>
<td>You can’t study all the time.</td>
</tr>
<tr>
<td>Larry hands her the ticket.</td>
<td>Barbara (smiling)</td>
</tr>
<tr>
<td></td>
<td>Perhaps...</td>
</tr>
</tbody>
</table>

Now the video has jumped ahead of the audio to the “effect” phase while the accompanying sound (Barbara’s reply) still lingers in the “cause” phase. If you now switch the phasing and have the audio progress to the effect phase with the video still lingering in the cause phase, you are dealing with predictive sound (see chapter 15).

Transitions

You can use a type of phasing for transitions from one scene to the next. For example, you can show a couple driving to a rock concert; the woman turns on the car radio and begins to groove with one of the concert band’s catchy tunes. Cut to: MS of lead singer and bass guitarist of the actual concert scene with the catchy tune now being continued by the band. Cut to: CU of woman, now in the audience, moving to the tune the way she did in the car.
16.35 Phasing

In phasing, the audio track of one scene (as event 1 for example) extends into the next scene (event 2) or changes to the next scene (event 2) while the video portion still shows the previous scene (event 1).

Or we may see a young man sitting in a bus, watching the rhythmic play of alternating sunlight and shadows caused by the passing trees. A drumbeat parallels the syncopated rhythm of the sunlight flashes. It increases in volume and texture and carries through the next series of cuts: (1) the young man crossing a busy intersection almost synchronous to the same syncopated beat, (2) rushing to the beat along the crowded sidewalk, (3) entering the backstage door of a concert hall, (4) sitting at the piano, looking at the audition judges, and (5) playing his extemporaneous piece—you guessed it—in the syncopated rhythm of the sunlight flashes of the bus ride.

Of course, you can make the sound track as complex as you want (by adding traffic sounds), but the primary transitional device remains the syncopated beat. In such polyphonic structures, it is the audio that sustains the transitions for the various video scenes.

Multiple texts The text of a story is what a character says and does. A subtext is what the character wants to say or do but keeps hidden; it can also refer to the character’s psychological makeup. Multiple texts are two or more dialogue tracks that run simultaneously or in a phasing mode. You can use multiple texts for single-screen presentations or, more commonly, for multiscreen productions. The following is an example of multiple texts playing simultaneously with a single-screen scene.

<table>
<thead>
<tr>
<th>Video</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-shot of couple.</td>
<td>Track 1 (main text)</td>
</tr>
<tr>
<td></td>
<td>She: How did it go today?</td>
</tr>
<tr>
<td></td>
<td>Track 2 (slightly delayed subtext)</td>
</tr>
<tr>
<td></td>
<td>She: I don't really care.</td>
</tr>
<tr>
<td>CU of man.</td>
<td>Track 1</td>
</tr>
<tr>
<td></td>
<td>He: Oh, pretty well!</td>
</tr>
<tr>
<td></td>
<td>Track 2 (slightly delayed)</td>
</tr>
<tr>
<td></td>
<td>He: What do you care, anyway?</td>
</tr>
<tr>
<td></td>
<td>Track 3 (simultaneous with 2)</td>
</tr>
<tr>
<td></td>
<td>She: Liar!</td>
</tr>
</tbody>
</table>
As you can see from the multiple dialogue, the two people do not feel as civilized toward each other as they pretend to be in the main dialogue track (track 1). Such a multiple-text technique is an obvious communication device and should be treated with discretion and deftness. Depending on how much of the event complexity you want to communicate, you can emphasize track 1, 2, or 3 or you can play them all together at equal volume. If you play all three tracks at the same time and the same loudness, the audience will no longer be able to follow exactly what is being said. Instead of communicating specific information, you will be providing the audience with a fabric of speech sounds that may reflect the complexity of the moment better than any single track could.

You can also emphasize a specific track by running its volume relatively high while keeping the other tracks at a lower volume. In effect you are applying the figure/ground principle. By emphasizing track 1, you communicate primarily the “outer” event—a woman asking a man about his day. But if you emphasize track 2 or 3, you shift from an outer-event orientation (plot) to an inner one (feeling); you add to the horizontal vector (event progression) a vertical one (event complexity). The advantage of such multiple texts is that you can provide the event complexity through sound while keeping the visual event relatively simple and straightforward.

**Multiple screens.** You can achieve a truly polyphonic structure by using multiple screens. In such a presentation technique, each screen can pursue a different story (voice) that relates to the others thematically or through the interaction of characters. Each screen may have its own sound track, or you can use an identical sound track for all three (or more) screens. Sometimes the various sound tracks get muddled into a dialogue fabric, but they allow the audience to associate with one or another phrase, sentence, or utterance.

If you construct the dialogue tracks of the various screens so that they are sequential rather than simultaneous, the audience will be better able to follow the meaning of each voice. For example, if you have a three-screen setup, the person on the left screen may ask, “How do you feel?” and the person on the right screen may answer “Fine,” while the person on the middle screen may simply listen. Thus you can establish a relationship among the three screen events even if the scenes are not connected by event location and event time. Because the dialogue in the three screens is connected through the narrative, we tend to connect the visual events on the separate screens however different they may be.

As with the single-screen presentation technique, there is a limit to how many dialogue tracks we can discriminate among, even in multiscreen presentations. When played simultaneously, the tracks might lose their informational function and yet blend into a sound configuration that serves as a powerful emotional intensification device. You may want to arrange the video and audio tracks of multiscreen presentations in a fuguelike way in which a particular theme “flees” from screen to screen, with the other screens and especially the other audio tracks providing contrapuntal contexts.

**PICTURE/SOUND MATCHING CRITERIA**

Ideally, you should conceive pictures and sound together as a unit, trying to see and hear the screen event simultaneously as an aesthetic whole. In practice, however, such a complete preconception of the whole event is not always possible. Even if the audio portion of your show consists mainly of literal, source-connected sounds, you will most likely need to add some music to intensify the overall screen event.

You probably find that you think in pictures first and then try to locate the appropriate supporting sounds, such as background music. Or you may have a
piece of music, such as a popular song, for which you need to find appropriate visuals. But what specific types of music should you use? What if your unfail-
ing instinct fails you just when you need it most? What you obviously need is more reliable criteria for selecting the appropriate music for the more common video events. Although they are hardly foolproof and may even seem offensive to your artistic integrity, you may find some guidance in applying one of the four basic picture/sound matching criteria: historical/geographical, thematic, tonal, and structural.

Historical/Geographical

Historical matching means that you pair pictures with music that was created in approximately the same historical period. For example, you would match the picture of a Baroque church with the music of Johann Sebastian Bach; an eighteenth-century scene in Salzburg, Austria, with music by Wolfgang Amadeus Mozart; a scene of 1960s London with music by the Beatles; or the brand-new, supermodern museum building with music by a contemporary composer.

Geographical matching means that you select music that is typical of the geographical area depicted in the scene. You could, for example, match a scene that plays in Vienna with a waltz, a scene in Japan with traditional koto music, or one in the American South with Dixieland jazz.

Once again a word of warning: matching the sights and the sounds of historical periods does not automatically make for effective and smooth picture/sound combinations. For example, the precise, carefully structured music of Bach does not necessarily fit the light, flamboyant Baroque architecture. Also, geographical matching is such an obvious aesthetic device that it might annoy or even insult the sensitive viewer. For example, to introduce an interview show featuring a famous scholar from China with what we consider to be typical Chinese music, or the ambassador from Austria with a waltz, would certainly offend the guests and annoy the viewers. On the other hand, if you do a documentary on Tibet, you might as well have Tibetan music and the sounds of prayer bells under the opening scenes and titles.

Thematic

When using thematic matching for video and audio, you select sounds that we are accustomed to hearing at specific events or locales. For example, when we see the interior of a church, we probably expect organ music. Or when you show a parade or a football stadium full of people, marching-band music is the thematically correct choice.

Tonal

Tonal matching requires sounds that fit the general tone, that is, the mood and feeling of the event. If you show a sad scene, the music should be sad, too. Similarly, you can match a happy scene with upbeat sounds. Romantic music that engulfs the lovers’ tender moonlit embrace is another familiar example of tonal video/audio matching. The difference between thematic and tonal matching is that in thematic matching you choose music you might hear at a specific event or location; in tonal matching you select music according to how the event feels.

Structural

When using structural matching, you parallel pictures and sound according to their internal structure. For example, take a look at the following figures and see whether you can “hear” them. **SEE 16.36–16.40**
16.36 Structural Matching 1

How does this picture sound? Loud or soft? Fast or slow? Does it have a simple or a syncopated beat? Is it more polyphonic than homophonic?
Although this picture shows a pattern similar to the one in figure 16.36, it contains other visual elements that may cause you to “hear” a different sound track.
16.38 Structural Matching 3

Compare the sound of this picture to that of figure 16.37. Do you hear a difference? Try to verbalize it.

What audio characteristics can you isolate that are unique to this picture?
16.39 Structural Matching 4
What specific sound characteristics does this picture have? Does it sound different from figures 16.36 and 16.37? What specific sound vector field would you have to create to match the visual vector field of this image?
How did you do with “listening” to these pictures? Could you assign each picture a specific type of music? Did they sound different from one another?

Now go back and try to identify some of the specific pictorial characteristics that prompted you to select a certain type of music. Was it the direction and arrangement of dominant vectors within the pictures? The relative harshness or softness of the lines? The graphic weight, degrees of balance, or relative shadow falloff? Does the picture feel heavy or light? Regular or irregular? Simple or complex? Which ones were relatively fast or slow? Did the pictures suggest a specific rhythm? Were you now able to “hear” Applebaum’s graphic score?

You probably went more by how the pictures felt to you and whether they were round, soft, hard, brassy, simple, or complex rather than by a careful vector analysis. Most likely, that was all you needed to accurately match the video and the sound structures. Nevertheless, even with your intuitive approach, you applied—however quickly and subconsciously—a series of vector analyses and comparisons. What you were actually doing was analyzing the visual vector fields and constructing aural vector fields of similar characteristics.

**STRUCTURAL ANALYSIS**

Just in case your intuition fails you, or the matching task becomes too complex for quick guessing, you need to apply specific criteria to analyze the video and audio vector fields for compatibility. To isolate the dominant structural element within a picture or picture sequence and to grasp the overall structural tendency (simple, complex, directional, confused) is not always an easy task. Sometimes
a picture or picture sequence just does not seem to “sound” right, or it seems so indistinct that any number of musical structures could fit it equally well or badly. In such a situation, you must be patient enough to analyze the picture sequence step-by-step according to the fundamental and contextual image elements developed throughout this book: light and color, space, time/motion, and sound. You should then be able to establish primary connections between the major visual vector fields or dominant aesthetic elements and similar aural ones.

Exactly how does such a structural analysis work? And how can you translate pictorial characteristics into music? There is no single or simple answer. Structural matching, like any other matching criterion, depends on many contextual variables, such as the desired overall effect, what immediately preceded the picture sequence and what follows it, and whether you intend to create montage effects. Most likely, you will discover a specific rhythm (fast/slow, simple/complex) in the video that you can then match with the rhythm of your music or audio effects.

The following table shows how you might approach a structural analysis of the video and its translation into musical terminology for audio matching. In fact, this table is an apt summary of the major elements of the five aesthetic fields and how they might relate to one another.  

To witness the structural power of music, take any video sequence and run some arbitrarily selected music with it. So long as the tempos of the primary and tertiary motion of the video and music are similar, you will be amazed at how frequently the video and audio seem to match structurally. Apparently, we use the tempo—the basic beat—and the relative complexity of the musical piece as the primary reference and expect the video to follow suit. If the visual beats do not match the aural ones, we apply psychological closure and try to make them fit until they drift too far apart. Again, the basic beat—the rhythm and tempo of the music—seems to be the primary agent that determines whether the video portion is a match or a mismatch. Once the tempos are similar, you can look for further matching criteria as indicated in figure 16.41.

See 16.41
## 16.41 Video/Audio Structural Analysis

<table>
<thead>
<tr>
<th><strong>Aesthetic Field</strong></th>
<th><strong>Video</strong></th>
<th><strong>Audio</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Light</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Directional</td>
<td>Staccato</td>
</tr>
<tr>
<td></td>
<td>Nondirectional</td>
<td>Legato</td>
</tr>
<tr>
<td>Mode</td>
<td>High-key</td>
<td>Major</td>
</tr>
<tr>
<td></td>
<td>Low-key</td>
<td>Minor</td>
</tr>
<tr>
<td>Falloff</td>
<td>Fast</td>
<td>High-contrast (loud/soft)</td>
</tr>
<tr>
<td></td>
<td>Slow</td>
<td>Low-contrast (even)</td>
</tr>
<tr>
<td><strong>Color</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>High</td>
<td>Loud</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Soft</td>
</tr>
<tr>
<td>Hue</td>
<td>Warm</td>
<td>Major</td>
</tr>
<tr>
<td></td>
<td>Cool</td>
<td>Minor</td>
</tr>
<tr>
<td>Saturation</td>
<td>High</td>
<td>Brass, strings</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Flutes, reeds</td>
</tr>
<tr>
<td>Brightness</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Space</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screen size</td>
<td>Large</td>
<td>Loud (high-energy)</td>
</tr>
<tr>
<td>Shot size</td>
<td>Large</td>
<td>Soft (low-energy)</td>
</tr>
<tr>
<td></td>
<td>Small</td>
<td>Near</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>Far</td>
</tr>
<tr>
<td>Graphic weight</td>
<td>Heavy (close-ups)</td>
<td>Complex (accented)</td>
</tr>
<tr>
<td></td>
<td>Light (long shots)</td>
<td>Simple (unaccented)</td>
</tr>
<tr>
<td>General shape</td>
<td>Regular</td>
<td>Consonant</td>
</tr>
<tr>
<td></td>
<td>Irregular</td>
<td>Dissonant</td>
</tr>
<tr>
<td>Balance of objects within frame</td>
<td>Dynamic (high-tension)</td>
<td>Chords and beat</td>
</tr>
<tr>
<td></td>
<td>Static (low-tension)</td>
<td>Complex (timbre, chords)</td>
</tr>
<tr>
<td>Texture</td>
<td>Heavy</td>
<td>Consonant</td>
</tr>
<tr>
<td></td>
<td>Light</td>
<td>Dissonant</td>
</tr>
<tr>
<td>Field density (number of elements in single frame)</td>
<td>High</td>
<td>Chord tension</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High (dissonant)</td>
</tr>
<tr>
<td>Field density (number of successive elements within given period)</td>
<td>High</td>
<td>Low (consonant)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Chords</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simple</td>
</tr>
<tr>
<td>Harmonic density</td>
<td>High</td>
<td>Chords</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Complex</td>
</tr>
<tr>
<td>Melodic density</td>
<td>High</td>
<td>Simple</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simple</td>
</tr>
</tbody>
</table>
### 16.41 Video/Audio Structural Analysis (continued)

<table>
<thead>
<tr>
<th>Aesthetic Field</th>
<th>Video</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space</strong> (continued)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Aesthetic energy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vector magnitude</strong></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Vector field energy</strong> (total energy communicated)</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Dynamics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Loud</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Soft</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sound vector field energy</strong></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Melodic or contrapuntal density</strong></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Melodic line</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Definite</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vague</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Melodic progression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Definite</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vague</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sound vector orientation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Harmonic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Complex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Simple</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Melodic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vague</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Field complexity in single frame or shot</strong></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Field complexity in successive frames or shots</strong></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Graphic vectors</strong></td>
<td>High-magnitude</td>
<td>Low-magnitude</td>
</tr>
<tr>
<td><strong>Index vectors</strong></td>
<td>High-magnitude</td>
<td>Low-magnitude</td>
</tr>
<tr>
<td><strong>Principal vector orientation</strong></td>
<td>Horizontal</td>
<td>High-magnitude</td>
</tr>
<tr>
<td></td>
<td>Vertical</td>
<td>High-magnitude</td>
</tr>
<tr>
<td><strong>Change in field of view (zooms)</strong></td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td><strong>Vector continuity</strong></td>
<td>Good</td>
<td>Bad</td>
</tr>
<tr>
<td><strong>Transitions (cuts, dissolves)</strong></td>
<td>Seamless</td>
<td>Conspicuous</td>
</tr>
<tr>
<td><strong>Rhythm</strong></td>
<td>Complex</td>
<td>Simple</td>
</tr>
<tr>
<td><strong>Tertiary motion</strong></td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td><strong>Volume and tempo</strong></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Sound rhythm</strong></td>
<td>Even</td>
<td>Uneven</td>
</tr>
<tr>
<td><strong>Dynamics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fast crescendo and diminuendo (fast attack and decay)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slow crescendo and diminuendo (slow attack and decay)</td>
<td></td>
</tr>
<tr>
<td><strong>Melodic progression and rhythmic continuity</strong></td>
<td>Even</td>
<td>Uneven</td>
</tr>
<tr>
<td><strong>Modulation (change from one key to another)</strong></td>
<td>Extreme</td>
<td>Conservative</td>
</tr>
<tr>
<td><strong>Sound rhythm</strong></td>
<td>Complex</td>
<td>Simple</td>
</tr>
<tr>
<td><strong>Beat</strong></td>
<td>Fast</td>
<td>Slow</td>
</tr>
</tbody>
</table>
When watching a video production or a film without sound, you will become surprisingly aware of the visual structure, especially the tertiary motion. The rhythmic beat of tertiary motion, or lack of it, shows up with undue prominence without the benefit of the accompanying sound track. When watching the same sequence with sound, however, you may not even be conscious of the tertiary motion employed in the video sequence. Again, the organic structural power of music greatly facilitates the pictorial vector flow. Exactly why will remain as much a mystery as the power of music itself.

**Summary**

In structuring the five-dimensional field, we concern ourselves with the elements of sound, basic sound structures, picture/sound combinations, and picture/sound matching criteria.

The elements of sound, or basic sound attributes, include pitch, timbre, duration, loudness (dynamics), and attack/decay.

Pitch refers to the relative highness and lowness of a sound measured against an agreed-upon scale; the pitch of a tone is perceived and measured by its frequency. Timbre describes the tone quality or tone color, whether a certain note is played by the violin or the trumpet; timbre depends on the number of overtones that vibrate with the fundamental tone. Duration refers to how long a sound can be heard. The loudness (dynamics) of a tone is its apparent strength as we perceive it; you can play a tone either loudly or softly. The variations of loudness are the dynamics of the sound. The attack/decay is part of the dynamics and duration of a sound. Attack refers to how fast a sound reaches a certain loudness level. Decay is the final fading process. The period during which a tone remains at its maximum loudness is called the sustain level. A sound envelope includes the whole tone from initial attack to final release.

The basic sound structures include melody, harmony, homophony, and polyphony. Melody is a series of musical tones arranged in a consequent succession. A melody forms a horizontal vector. Harmony is the combination of simultaneously played notes, which form a vertical vector. Harmonic combinations of three or more tones are called chords.

Homophony refers to a musical structure in which a single, predominant melody is supported by corresponding chords. In a homophonic structure, the melodic line (horizontal vector) can stand alone, but the chords (vertical vectors) make sense only in the presence of the melody.

Polyphony stands for “many voices” and refers to two or more melodic lines (horizontal vectors) that, when played together, form a harmonic whole. When played separately each voice forms a self-sufficient entity. Together they form a vertical harmonic structure. Most polyphonic music is written in counterpoint, which means that the elements of various voices (vector direction, pitch, timbre, dynamics, and rhythm) are countering one another to produce structural tension.

Imitation, a common structural element in polyphony, means that a short theme of one voice is repeated in the other voice or voices. The canon (round) is a direct imitation. The fugue introduces a theme that appears in different voices.

The picture/sound combinations should form a synergistic structure in which picture and sound reinforce each other. In homophonic combinations, the picture dominates and is supported by sound or vice versa. In polyphonic combinations, picture and sound seem to develop independently as melodic lines yet they combine vertically into an intensified audiovisual experience. Multiple texts refer to several sound tracks of dialogue that are run simultaneously or in a phasing mode.
There are four basic picture/sound matching criteria: historical/geographical, thematic, tonal, and structural. In historical matching you select music that was played in the historical period of the visual event. Geographical matching means that the music chosen has its origin in the geographical area depicted. In thematic matching you select sounds that you expect to hear at a particular event or locale. When using tonal matching, you choose sounds that fit the general mood and feeling of the pictorial event. Structural matching means that we parallel pictures and sound according to the internal structure, that is, their specific vector fields. The major attributes for structural matching seem to be similar rhythms and tempos of sound and picture.

Detailed structural analysis involves analyzing the picture sequence according to the fundamental and contextual image elements of light and color, space, time/motion, and sound. You then establish primary connections between the major visual vector fields or the dominant aesthetic elements and similar aural ones.

NOTES

1. The background music you sometimes hear in elevators, waiting rooms, and supermarkets has a drastically limited range of dynamics. The music sounds equally soft throughout and is therefore unobtrusive. All other elements (melody, timbre, harmony, and rhythm) are kept intact.


3. Listen to a relatively simple polyphonic composition (such as J. S. Bach’s Two-Part Inventions) and follow the theme throughout the piece. Note the counterpoint that works opposite the theme or the slight variations of the theme itself when it is reintroduced in each of the two voices. If you are a skilled listener, follow the themes in Bach’s The Well-Tempered Clavier and in his fugues.

4. “The score, while unconventional, does not invite improvisation; instead, players are charged with the task of predetermining a thoughtful means to sonify (with exactitude) an inferred musicality.” From the foreword to Medium for quartet by composer Mark Applebaum.
Glossary

above-eye-level key-light position The principal light source (key light) strikes from above the object’s eye level.

accelerated motion Object motion on-screen appears faster than normal. In film and digital video, the division of object motion into relatively few “at” positions, each differing considerably from the other. The frame density is low.

achromatic Its basic meaning is “without chroma” (color). Usually, it refers to totally desaturated colors (having no hue) that show up white, black, and various shades of gray (brightness steps). The grayscale is an achromatic scale ranging from white to black.

additive color mixing The mixing of colored light. Usually, the mixing of the light primaries—red, green, and blue (RGB).

aerial perspective One of the graphic depth factors: three-dimensional emphasis by means of selective focus in a shallow depth of field. Only a relatively short section of the z-axis is in focus, with everything else out of focus. See also graphic depth factors.

aesthetic energy The energy we perceive from aesthetic phenomena, such as color, sound, and motion. Can be expressed as vector magnitudes.

aesthetic motion paradox An object can be in motion and perceived at rest, or at rest and perceived in motion.

ambient sounds Background sounds that normally occur in a specific environment.

anaglyph A 3D method in which the right-eye and left-eye views are colored in red and cyan (or any other complementary colors) and superimposed. When you look at the double red/cyan image with stereoscopic glasses, the red filter for the left eye lets you see the red-colored image but blocks the cyan one; the cyan filter for the right eye lets you see the cyan-colored image but blocks the red one.

analytical montage The selection of key elements of a single event (analysis) and their proper sequencing (synthesis).

angles Variety of camera viewpoints. Angle can also refer to a specific approach to a story.

applied aesthetics The branch of aesthetics that deals with sense perceptions and how to influence them through the fundamental image elements of light and color, space, time/motion, and sound.

applied media aesthetics Same as applied aesthetics except that its focus is on video, film, and other electronic audiovisual media.

articulating the z-axis To place objects or people along the z-axis to serve as three-dimensional space modulators.

aspect ratio The relationship of screen width to screen height: 4 × 3 (1.33:1) for the standard video screen; 16 × 9 (1.78:1) for HDTV; and between 1.85:1 (5.55 × 3) and 2.35:1 (7 × 3) for wide motion picture screens. The display screens of mobile media range from the standard 4 × 3 aspect ratio to various vertical ratios.
asymmetry of the frame  The right and left sides of the video, film, and computer screen are unequal in visual prominence. The right side commands more attention than the left.

“at-at” theory  Adapted from Zeno’s concept of motion, consisting of a series of static “at” positions in space and time, each differing to some degree from the previous one. Film shows a specific “at” position in each frame.

attached shadow  A shadow that is on the object itself. It cannot be seen independent of (detached from) the object. Attached shadows help us primarily with interpreting an object’s basic shape and texture.

attack  The speed with which a tone reaches a certain (usually maximum) level of loudness. See also decay.

background  A basic structural element of the three-dimensional field—the depth plane farthest from the camera, marking the end of the z-axis.

background light  Illumination of the set pieces and backdrops. Also called set light.

back light  Illumination from behind the subject and opposite the camera.

balance  Relative structural stability of objects or events within the screen. Specifically, the distribution of vectors and graphic weight into static (stable and unlikely to change) and dynamic (asymmetrical and less stable) pictorial structures.

below-eye-level key-light position  The principal light source (key light) strikes from below the subject’s eye level. Also called horror lighting and reverse modeling.

binaural sound  Sound recorded by two microphones that are attached to an artificial head approximately where our ears are and played back through separate headphone channels. It emphasizes the z-axis (depth) dimension.

biological time  An internal clock that tells us when to feel awake or tired. A type of subjective time that is measured quantitatively (when to do certain things).

brightness  The color attribute that indicates how light or dark a color appears in a black-and-white photograph. Technically, brightness is one of the three major attributes of color that indicates how much light is reflected from a colored surface. Also called lightness and value.

brightness constancy  The stabilization of brightness values by our mental operating system so that we perceive white as white and black as black regardless of the actual reflectance values. Also called lightness constancy.

cameo lighting  Lighted objects set off against a plain, dark background. Foreground figures are illuminated with highly directional light, and the background remains unlighted.

canon  The purest and most obvious form of musical imitation. Both the theme and the melody are repeated verbatim by the other voices. Also called round. See also imitation.

cast shadow  A shadow produced by an object and thrown (cast) onto a surface (part of the object itself or another surface). The cast shadow may be object-connected (shadow touches the object producing it) or object-disconnected (shadow does not touch the object producing it). Cast shadows help us locate an object relative to its surroundings.

character time  The objective- and subjective-time elements concerning the character’s actions and feelings.

chiaroscuro lighting  Lighting for light/dark contrast (fast falloff) to emphasize volume and specific areas.

clock time  The “at” position in the time continuum when an event occurs. See also objective time.

collision montage  An idea-associative montage that clashes opposite events to express or reinforce a basic idea.

color  Specific wavelengths within the visible light spectrum, which we interpret as various hues.

color attributes  The three color sensations: hue, saturation, and brightness.

color constancy  Perceiving a color as uniform despite variations.

color energy  The relative aesthetic impact a color has on us; the relative energy a color emits within its contextual field.

color harmony  Hues that go well together. Specifically, the balanced energy of colors. Color harmony is most easily achieved with high-energy colors (figure) set off against a low-energy color background (ground).
**color model**  A graphic representation of the integration of the three color attributes: hue, saturation, and brightness.

**color temperature**  The relative bluishness or reddishness of white light, measured in Kelvin degrees. Bluish light has a high color temperature; reddish light has a low one. The video camera must be adjusted to the color temperature of the prevailing light. See also white balance.

**comparison montage**  An idea-associative montage that compares seemingly dissociated yet thematically related events to establish or reinforce a basic idea.

**complexity editing**  The building of an intensified screen event from carefully selected event essences. Montages result from complexity editing.

**context**  The environment in which we perceive and evaluate specific perceptual phenomena. Every aesthetic element operates within, and is dependent on, the context of all others.

**contextualism**  A branch of philosophy that includes, rather than excludes, the environment (context) in the process of clarifying, intensifying, and interpreting experience.

**contextualistic aesthetics**  How the various fundamental image elements (light and color, space, time/motion, and sound) operate in various contexts and in relation to one another. See also contextualism.

**continuing vectors**  Vectors that succeed each other in the same direction.

**continuity editing**  The assembly of shots that ensure vector and vector field continuity. Its principal function is the clarification of an event.

**convergence**  See point of convergence

**converging vectors**  Vectors that point or move toward each other.

**counterpoint**  A specific polyphonic technique in which the various voices (horizontal vectors) encounter each other. In media aesthetics the musical counterpoint of note against note is expanded into vector against vector.

**cross shooting**  Similar to over-the-shoulder shooting except that the camera-near person is out of the shot. See also over-the-shoulder (O/S) shooting.

**cut**  The instantaneous change from one shot (image) to another.

**cutaway**  A shot of an object or event that is peripherally connected with the overall event and that is (ideally) neutral as to screen direction. Used to intercut between two shots in which the screen direction is reversed.

**decay**  The speed with which a sound fades to where it can no longer be heard. See also attack.

**deductive visual approach**  Moving from an overview to event detail. A deductive method stresses the analysis and breakdown of a complete video program, film, or computer display into its major aesthetic elements.

**depth of field**  Area along the z-axis that is in focus. In a great depth of field, most or all objects located along the z-axis at different distances from the camera are in focus. In a shallow depth of field, only objects that lie within a short section of the z-axis are in focus; all others are out of focus. It is dependent on the focal length of the lens, lens aperture, and distance from camera to object.

**desaturation theory**  The more desaturated the colors of a scene, the more internal it becomes and the more the viewer is compelled to participate. Color desaturation renders the scene low-definition.

**dialectic principle**  The juxtaposition of opposing or contradictory statements or events to resolve the contradictions into universally true axioms or an event synthesis (new event or idea).

**diegetic sound**  Literal sounds that “occupy story space,” that is, are part of the story. See also literal sound.

**digital cinema**  Refers mainly to the use of high-definition digital video cameras instead of traditional film cameras for the acquisition of images. Can also refer to the whole electronic production, postproduction, distribution, and projection processes. It differs from HDTV in that it can work with higher-resolution images, such as 4K (4,000) or more pixels per horizontal line.

**digital video effects (DVE)**  Visual effects created by a computer or other digital video effects equipment.

**dissolve**  A gradual transition from shot to shot in which the two images temporarily overlap.

**diverging vectors**  Vectors that point or move in opposite directions.

**dramatic agent**  Any object or action within a scene that contributes directly to its intensification.

**dramaturgy**  The art of dramatic narrative and composition. More generally, the whole structure of a play—the total orchestration of dialogue, action, and various aesthetic elements.
**duration**  
The running time of a scene, sequence, or total film or video production. In music, refers to how long we perceive a sound.

**dynamic balance**  
An asymmetrical balance where the graphic weight and the vectors are not equal on both sides of the screen. The aesthetic energy is increased because the asymmetrical distribution of graphic elements and vectors causes some tension.

**dynamics**  
The variations of perceived loudness of a sound.

**editing**  
Selecting significant event details and sequencing them into a comprehensive whole—building a screen event.

**editing syntax**  
Prescribes how to arrange individual shots so that they combine into meaningful scenes and sequences to tell a story. See also *continuity editing* and *complexity editing*.

**encoding**  
Translating an idea into a message for a specific communication medium.

**envelope**  
The total duration of a tone, from initial attack to final decay. Also called *sound envelope*.

**event density**  
A great number of event details that occur within a specific clock time period. Can also be used to describe the complexity of an event.

**event intensity**  
The relative energy and relative significance we perceive about a specific event.

**experience intensity**  
The number of relevant experiences to which we are subjected simultaneously or in rapid succession and their relative depths.

**external vector**  
Force with a direction and a magnitude operating outside of us.

**eye level**  
The plane parallel to the ground, emanating from the eye of the observer. Eye level and the horizon line lie on the same plane regardless of how high the observer is from the ground.

**fade**  
The gradual appearance of a picture from black or its disappearance to black.

**falloff**  
(1) The brightness contrast between the light and shadow sides of an object. (2) The speed (degree of change) with which the brightest part of an object turns into dense shadow. *Fast falloff* means that the lighted area changes abruptly into dense shadows; the contrast is high. *Slow falloff* means that a very gradual change takes place from lighted area to shadow area or that very low, if any, contrast exists between light and shadow areas. *No falloff* means that there is no contrast—all visible sides are equally illuminated.

**field of view**  
The territory a shot includes, ranging from extreme long shot (ELS) to extreme close-up (ECU).

**figure/ground principle**  
Our tendency to organize a scene into figures that lie in front of a background. In so doing we perceive the ground as being more stable than the figures. In sound design *figure/ground* means that you choose the important sounds to be the figure while relegating the other sounds to the background.

**fill light**  
Additional light on the opposite side of the camera from the key light, used to illuminate shadow areas and thereby reduce falloff. Usually accomplished by floodlights.

**first-order space**  
Video space as defined by the borders of the video screen (x- and y-axes) and the illusory z-axis. See also *second-order space*.

**flat lighting**  
Omnidirectional illumination from no particular single source. The falloff is slow or nonexistent.

**forced perspective**  
An exaggerated linear perspective, making us perceive parallel lines converging more drastically than in normal vision. Wide-angle lenses create a forced perspective.

**foreground**  
A basic structural element of the three-dimensional field—the depth plane closest to the camera, marking the beginning of the z-axis.

**frame density**  
The sampling rate of a motion, that is, the number of “at” positions used to divide a single motion.

**fugue**  
A musical theme or subject stated first in each of the voices (usually four) and then restated verbatim or in a slightly changed form at various times in all voices. The theme is virtually chased from voice to voice throughout the fugue, relating vertically to the counterpoints of the other voices.

**geographical matching**  
Sound and pictures originate in the same geographical area.

**gestalt**  
A complete configuration that we perceive through psychological closure. The perceived pattern is generally different from and often more complete than the sum of its parts. In a gestalt all elements operate in relation to the whole.

**golden section**  
A classical proportional system in which the smaller section of a line is to the greater section as the greater is to the total length of the line. Especially effective when a prominent horizontal line is divided by a vertical one at the golden section point. It creates a dynamic balance. Also called *divine proportion* and *golden mean*.
**white balance**  The adjustment of the three color signals in the video camera to show a white object as white regardless of the relative color temperature of the light that illuminates the object.

**wide-screen format**  The most common aspect ratio used for wide-screen film—1.85:1. This means that for every unit of screen height there are 1.85 units of screen width. The standard video aspect ratio of 1.33:1 is considerably narrower.

**window**  The screen or any other 2D surface on which the 3D image is displayed.

**windowboxing**  Technique for changing the aspect ratio by placing a smaller picture in the center of the actual display screen, with the leftover space of the 16 x 9 frame surrounding it.

**window violation**  An image that is cropped by the edges of the frame and then projected into zv space as an incomplete image.

**wipe**  The transition in which a second image, framed in some geometric shape, gradually replaces all or part of the first one.

**z-axis**  The axis in the coordinating system that defines depth. Also the imaginary line that extends from the camera lens to the horizon.

**z-axis blocking**  Arranging the event (people and things) along the z-axis or in close proximity to it.

**z-axis index vector**  Someone looking or pointing directly at the camera.

**z-axis motion vector**  Movement along the z-axis (toward or away from the camera).

**z-axis vector**  An index or motion vector that points or moves toward or away from the camera.

**z-h-axis**  Stands for z-horizon axis. The line (space) from window (screen) to horizon. Also called positive parallax.

**zv-axis**  Stands for z-viewer axis. The line (space) from window (screen) to viewer. Also called negative parallax.

**zero time**  A high-magnitude subjective-time vector that occupies only a spot (practically zero length) in the objective time continuum.
3D. See stereo 3D
4 × 3 aspect ratio (standard TV), 88–91, 105
5.1 surround sound, 319, 320, 329, 330
7.1 surround sound, 319, 320, 329
8½, 32
12-tone scale, 342, 343
16 × 9 aspect ratio (HDTV), 88, 90–93, 94, 96, 97, 105, 106, 144, 234

ABC positioning, 375
above-eye-level lighting, 31
above-eye-level point of view, 231
abstract expressionism, 248
abstraction, deductive, 13
abstraction, inductive, 12, 13
accelerated motion, 280–283, 284, 291
accommodation, 177
achromatic colors, 56
action continuity, 389
active-glasses system, 180–181
Adams, Donald, 133n12
additive color mixing, 59–62
additive primary colors, 59
aerial perspective, 163–170, 173–176, 186n2
aesthetic context, 9–11
aesthetic energy, 68–69
aesthetic entropy, 265
aesthetic motion paradox, 272–274, 285n4
aesthetics
accelerated motion, 280–283, 284, 291
applied media, 3–16
colors, 55–57, 68–69, 78–83
complexity editing, 408–409
dissolve, 302–305
film look, 66
film structure, 278, 279
human condition, and, 408–409
size, 98
slow motion, 277, 280–283
sound continuity, 327–328
stereo 3D predicament, 215–216
two-dimensional field, 150–151
zoom vs. dolly, 299
Akutagawa, Ryūnosuke, 245n9
Allen, Woody, 83, 85n12, 237, 318
Alten, Stanley R., 367n2
ambisonic sound, 321, 333
anaglyph, 180, 181, 214, 219n13
analytical montage, 396–400, 403, 406–407
angles. See camera angles
apparent-motion phenomenon, 278, 284
appearance continuity, 390
Applebaum, Mark, 352–353, 362, 367
applied media aesthetics
art and, 5
contextual perceptions, 5–8
defined, 4
image elements, 12–15
medium as structural agent, 11–12
method of presenting, 12
power of context, 8–11
responsibility, 14–15
area, 68–69. See also two-dimensional field
Aristotle, 103, 261, 269n20
Arnheim, Rudolf, 35n1, 68, 71n15, 132–133n4, 142, 159n5, 269n26, 285n9, 332n1
articulating the z-axis, 185, 195–199, 211–212
artifact, 66
artificial landscape, 401
Art of Film Music, The (Burt), 324, 333n12
aspect ratio, 87–98
4 × 3 (standard TV), 88–91, 105
16 × 9 (HDTV), 88, 90–93, 94, 96, 97, 105, 106, 144, 234
defined, 88
flexible, 94–98
horizontal, 90
horizontal orientation, 87–88
matching, 94–96
mobile media, 92–93
pull of entire frame, 116
pull of side edges, 112, 114, 115, 132n2, 132
pull of top edge, 112, 113, 114
Index

aspect ratio (continued)
screens within the screen, 97–98, 154
secondary frames, 96–97
standard, 88–89
vertical, 90
associative context, 8–9
asymmetry of the frame, 117–120
“at-at” motion of film, 278, 285n4, 403
attached shadow, 22–23
attack, 339–341
attraction of mass, 116
Audi, Robert, 285n4
audio continuity, 391–392
audio/video balance, 329
audio/video montage, 406–408
Augustine, Saint, 257, 268n14
Autumn Sonata,
Bach, Johann Sebastian, 346–351, 357, 367n3
background
aerial perspective, 169–170, 173–176
cameo lighting, 42–44, 46
chiaroscuro lighting, 38–44
flat lighting, 44–46
high-key lighting, 30
limbo lighting, 46
low-key lighting, 30–31
Rembrandt lighting, 42–43, 47
screen volume, and, 189–190
silhouette lighting, 46–47
surrounding colors, 65–66
time orientation, 28–30
background light, 38–39, 40, 41, 43, 46, 47
back light, 27, 37–38, 39, 41, 43, 44, 45, 47, 51n3, 64. See also lighting
BAC positioning, 375
bad cut, 301
balance
dynamic, 142–143, 146, 158
golden section, 143–144, 159n7
labile, 146, 222, 241, 302, 395
rule of thirds, 144–145, 159n7
stages of, 142–146
static, 142–143, 146
unbalanced screen space, 147
Barbatsis, Gretchen, 159n13
barn doors, 47, 48
Barnes, Susan B., 159n2
Bary, Ann Marie, 309n10
baselight, 40, 41, 64
Battleship Potemkin, 282
Bauhaus, 12, 17n18–19, 27, 35n5, 80
Beech, Michael, 219n17
Beer, Johnannes, 159n9
Begleiter, Marcie, 245n14
below-eye-level lighting, 30, 31–32
below-eye-level point of view, 231, 232, 239, 241
Belton, John, 107n5, 332n1, 333n6
Bergman, Ingmar, 83, 261
Bergson, Henri, 268n10
Best Intentions, The, 83
Binocular depth, 3D, 161, 176–177, 185
biological time, 251, 255–256
Birth of a Nation,
Binet, John, 285n4
Bosklo, John, 268n3, 268n7
Bosklo, Anna, 159n8
Brecht, Bertolt, 218n9
brightness, 56–57, 66–67, 69
brightness constancy, 66–67
Brown, Blain, 35n6
Bruner, Jerome S., 17n10
Bucy, Erik, 219n11
building screen space. See visualization
Burnham, Jack, 85n5
Burns, Gary, 218n8
Burt, George, 324, 333n12
Button, Bryce, 410n1
cameo lighting, 42–44, 46, 47
camera. See also camera angles; lens
brightness, 47, 48, 57, 66, 70
canting, 10, 92, 111
color composition, 76–77
color constancy, 66–67
HDTV, 124
moving, 98, 121, 280, 297, 298
placement, 238–239, 377–378, 385, 386–389
subjective, 234–237, 241, 244, 245n12, 292
surrounding colors, 65–66
3D, 177, 178, 181, 297
white-balancing, 65
camera angles
continuity, 237–243
event intensification, 239–241
height in plane, 166
multiple viewpoints, 238–239
style setting, 241–242
visualization, 224, 230
canon, 349, 350, 366
canting, 92, 111
Capra, Fritjof, 269n18
Caravaggio, Michelangelo Merisi da, 38
Castel, Louis-Bertrand, 80
cast shadow, 22–24
Cathcart, Robert, 219n21, 245n7
character’s point of view, 234–236
character time, 292–293
chiaroscuro lighting
analysis, 38–40
cameo, 42–44, 46
defined, 37
directional function, 40, 41
emotional, 42
organic function, 40
Rembrandt, 42, 43, 47
spatial/compositional function, 41–42
thematic, 42
chord, 335, 343, 344, 345, 352, 353, 364, 366, 400
Christo, 250
chroma, 56
chroma-key backdrop, 45, 64
chromatic colors, 56, 84
chromatic scale, 342
cinema. See digital cinema
circadian cycle desynchronization, 255
Citizen Kane, 218n2
Clairborne, Robert, 268n7
Clemens, Samuel L. (Mark Twain), 317
Clooney, George, 219n14

424
close-up (CU), 228–229
  4 × 3 aspect ratio, 90, 91
  continuity, 238
cross shooting, 232–233
deductive visual approach, 224–226
event intensification, 239–241
  field of view, 229–230
  figure/ground, 121, 326, 327
  inductive visual approach, 227–228
  lateral motion, 275
  lens choice and blocking, 275–277
  low-definition image, 124
  on mobile media display, 104
  multiple viewpoints, 239
  present as subjective time, 258
  primary motion, 294
  relation to screen area, 100, 101
  relative size, 165
  shot time, 290
  close-up sound, 327, 332
  closure, 81, 123–126, 147–151
  facilitating, 125, 147
  graphic cues, 148–149
  illogical, 150–151
  psychological, 123–126
  3D, in, 158
  CMY (cyan, magenta, yellow), 59, 63
  cold colors, 68, 70, 77, 78, 79
  comparison idea-associative montage, 402–404, 407–408
  collision montage, 402–404, 407–408
  color
  achromatic, 56
  additive mixing, 59–62
  aerial perspective, and, 170
  aesthetic energy, 69
  aesthetic perception factors, 55–57
  attributes, 55–57
  brightness, 66–67
  cold, 68, 69, 77, 78, 79
  compatible, 59
  complementary, 61–62
  compositional function, 76–78
  constancy, 66–67
  continuity, 390
  defined, 53–54
  desaturation theory, 81–82
distortion, 45, 66, 79, 125, 208, 214
dramatic agent, as, 79
  energy, 68, 69, 77–78
  expressive function, 78–83
  feelings, and, 67–68
  grayscale, 57
  harmony, 73, 76
  informational function, 73–74
  juxtaposition, 66
  L*a*b model, 58
  light environment, 64
  mixed mixing, 63
  mixing, 59–62
  models, 57–59
  mood, 80–81
  perception, 54–59
  primaries, 59–60
  relativity, 63–67
  sound, and, 80
  subtractive mixing, 62–63
  surface reflectance, 64
  surrounding, 65–66
  symbolism, 74–75
  temperature, 65
  vibrations, 67
  warm, 68, 69, 77, 78, 79
  color attributes, 55–57
  color constancy, 66–67
  color continuity, 390
  color energy, 68, 69, 77–78
  color harmony, 73, 76
  colorizing film, 82–83
  color models, 57–59
  color temperature, 65
  color vibrations, 67
  Colour-Music: The Art of Light (Klein), 80
  communication
  model, 12
  responsibility, 14–15
  sound and noise, 312
  Communtiz-Page, Georgette, 159n13
  comparison idea-associative montage, 401–402, 407
  comparison montage, 401–402
  compatible color, 59
  complementary colors, 61–62
  complexity editing
  aesthetics, 408–409
  analytical montage, 396–400, 403, 406–407
  audio/video montage, 406–408
  collision montage, 402–404, 407–408
  comparison montage, 401–402, 407
  idea-associative montage, 402–408
  metric montage, 405–406
  sectional analytical montage, 398–400, 407
  sequential analytical montage, 396–397, 406–407
  visual dialectic, 404–405
  composition
  color, 76–78
  line, and, 6
  rule of thirds, 144–145
  shot sequence, 136
  structuring two-dimensional field, 135
  unbalanced, 147
  unusual, 151–152
  compressing, 172
  conceptual art, 250
  Confeessions (St. Augustine), 257
  connected frames, 157
  constancy
  brightness, 66–67
  color, 67
  size, 98
  context
  aesthetic, 9–11
  associative, 8–9
  power of, 8–11
  sound, 315
  temporal and spatial, 155–157
  time, 155
  contextual perception, 5–8
  continuing graphic vectors, 130
  continuing index vectors, 130, 371
  continuing motion vectors, 381–382
  continuing vectors, 130, 131, 371
  continuity
  action, 389
  appearance, 390
  audio, 391–392
  camera angles, 237–238
closure, and, 125, 126
  color, 390
depth, 391
dissolve, 302–305
  editing pace, 391
  motion vector, 381–388
  off-screen, 374
  position, 377
  rhythmic, 325
  sound, 327–328
  subject, 389–390
vector field, 127–131
directions, 129
graphic, 127, 128, 370–371
index, 127–128
motion, 128–129
types, 127
vector line, 377, 380, 385–388
vector magnitude, 129, 264–265
velocity, 283
velocity change, 283
Venkatesh, Svetha, 17n20, 133n15
vertical aspect ratio, 90
vertical screen orientation, 109–111
vertical time vector, 253–255, 291
viewer-to-window distance, 211–213
viewpoint, 228–230
virtual reality (VR)
sound, 320, 333n9
visible radiant energy, 20
visible spectrum, 20, 54
visual dialectic, 404–405
visual fluff, 95
visualization
angles, 237–243
basics of, 221–222
deductive approach, 224–226
field of view, 228–230
inductive approach, 227–228
point of view, 230–237. See also point of view (POV)
storyboard, 243
viewing approach, 222–224
visual pollution, 196
visual rhythm, 126, 329
visual syntax, 311, 369, 370, 392
volume. See three-dimensional space
volume duality, 190–196, 200, 211, 216, 297
von Helmholtz, Hermann L. F., 80
von Sternberg, Josef, 190
von Weber, Carl Maria, 323
VR. See virtual reality
Wagner, Richard, 323
Ward, Peter, 95, 107n6, 159n7
warm colors, 68, 77, 79, 81
Wattie, John, 187n8, 219n13, 219n16
Weaver, Warren, 17n16
web page asymmetry, 119, 120
wedding switch, 379
Weis, Elisabeth, 332n1, 333n6, 333n16
Welles, Orson, 218n2
Well-Tempered Clavier, The, 349, 350, 351, 367n3
Wertheimer, Max, 126, 133n8, 159n10, 285n3
Wheatstone, Charles, Sir, 187n9
white balance, 65
white light, 65
white sunlight, 20, 65
whole-tone scale, 342, 343
aerial perspective, 169–170, 173–176
depth characteristics, 170
distortion, 198–199
height in plane, 166–167
linear perspective, 167–169, 172
motion, 276–277
overlapping planes, 164–165, 170–171
relative size, 165–166, 171–172
wide-angle-lens distortion, 198–199
wide-screen format, 89
wide-screen letterbox, 95
wide-screen motion picture aspect ratio, 89
Wilfrid, Thomas, 80
Willats, John, 219n10
Willett, John, 218n9
Wilson, Benjamin P., 333n7
windowboxing, 94, 95, 97, 106, 204, 205
window violation, 211, 212–213, 214, 215, 216, 217, 219n17, 219n20, 234, 244
wing frames, 156
Wingler, Hans M., 17n19
wipe, 305
Wittig, Susan, 107n1, 218n9
Wol, Mike, 159n3
Wolfe, Thomas, 268n6
Wölfflin, Heinrich, 132n4
Woods, Katherine, 411n16
Wright, Frank Lloyd, 148
Wright, Richard D., 269n17
XCU. See extreme close-up (ECU)
y-axis, 161, 162, 181
Yeh, Tiffany, 245n8
Zakia, Richard, 17n11, 133n11
Zalta, Edward N., 268n2
Zavarin, Valentina, 107n1, 218n9
z-axis
articulation, 185, 195–199, 211–212
defining sound, 320
double, 163, 203
index vector, 127–128, 131, 377
motion vector, 128–129, 276–277, 384–385, 386–388
perceived speed, 274–277
3D space, 161–163, 211. See also three-dimensional space
vector, 129, 155, 207, 376
vectors in split screen, 155–156
wedding switch, 379
z-axis index vector, 127–128, 131, 377
z-axis vector, 129, 155, 207, 376
Zener, Karl, 133n12
Zeno of Elea, 285n4
zero point in time, 254
Zettel, Erika, 403, 411n8
Zettl, Herbert, 16n9, 17n20–21, 35n2–4, 51n2, 51n4–7, 71n7, 71n10–13, 85n10, 107n4, 133n14, 186n3, 218n8–9, 219n22, 245n1, 245n13, 269n27, 285n7, 309n1, 309n6, 309n9, 333n14, 333n16, 393n1–5, 393n7–8, 410n3
Zingrone, Frank, 17n13
Zone, Ray, 187n9
zooming and dollying, 277, 295, 296–299