# 1 Introduction

Clearly everybody knows the word "multimedia," yet when people think of it, they usually think of different things. For some people, multimedia equals entertainment. For other people, multimedia equals Web design. For many computer scientists, multimedia often means video in a computing environment. All these are narrow perspectives on multimedia. For example, visual information definitely dominates human activities because of the powerful visual machinery that we are equipped with. In the end, however, humans use all five senses effectively, opportunistically, and judiciously. Therefore, multimedia computing should utilize signals from multifarious sensors and present to users only the relevant information in the appropriate sensory modality.[1]

This book takes an *integrative systems* approach to multimedia. Integrated multimedia systems receive input from different sensory and symbolic sources in different forms and representations. Users ideally access this information in experiential environments. Early techniques dealt with individual media more effectively than with integrated media and focused on developing efficient techniques for separate individual media, for example, MPEG video compression. During the past few years, issues that span multimedia have received more central attention. Many researchers now recognize that most of the difficult semantic issues become easier to solve when considering integrated multimedia rather than separate individual media.

In the early days of computing, science fiction writers envisioned computers as robots that would effectively use audiovisual data. Later, people dreamt of systems that could organize audio files, images, and video. Now people want to share perceptual experiences independent of time and distance. Within the next few years, most of the data stored on computers – in terms of storage size and bandwidth requirements – will be audiovisual. Therefore, the fundamental media for computing and communications will also be audiovisual and will increasingly include tactile components.

Handling multimedia content requires incorporating concepts and techniques from various disciplines – from signal processing, from communication theory to image data-

[1] In this book, we use the notions of sensory modality and media interchangeably, defined as data generated from one particular type of sensor.

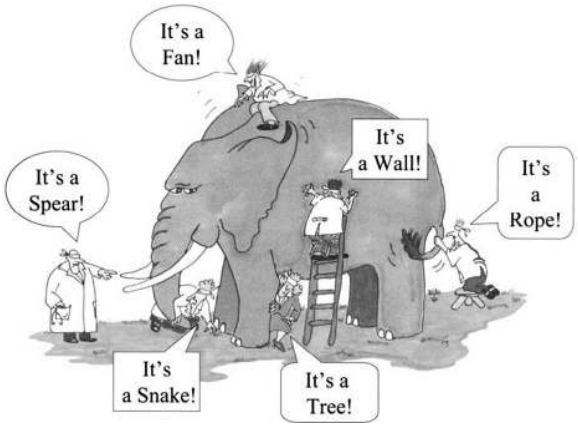


**Figure 1.1.** Multimedia is like an elephant. Looking at it from a limited perspective leads to many completely wrong characterizations.

bases, and from compression techniques to content analysis. Multimedia computing has consequently evolved as a collection of techniques from different disciplines.

Because of the diversity of types of information and the evolution of technology, application development, as well as multimedia research, has evolved in a way like the elephant in the fable about the elephant and the six blind men (see Figure 1.1). In this fable, each blind man has a limited perspective. In real life, people impose limitations of perspective in many ways and hence – though naturally endowed with multiple sensory and cognitive faculties – functionally behave like these blind men portrayed in the cartoon: each engineering and research discipline perceives multimedia from its own limited viewpoint. This has resulted in a skewed development of the field, where multimedia is perceived as multiple monomedia fields.

We (humans) use our five senses (sight, hearing, touch, smell, and taste) together with our abstract knowledge to form holistic experiences and extract information about the world. Multimedia computing aims to develop communication techniques to share holistic experiences from multiple sources and modalities of data and to extract useful information in the context of various applications.

This fragmented perspective of multimedia has slowed progress in understanding and processing multimedia information, although the hardware used for processing it – ranging from sensors to bandwidth – has advanced rapidly. Multimedia computing should leverage correlated and contextual information from all sources to develop holistic and unified perspectives and experiences. It should focus on full *multi*sensory experiences rather than partial experiences, such as listening to an audio-only sports commentary.

This book presents emerging techniques in multimedia computing from an experiential perspective in which each medium – audio, images, text, and so on – is a strong component of the complete exchange of information or experience. Humans are the best functioning example of multimedia communication and computing – that is, we understand information and experiences through the unified perspective our five senses offer. Our goal in this book is to present current techniques in computing and communication

that will lead to the development of a unified and holistic approach to computing using heterogeneous data sources.

By describing the properties of perceptually encoded information, presenting common algorithms and concepts for handling it, and outlining the typical requirements for emerging applications that use multifarious information sources, this book introduces the fundamentals of multimedia computing. It serves as an introduction to engineers and researchers interested in understanding the elements of multimedia and their role in building specific applications.

## ORGANIZATION OF THIS BOOK

We organized this book to present a unified perspective on different media sources for addressing emerging applications. The chapters in this book are organized for linear reading even though the reader may choose to skip some of them. Often, we are not able to explain a topic in full detail; in those cases, we provide pointers to literature. Our main goal is to present concepts, techniques, and applications that will be useful in building integrated multimedia systems. The nature of the field mandates that this includes not only concrete algorithms, but also high-level concepts and strategic knowledge that illustrate the big picture of the problems to be addressed. We believe that the holistic viewpoint presented in this book is essential for understanding, using, and communicating emerging applications in multimedia.

### Chapters 1–3 Defining Multimedia Systems

Evolution of technology related to communication and computing has resulted in fragmentation and limited perspectives in multimedia computing. The current stage of the multimedia field brings to mind the parable about the six blind men and the elephant; we therefore track the evolution of the field briefly and then define multimedia systems and discuss their main elements in these chapters. The big picture concepts in these chapters will help us discuss all elements concurrently in the subsequent chapters without losing the whole-system perspective.

### Chapters 4–6 Nature of Perceptually Encoded Information

Like humans, multimedia systems gain information and experience through a variety of sensory and other sources. Understanding the relationships among data, information, knowledge, insight, and experience is crucial to being able to use these sources judiciously. We discuss basic elements of information and data source types, mostly focused on but not limited to audio and visual data, in the context of multimedia systems. These areas are well established and many other sources provide details on every aspect of representation and processing. Our goal here will be to present the essential elements from those areas and direct readers to sources for more information.

### Chapters 7–11 Fundamental Properties of Multimedia Applications and Systems

Once multimedia data is acquired through sensors, it needs to be transmitted, stored, and reproduced. Users often use production environments to edit and create multimedia presentations out of the raw sensor data. In this part of the book, the fundamentals concepts of multimedia systems and applications will be discussed.

### Chapters 12–15 Compression

Multimedia data is usually continuous time-dependent data that is significantly more voluminous than other data. Often, though, a system's input, processing, and output elements are at different locations. A large volume of data must therefore be communicated to different locations. The cost of these communications, as well as storage, must be reduced using data-compression techniques. The field of data compression has matured and been largely implemented in industrial products because most of these techniques have responded well to multimedia systems' needs. When working with multimedia data, it is therefore almost inevitable to think about methods to handle compressed data. This part of the book presents fundamental algorithms and ideas and provides sources for the reader to go into further details based on her interest.

### Chapters 16–19 Organization and Analysis of Multimedia Content

Most multimedia systems are for collecting, processing, creating, and storing data to extract information and knowledge from it. Most applications of multimedia systems are for accessing, sharing, and using this data, information, and knowledge. In a data management system, structuring techniques are used to index and organize data for efficient access. Organizing multimedia data for search and navigation has been a challenge. Even organizing individual components such as audio, images, and video presents difficulties. During the past few years, the types of sensors used, as well as the volume of data acquired, has continuously increased, making this problem even trickier. This part of the book introduces techniques and approaches to structure, index, and access multimedia data.

### Chapter 20 The Future of Multimedia

Multimedia is a very dynamic and ever-changing field. Therefore, we end this book with an outlook and with techniques that can help with finding a strategy to predict the development of multimedia computing that allows the readers to position themselves in the big picture.

## RECOMMENDED READING

The story of an elephant and six blind men has been used many times by different authors, most famously by John Godfrey Saxe, to bring an ancient viewpoint from India. This story

illustrates how a limited perspective may lead to erroneous understanding of a complex object or concept. James Gleick's book *The Information* is a very good exposition to the nature of information both from a mathematical and from an evolutionary perspective of information. Current multimedia systems have now evolved to the stage where, for the first time in history, those working with such systems can seamlessly combine different modalities. Another very good read for students of multimedia is Samuel Hayakawa's *Language in Thought and Action*. This book explains how languages evolve and is very insightful.

## Further Reading

James Gleich. *The Information: A History, a Theory, a Flood*. New York: Pantheon Books, 2011.

Samuel Ichiye Hayakawa. *Language in Thought and Action*. Harcourt Brace Jovanovich, 1978, Orlando, Florida.

John Godfrey Saxe. *The Blind Men and the Elephant* (retold).

# 2 Multimedia: A Definition

Only a few inventions in the history of civilization have had the same impact on society in so many ways and at so many levels as computers. Where once we used computers for computing with simple alphanumeric data, we now use them primarily to exchange information, to communicate, and to share experiences. Computers are rapidly evolving as a means for gaining insights and sharing ideas across distance and time.

Multimedia computing started gaining serious attention from researchers and practitioners during the 1990s. Before 1991, people talked about multimedia, but the computing power, storage, bandwidth, and processing algorithms were not advanced enough to deal with audio and video. With the increasing availability and popularity of CDs, people became excited about creating documents that could include not only text, but also images, audio, and even video. That decade saw explosive growth in all aspects of hardware and software technology related to multimedia computing and communication. In the early 1990s, PC manufacturers labeled their high-end units containing advanced graphics *multimedia PCs*. That trend disappeared a few years later because every new computer became a multimedia computer.

Research and development in multimedia-related areas has been around for much longer. Research in speech processing, speech compression, and speech recognition was fueled first by telephony and then by digital sound applications. Image and video processing and compression have also been active research and development areas because of digital photos and video.

Before 1990, much of the research in audio and video compression, storage, and communication was driven by broadcast and consumer electronics related to entertainment applications. During the 1990s, combining these sources in a computing environment emerged as a clear possibility. As a result, research in all areas of audio and video received significantly greater emphasis.

In the following, we describe the historic evolution of communication to relate it to current multimedia computing. The current technology in multimedia computing and communications is only a stage in the evolution of humans' desire to share their experiences and to extract knowledge from those experiences. A clear understanding of the goal of multimedia computing helps in developing efficient approaches to deal with it.

## COMMUNICATION IN HUMAN SOCIETY

The ability to effectively communicate complex facts and interrelationships is one of the main features that distinguishes humans from animals and has been a major force in human evolution. Communication lets us share experiences and create, maintain, sustain, and propagate knowledge. This innate desire resulted in several influential inventions that determined the progress of human civilization. Communication across space allows participants to exchange information independent of their current location, and communication across time allows observers to experience an event over and over again without having to be there at the exact moment. Mankind's quest for communication of experiences across both time and space is one of the defining foundations of multimedia. A third factor is the desire to communicate to as many people as possible, thus duplicating communication easily. Communication thereby becomes invariant to the number of addressees in space and time. As Table 2.1 shows, human civilization has seen many influential inventions related to communicating experiences across space, time, and addresses.

Human communication exists in many forms, including facial and body gestures, olfactory signals, and of course spoken language. Out of these, spoken language is the most capable of conveying the most complex facts, that is, the information density is very high. For most of humans' existence, though, spoken language only consisted of analog sounds uttered with the speech-producing infrastructure in the throat. Eventually, people realized that experiences were important and that they should somehow store these events for sharing with others. Initially, drawings and paintings would preserve these experiences, but they were not precise enough to unambiguously convey complex facts, and they were too cumbersome to produce. This resulted in the invention of written language as a system for representing language to share experiences. Bulky preservation techniques such as stone tablets gave way to more practical storage devices and writing methods. Next came the development of paper and ink, and still more people began using the stored experiences that others had painstakingly recorded.

Then came one of the most influential inventions in our history: Gutenberg's movable printing press. This invention enabled mass communication for the first time and revolutionized society. By making creation, storage, and distribution of documents more manageable, this technology resulted in easy sharing of knowledge in society. Our current education system, our reliance on documents (such as newspapers) as a major source of communication, and on libraries as public, government-supported institutions dedicated to storing knowledge, stem from that one invention that appeared more than five hundred years ago.

The telegraph, which allowed instantaneous communication of symbolic information over long distances, began to bring the world closer. This invention signaled the beginning of the global village. Telephones let us return to our natural communication medium – talking – while retaining the advantages of instantaneous remote communication. People could experience the emotions of the person on the other end of the connection – something symbol-based methods of writing and the telegraph could only hint at.

Radio ushered in the wireless approach to sound and popularized sound as a medium for instantaneous mass communication. Movies and television took communication a

**Table 2.1. Communications-related inventions in human civilization**

| Invention | Resulting Application | Invariance |
|---|---|---|
| Spoken Languages | Natural communication | None |
| Written Languages | Symbolic record of language | Time |
| Paper | Portability | Time and space |
| Print | Mass distribution | Time, space, addressee |
| Telegraph | Remote narrow communication | Space |
| Telephone | Remote analog communication | Space |
| Radio | Broadcasting of sound | Space, addressee |
| Movies | Recording of sight and sound | Time, space, addressee |
| Television | Broadcasting of sight and sound | Space, addressee |
| Recording media | Recording | Time, space, addressee |
| Digital media | Machine enhancement and processing | Time, space, addressee |
| Internet | Personalized reception | Time and space (sometimes addressee) |

step further by combining our senses of sight and hearing and making communication more natural. It was the first medium that let us experience with more than one sense and as such was able to more effectively key into our emotions. Video communication's popularity is clearly due to its use of our two most powerful senses working in harmony.

Storage and distribution technologies, such as magnetic tape, allowed the storage, preservation, and distribution of sound, again bringing us closer to natural experience. Video recording enhanced this experience and advanced experience sharing to the next stage. Digital media further improved the quality of our experience by making it possible to copy and share information with controllable loss in quality. Finally, the Internet took information availability to a new dimension, providing us with experiential accounts of an unprecedented variety by combining different media appropriate to the message to be communicated.

## EVOLUTION OF COMPUTING AND COMMUNICATION TECHNOLOGY

The changes in the landscapes of both computing and communications have been overwhelming during the past few decades.

Just a few decades ago, a computing center was one of the most important buildings on a university or corporate campus. Access to this building, particularly to the "holy" room in which the computer operated, was highly restricted. A computer occupied several rooms,

**Table 2.2. Comparison of early computers with those of a typical handheld in 2011**

| Computer | Processing unit | Operating system | Core memory | Secondary memory |
|---|---|---|---|---|
| 1960s-era computer (IBM1620) | Could not do arithmetic, used look-up tables | No OS; human monitors controlled everything | 60 Kbytes | 2M characters |
| Handheld Experiential Computer (iPhone 4S) in 2011 | Dual Core A5 Chip, 800 MHz | iPhone OS | 512 Mbytes | 64 Gbytes |

**Figure 2.1.** A 1960s-era computer at NASA.(Source: U.S. National Archives and Records Administration)

if not floors, of a building; needed air conditioning; and required a specialized and trained staff to interact with it. These computers cost millions of dollars. Figure 2.1 shows a popular computer from the late 1960s and early 1970s. Table 2.2 lists important characteristics of the IBM 1620 computer, which is from that era.

Progress in processing, storage, networking, and software technology has changed computing beyond anyone's expectations. Today, most people carry computers in their pockets that are several orders of magnitude more powerful and sophisticated than the 1960s-era computer. Table 2.2 compares a current handheld to the early computer. Although the newest handhelds are more powerful and sophisticated, they cost several thousandths of what the older version cost, are easy to carry, and are much less affected by climate. Moreover, just about anyone can operate them, using them to solve everyday computing and communications needs.

Communications technology has experienced a similar overwhelming transformation. We've already discussed the historical perspective. Here, we focus on short-term technological improvements in one medium.

Consider the telephone. In its very early incarnations, the telephone had limited use. Only a few people could afford to have one in their homes. Moreover, a house had one phone, and when you called someone you had to shout into the mouthpiece. During a long-distance call, latency made communication difficult. Either both parties spoke at the same time or waited for the other, while an expensive meter ticked off seconds. Usually, people spent more time shouting "Hello! Hello!" than having a meaningful conversation. Now, users can talk on a phone while walking, running, driving, or flying in an airplane. Signal reception is so clear that you can whisper to a person on the other side of the globe. More important, not only is your phone a voice communication device, but it is also your connection to a computer network, a camera, a calendar and address book, an entertainment center, and a video communication device.

## WHY *MULTI*MEDIA COMPUTING?

To understand computing technology's evolution to its current state, as well as to project its future evolution, consider the applications that have been and will be driving the technology's development.

The first computer applications performed numerical computations using data in scientific applications, hence the name *computer*. Business was the next major driving application with so-called data processing as the driving operation. It brought alphanumeric processing and databases in focus for development. Major networking advances resulted in enterprise computing based on the traditional distributed processing approaches that eventually culminated in the Internet.

Personal computers were a major influence on computing. PCs ended reliance on a powerful central computer and brought several applications, including word processing, spreadsheets, and electronic mail, into the consumer space. Combining personal computing and Internet connectivity led to one of the most amazing communication revolutions that human civilization has seen so far: the World Wide Web. The progress continued, and laptop computers replaced most PCs. Laptops are now being replaced by even more personal and sentient computing devices – tablets and mobile phones. These mobile devices can be used for computing, communication, and much more. Moreover, they can use audio and visual mechanisms equally effectively as traditional alphanumeric computing.