Chapter 11. Multimedia, Hypermedia, and Entertainment Technology

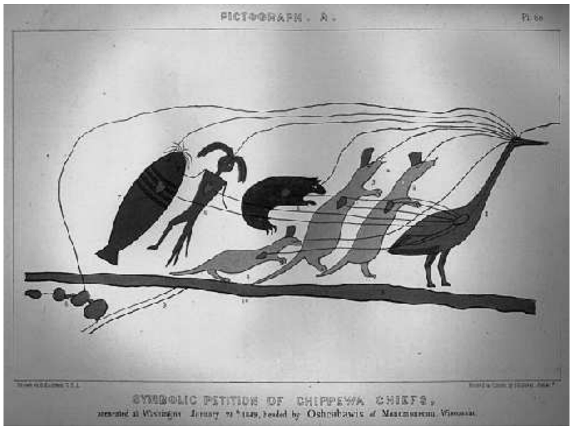


Figure 11.1: What is the meaning of an image? Pictograph (from www.library.wisc.edu/etext/) that represents a peace party of Native Americans represented by their totem animals. Their unity of purpose is shown with lines connecting their heads and hearts. Understanding such an image requires considerable cultural knowledge. (check permission)

11.1. Overview

Multimedia is content beyond text. It often emphasizes affect (4.6.0)[2]. Its scope has spread far beyond a few images and sound files by the rise of consumer electronics. Digital convergence. We can distinguish between multimedia and hypermedia.

On one hand, these technologies are rather different. On the other hand, they share many issues.

Processing, metadata, libraries.

We go beyond traditional multimedia to consider related issues such as 3-D copying. Mashups.

Multimedia surrounds us much more than it did in the past. Television programs have gotten much more complex. Television viewing habits. Social viewing. Video in applications ranging from meeting archives (5.6.4) to media spaces (5.6.6).

Representation for multimedia at several levels. Storyboard. Semantic annotation with identification and recognition.

The focus of the experience of multimedia is often different than for text resources. In particular, it is less often applicable for scholarship. Types of information needs for multimedia. Including for entertainment.

The Media Experience. Multimedia is a way of telling stories. Media only partially captures reality. Entertainment technology.

Comparative media studies.

11.1.1. Art

Art evokes an affective response (4.6.2). Art is often representational. Visual Art and Representation.

Social dimensions of art. Art museums (7.6.1).

Many levels of description and indexing for art. Images and culture. DOIs for art objects.

Paintings that emphasize linear perspective do not necessarily our subjective impressions of the visual world. Impressionist painting attempts to capture that (Fig. 11.2). Literal perspective. From representational art to abstract art.

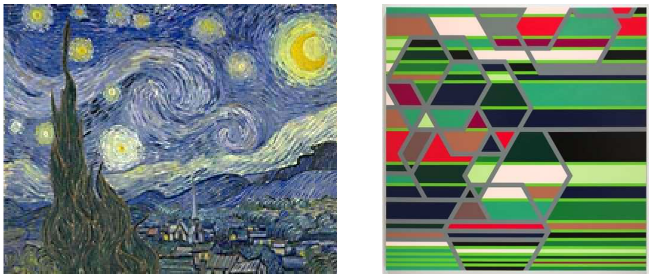


Figure 11.2: Impressionist (left) and abstract art (right). (check permissions)

Such representations capture aspects of the world, or perhaps just patterns. There may be a cultural meaning associated with images[31]. How should that affect indexing.

Multimedia and Culture

We have already seen print culture (8.13.6). Fan groups.

Participatory culture.

11.1.2. Multimedia Libraries

LSCOM (Large-scale Concept Ontology for Multimedia)

YouTube channels.

Figure 11.3: Museum of TV and Radio collection.

11.1.3. Processing Multimedia

Types of indexing for multimedia.

Segmentation of multimedia objects.

Fixity of multimedia objects changes their nature.

Common issues for processing across multimedia types, segmentation, frequency.

Interactivity and hypermedia.

Annotations for the internal structure of complex media.

11.1.4. Representations for Multimedia

Hypermedia Models

Synchronization.

11.1.5. Interactive Hypermedia

Interaction design (4.8.1). Need to manage the user's attention.

Interactive collages.

Representing Multimedia and Hypermedia

We have emphasized representations throughout this text. What expressiveness is required for a complete multimedia model? Coordination among the media in a hypermedia presentation.

The unique aspect of hypermedia is interactivity. There are also issues of sequence and coordination in hypermedia structure. We would like to have device-independent multimedia.

State and Language-Based Hypermedia Models State machines (3.10.1). Discrete state hypermedia. be described in sequences of states. Features that might be needed include looping, concurrency, synchronization points, and alternate or optional paths. The real trick is to determine how the model is structured. ATNs (6.5.1) are a better representation for multimedia.

Ultimately, we might use the models we have described earlier. From semantic graphs to multimedia presentations[23].

Temporal Scripting Languages manage events in time. Sometimes scripting languages are created for the interaction of media. These are different from languages for the interaction of objects in the media. For instance, they may support looping. Animation languages.

Multimedia and Hypermedia Authoring

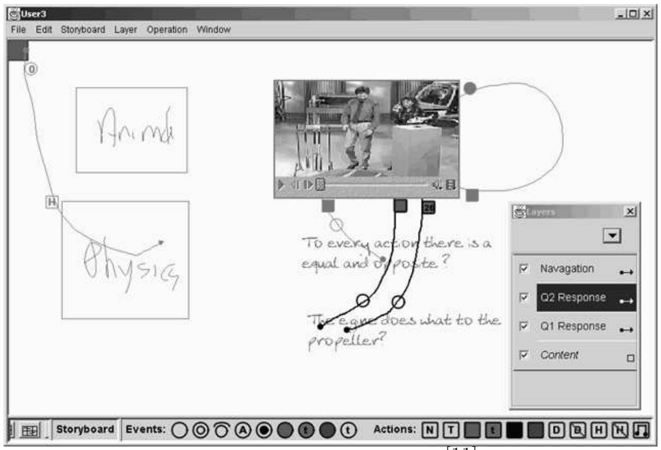
The definition of the word "authoring" has changed over time. While it once meant simply to write, it is now used much more freely to describe the act of creating a multi- or hyper-media information object. Authored objects of this sort typically included many forms of media, such as text, video, audio, hypertext, and animation. As was noted earlier, authoring is a design activity (3.5.4); there is purpose and rationale behind the structure of the information object, or the way in which the various media modes interact and work together. The ultimate purpose of most multimedia authoring projects is to create structured content, be it for logical, aesthetic, or educational ends.

Multimedia authoring is a complex activity. It can be conceived of as combining the difficulties of writing, drawing, sound engineering, and video editing. Perhaps the greatest difficulty is specifying the synchronization of these various mediums. One tool for aiding in this synchronization is storyboarding.

Storyboarding has been used for a long time in movie making and advertising, to name only two examples, to visually organize the interaction and transitions between different conceptual elements or narratives. In the same way, storyboarding is used to synchronize or organize the different media applications in a multimedia object. While it is similar to a navigation map, a multimedia storyboard may not contain all the eventual detail, such as hyperlinks, of the authored object, but it will contain sketches of the layout of the different pages. In many ways, a storyboard for a multimedia object is far more complex than one for a movie or other linear formats because of the interactivity between the different mediums.

Movement through a multimedia object does not progress only in one direction, but skips and jumps from place to place; this makes its organization, visual or otherwise, difficult. Just as multimedia comes in many versions, so too are there many types of storyboards. The main characteristic of generalized multimedia storyboards, however, is the ability to specify interaction between media; for instance, how links should be specified and organized (2.6.3). Many storyboarding tools utilize intuitive, user-friendly interfaces, such as drag-and-drop authoring[11](Fig. 11.4) for organizing material, rather than formal or technical descriptions for media placement. Multimedia content may be evaluated by a person interacting with the environment via "walkthroughs" [20].

Path through a multimedia application matches information access. Layout and use-cases.



To every action there is a equal and opposite?

The engine does what to the propeller?

Figure 11.4: Drag-and-drop authoring for multimedia[11]. (check permission)

Multimedia Desktops

Web-based portals and even personal computer desktops (3.5.4) can be seen as multimedia environments in which users can interact with various information services. A workspace can collect work in progress (4.11.2) and many computer applications, such as text editors or even the entire desktop interface, are types of discrete-state multimedia objects.

These objects exist within a larger environment, and are controlled by defined multimedia events: "tool was opened," "data was entered," "tool was closed". Although each of these tools or objects may operate independently, they nay be required to interact and to function together, and in a larger sense they all need to obey the laws that govern the operating environment. Multimedia environments and the tools that populate them can be quite extensive, and when developing environments such as this it often becomes necessary to design exactly how the interface will work.

11.2. Visual Information, Visual Languages, and Visualization

2-dimensional visual materials can convey meaning in several ways. Images approximate our visual perception and capture spatial relationships and colors inherent to objects in the world. Images are merely representations of those objects and relationships, and may not contain all the intellectual or emotional context that comes from visualizing first-hand. The context that a representation conveys is therefore dependent in part on the viewer and in part on its creator; the creator to present the image in the fashion that they believe conveys the message they want, and the viewer to contextualize the image according to their own experience. This dichotomy allows images to communicate an enormous range of information in a purely visual format, from images that simply duplicate an image of the world, to those visualizations that seek to use a representation of the physical world to convey more abstract concepts.

Human beings primarily use vision to understand the world — this makes the visualization of images and data structures a particularly effective means of communicating information. Imaging and visualization systems can be combined with descriptive abilities to further enhance this effectiveness — note how textbooks always use images and figures to reinforce education. We are only now beginning to tap into the power of visualization to analyze large amounts of data and data libraries. Other means of using imaging and visualization to enhance information science and systems include schematics that show visual representations of abstract relationships and spatial information systems that deal with the position of objects in space.

Cognition and images. Brain processing and images (-A.12.2).

11.2.1. Representations and Images

Pictures record information. Before the widespread use of the printing press most recorded communications were pictorial. This form of communication proved to be very effective in large part due to the ability of images to convey complicated messages. A photographer or painter has many ways to draw the viewer's attention and convey meaning, often very subtly — images may be posed or edited to present a particular viewpoint or perspective, and they can be used sequentially to narrate a story. Indeed, many images or sets of images are very language-like. In pictorial presentations, one can perceive meaningful units (4.2.1) that function as a whole. Composition, or the way these units or objects are arranged (or the manner in which they are viewed or perceived) can amplify the significance that may be inherent to the objects in the first place. In Fig. 11.5, the repetition of objects and the viewpoint from which they are seen creates a rhythm. Other visual information objects, such as schematics and maps, can be even more language-like than a single image (11.2.4). How do we understand and represent space in images[32].

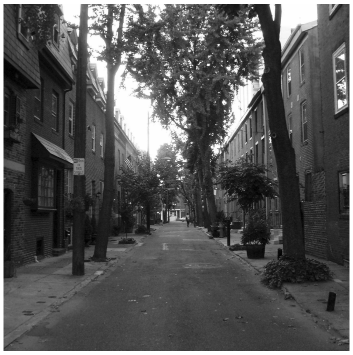


Figure 11.5: Note how the convergence of the parallel lines draws the eye.

Use of images in engineering, science, and medicine.

Attribution of painter[22].

11.2.2. Image Processing

More image processing (-A.2.3).

Edge detection.

Affine transformation.

Object recognition.

Photo sequencing.

11.2.3. Image Collections and Retrieval

Image metadata. It has proven difficult to develop useful descriptive systems for images. This may be because there are too many difficult tasks to be satisfied by any small set of descriptive systems.

Photo matching. Compare cellphone pictures to image database.

ARTSTOR. Medical images (9.9.2).

Value of metadata for image retrieval. Too many image retrieval tasks to be satisfied by any one metadata system.

Query by image content.

Computational photography.

Searching Images

Difficult to extract semantically meaningful features automatically. Described with metatadata. GWAP (2.5.4).

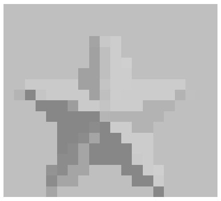


Figure 11.6: Image search.

11.2.4. Visual Languages

Visual objects can be highly structured. A language was defined earlier as the output of a lexicon and a grammar (6.5.2). Visual presentations can have some properties of a language. We may say there is a visual language, which includes a lexicon, a syntax, and semantics. This may occur within a single image but more often is found in composite images or sketches. We have already discussed the visual layout of documents (2.3.3). We will see similar structural implications for cinema (11.6.3) and even the structure of Chinese Opera. Visual language and visual literacy.

Visual Lexicon and Structure

As with any language, a visual language is composed of a lexicon and a structure. Icons are one part of the lexicon of the visual language (Fig. 11.7). The icons indicate functions they represent in different ways. Some use metaphor; a file system may be represented by an icon showing a filing cabinet. Some things are difficult to represent with icons and some icons are obscure. The selection of effective icons with arbitrary meanings can be a challenge for interface design. Some icon-based systems show text describing an icon's purpose when the cursor is placed over it.

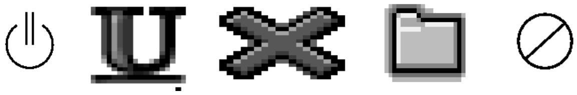


Figure 11.7: Examples of icons associated with computing. (redraw) (check permissions)

Visual Grammar and Syntax

Graphical design.

As with natural language, managing the focus of attention is necessary for visual language processing.

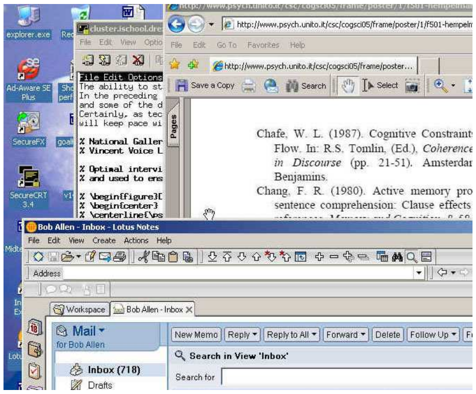
Cartoons is designed to highlight the significant points in a narrative.

Diagrams employ visual language. Design of visual language presentations. Gestalt principles (4.2.1). Layout.

Visually clustered concepts are viewed as more related. Concepts linked with a line, are believed to be particularly closely associated.

Visual parsing (Fig. 11.8). There is often a syntax-like structure in visual materials. An arrow on a display may indicate the movement; in other words, it functions as a verb. Lines show connections between concepts or to separate one set of concepts from others. These lines form a type of visual punctuation.

Visual similarity.



```
File Edit Options
The ability to st
In the preceding
and some of the d
Certainly, as tec
will keep pace wi

% National Galler
% Vincent Voice L

% Optimal intervi
% and used to ens

% \begin{figure}[
% \begin{center}
% \centerline{\ps
```

Chafe, W. L. (1987). Cognitive Constraint Flow. In: R.S. Tomlin, (Ed.), *Coherence in Discourse* (pp. 21-51). Amsterdar Benjamins.
Chang, F. R. (1980). Active memory pro sentence comprehension: Clause effects

Figure 11.8: A complex scene such as this screen dump of a computer desktop can be segmented by visually parsing its components.

Visual Discourse

Visual discourse. Creating meaning and impressions by associations. [?]

11.2.5. Information Visualization and Interactive Visualization Environments

Visualization presents representations of the attributes of entities and the relationships among those attributes. These are based on information structure but they also add interactivity for exploring data. There are two types of visualization: information visualization and data (or scientific) visualization. Data visualization focus on the display of quantitative values. Scientific visualization and visual analytics (9.6.5).

Information visualization concentrates on qualitative values. They can be thought of interfaces for database attributes. Schematics and diagrams typically show qualitative visual descriptions but information visualization adds interactivity. However, primarily a schematic is a model.

Browsing Hierarchies

Focus + Context. Visualization of hierarchies. People often lose the context of the information they are accessing. This is one of the reasons that books, such as this textbook, include chapter headings at the top of each page. Zooming text and zooming images.

Interactive visualization tools often allow users to view content at a high level and the focus is on details while keeping a context of the high-level content. For browsing hierarchical structures.

Zooming

Zooming allows the user to control the level of detail when examining objects. With logical zooming, the display simply highlights different aspects of a display in continuous or discrete steps. This makes most sense when there is a spatial relationship among the objects. Spatial hypertext (2.6.2)[??] Many conceptual systems are best understood at different levels of granularity. Consider, for instance, a user viewing a galaxy and zooming in to look at stars within that galaxy. While it is possible to zoom smoothly into an image of a physical system such as a view of outer space, information spaces have discrete steps. Powers of 10[5].

Logical or Semantic Zooming: Graphical zooming may be accompanied by richer description descriptions. “Semantic zooming”. One might “zoom” into the structure of the information. The relationships among objects can be highlighted with color and linking. Getting more meaningful detail in regions of a display.

Hierarchies are easy to understand and are widely used. When the relationships between entities are purely conceptual, only the category labels need to be displayed, as in tables of contents (TOCs) (2.5.5).

Menus display hierarchical actions or concepts without context. Menus can be seen as structured hypertexts (2.6.2). It is possible to generalize the logic of fisheye views to 2-D fisheyes and multi-foveal fisheyes. 3-D depth perception provides approximations of layered depth. Fig. **??** illustrates navigation in multi-scale space; this is one way of displaying visual context.

Lenses and Filtering

A traditional magnifying glass enlarges all parts of an image. By analogy, a visualization lens could give the user enlarged and/or re-focused views of graphics or images. Indeed, different types of physical lenses could be modeled (Fig. 11.9). Visualization lenses are not limited to enlarging the objects that are being viewed; different attributes of the objects in the display can be presented when the lens is positioned over them. Once users have found an area of interest in an information display, they may want to examine other attributes of that area of interest. such as focusing on Madrid in an interactive map, The same system could show other properties such as its population (9.10.5). As we will see later, a lens can also be used to extend the physical analog. 3-D lenses could show internal structures much as an X-ray does (11.10.1).

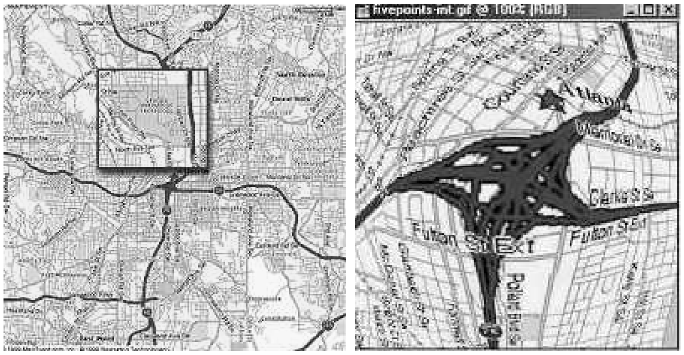


Figure 11.9: Bi-focal lens (left) and fisheye lens (right) for viewing details on a map of Atlanta[39]. (ACM check permission)

11.3. Audio

11.3.1. Sounds

Speech, music, heart beats, sighs, bird song, water dripping, glass breaking — we live in a world of sound. It is good then, that human beings are well suited to it. Not only can we hear a vast range of sounds (from 20 to 20,000 Hz), but our brains also learn to determine what sounds are appropriate for a particular environment. We use sound to communicate with speech, but also use sounds to represent general ideas, such as the sound of a sweeping broom to indicate that a file has been thrown in the "recycle bin". Audio is a very flexible medium; rarely do we use sound in isolation. Based on the biology of human beings, there may in fact be attentional and cognitive implications to modern multimedia applications. In this chapter, we will explore the many aspects of audio information: its capture, its storage, its forms, and its processing.

Acoustics.

Audio spectrogram.

Simulated audio environments.

11.3.2. Music and Sonic Arts

There are many types of musical experiences. I may whistle a popular song, listen to a live rock band, or go to the opera. Moreover, each piece of music may be performed in many ways. Music is sound with a rich structure. However, unlike speech the sound is generally not symbolic. Instrumental music does not usually convey information in the sense of helping a person to predict events. Music is

Figure 11.10: Ella Fitzgerald. (check permission)

highly structured at many levels, and that structure provides much of the aesthetic pleasure we have in listening to it. However, overly structured music can be tedious and some unpredictability is needed. Music comes in many genres from jazz to classical. Music is social. Music genre and mood detection. Musical production is often a social effort. Though, increasingly "music is a thing", a fixed recording [15]. The technology dramatically affect the content and the usage. The development of recording made music much more accessible to the masses and led to mass-market culture. Moreover, increasingly, music is integrated with information systems. Music and speech analysis use similar methods.

The Structure of Music

Music can be considered structured sound; some music is highly stylized, such as Western classical music, while other music is more free-formed, such as some jazz music. The structure of music can be compared to a building: the beat and rhythm provides the foundation with other sounds layered on top. Music can create listener involvement through theme and variation. A theme or structure is established within a piece and that theme is then varied to create a tension between a listener's expectations and their sense of novelty (4.6.2).

There are many other ways in which music is structured. Often, the structure of the music determines its style; it is its structure that defines it. In some cases, the structure of music may become so rigid that it may be modeled with a grammar (6.5.1)[12]. Fig. 11.11 shows a repetitive pattern that can be diagrammed like a sentence. Musical structure based on mathematical principles. While maintaining structure is important, it certainly does not have to be slavishly followed; deletion and variation from structure are essential aspects of art (4.6.2).

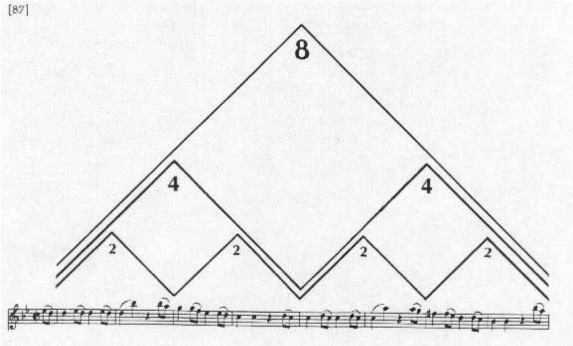


Figure 11.11: Music can highly structured. In some cases, that structure can be represented by regular expressions and grammars. Structure in music is particularly apparent in classical music. Observe the repeating pattern in the notes across several levels[12]. (check permission)

Representation and Markup for Music

Music can be as simple as a verse of "Happy Birthday" or as complex as Handel's Messiah with a 100-voice choir at Christmas. Thus, any representation of music must be scalable to different complexities. While a representation may include each of the 100 voices separately, this variety would require 100 different pieces for one single work. A representation such as that may be useful for a performer, but not so useful for a conductor. Similarly, it may not be necessary to represent all 100 voices to store the work and retrieve it later. Musical performance is dynamic, and not every detail of a piece of music needs to be defined. Instead, it is often preferable to represent the theme, or baseline of the music, and allow individual interpreters to scale it as they may.

Music markup and metadata.

CSound. Composition. Algorithmic and probabilistic approaches to music generation.

Musical Instrument Interfaces

Novel interfaces. KBow.

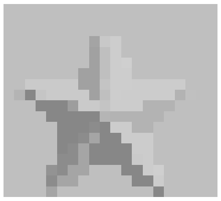


Figure 11.12: Musical interface controllers.

Music Communities

Social music.

Music Indexing and Retrieval

Music libraries.

Music metadata.

Query-by-humming.

11.3.3. Speech and Speech Processing

Speech is interwoven with language; speech communicates natural language using sound. There are many differences between natural language as expressed through text and speech. When a speaker can be seen, many factors, such as the speaker's eye contact and gestures, convey meaning and may be analyzed in conjunction with speech (11.4.1). Even purely verbal nuances, such as cadence, emphasis, and volume contribute to the meaning of speech.

We will look at elements of human speech, from the physical aspects, to linguistic theory, and on to certain speech information services. People don't always use speech in auditory communication; rather, we may communicate with a grunt or even a Bronx cheer.

Phonology and Phonemes

Phonology is the study of the generation and perception of speech sounds. Similar to language itself, speech is a system of small, discrete sound units that are combined to form larger, more complex structures. Phonology studies these basic units in an attempt to better understand not only how we use language, but how and from what it developed.

While many languages use sounds in a way that speakers from other languages are not used to, some languages use sounds that are completely strange to foreign speakers. The trilled r in Spanish as in the Spanish word "rojo" has no analog in English. An even more extreme case is the !kung language from southern Africa (a member of the Khoisan family of languages), which includes a series of unique

Type	Code	Example
Voiced	AA	Bob
	AE	bat
	AH	bought
	AO	boat
	AW	down
	AX	about
	AXR	butter
	AY	buy
	EH	bet
	EL	battle
	EM	bottom
	EN	button
	ER	bird
	EY	bait
	IH	bit
	IX	roses
	IY	beat
	L	let
	M	met
	N	net
	NX	sing
	OW	book
	OY	boy
	UH	book
	UW	boot
	W	wit
	Y	you
	Z	zoo
	R	rent

Type	Code	Example
Voiced-fricatives	DH	that
	JH	judge
	V	vat
	WH	which
	ZH	azure
Unvoiced-fricatives	F	fat
	S	sat
	HH	hat
	TH	thing
	SH	shut
	CH	church
Voiced Stops	B	bet
	D	debt
	H	get
	DX	batter
Unvoiced Stops	K	kit
	P	pet
	T	ten
Glottal Stop	(stop)	

Figure 11.13: Phonemes are the distinct sounds which carry the meaning of words in a language. Here are the phonemes for English as defined by[1].

tongue-click sounds (which are all represented by the ! symbol); this sound is absent from the regular speech patterns of all other known language families on earth.

There are several basic categories of phonemes, such as vowels, liquids, fricatives, plosives, and stops. Each of these phonemes is associated with a particular means by which the sound is produced. Vowels are generally the most distinctive; they are produced by air passing largely unimpeded through the vocal chords and the mouth, tongue and jaw forming tubular, or hollow shapes.

One of the biggest distinctions is between voiced and unvoiced phonemes. Voiced phonemes pass air through both the nasal cavity and mouth. Unvoiced phonemes, on the other hand, pass air only through the mouth. You can test this by holding your hand in front of your mouth and comparing the air movement produced by voiced versus unvoiced phonemes. This difference is why speech sounds strange when a speaker's nose is plugged.

When a person is speaking, the phonemes are not constant and regular. The variants of a phoneme are called "allophones". In part, this is due to the blending that happens when a speaker transitions from one sound to another; this is termed "coarticulation". Many of these differences are produced by the tongue, which has its own trajectories and mechanics that produce different sounds. Regional differences in pronunciation.

Just as morphological analysis determines the meaning units of written words (6.2.1), "morphemes" are the meaning units of spoken words. Such phonological analysis has its limits. Words such as "here" and "hear" are homophones; they sound the same, but are spelled differently and have different meanings. Even if their phonemes are correctly recognized, the particular meaning of a word can be identified only by considering the surrounding context. Lexical semantics was covered in (6.2.3).

Linguistic Markers in Speech

Linguistic markers affect the meaning through sound alone. this could take the form of pronunciation, accents, stress, cadence, intonation, tone, or duration. Fluent speech conveys meaning in what is said, and there how it is said. How something is said is important to its clarity. Emphasis also helps in conversation management. Vocal cues such as uptalk can indicate solidarity or even power relationships.

Prosody is the intonation, rhythm, and stress of speech. It is analogous to orthography (10.1.1)in written text. Prosody can change the nature of a statement. One common example is that questions tend to end in an upward or higher inflection.

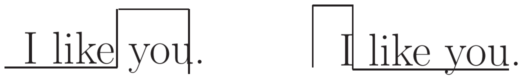


Figure 11.14: Prosody places the emphasis in spoken phrases. On the left, the speaker emphasizes that it is the listener they like, whereas on the right, the emphasis places the focus on the speaker.

Some spoken languages base the meaning of words on inflections which are known as “tones”. These are known as “tonal” languages. Although English has some tonal elements (an example is “too” vs. “to”) it is primarily a-tonal. In these languages, prosody becomes very important.

Inflection and discourse. Deception detection in speech [**?**]. Specifically, there is often a high-pitch from the stress.

Inflection and affect.

There are a variety of individual differences in speech. Indeed, a “voice model” can be developed that represents the characteristic speech sounds of an individual. There are many variables that make up an individual’s speech patterns. As a starting point, women generally have shorter vocal chords than men; their voices are generally of a higher pitch. Another factor may be accent; even within the same language, speakers often develop different “accents” which are characteristic patterns of speech. This effect is heightened when a person is speaking a second language. Because of the widespread use of television and radio, extreme forms of accents are heard much less frequently than they used to be. Socioeconomic and individual differences may also contribute to differences in speech patterns; these effects may be manifested in particular word choices and cadences. One example is “up-talk,” in which all statements sound like questions with a rising pitch.

Diction.

Disfluencies are hesitations and mistakes in speech. Many different types. They can be forced with tongue twisters.

Spoken Language

Speech processing has a wide range of possible applications, from live audio streaming, to speaker identification, to administrative and business uses. These technologies are only recently beginning to become widespread. As computer processing power and our understanding of human speech grows, there is no doubt speech processing will play a larger part in our day-to-day lives.

The top panel of Fig. 11.15 shows the raw speech waves for the phrase “Every salt breeze comes from the sea”. In the lower portion, the amplitudes are converted to frequencies to show a spectral analysis. The very dark bands of high-energy at the bottom of the display are the “formants” of that phrase. The formants are the most distinctive components of phonemes so that identifying them will improve speech recognition.

In particular, formants indicate phonemes. Phonemes are the building blocks of the sounds from which words are constructed. The sounds of the vowels in the words “bad” and “bed” clearly indicate a

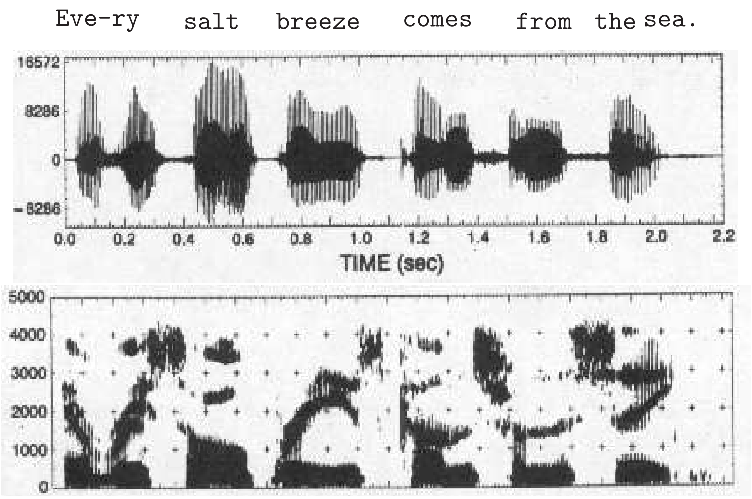


Figure 11.15: Raw speech amplitudes (top), converted to frequency spectrogram with spurious frequencies removed, the formants are clearly visible (bottom)[33]. (add text) (check permission)

difference in meaning. The basic speech sounds that make up a spoken language are categorized into phonemes. The commonly accepted English phonemes are listed in Fig. 11.13. The word "multimedia" is composed of the phonemes M-UH-L-T-AY-M-IY-D-IY-AE. Each language has its own sets of phonemes that define the way that language is spoken, even though the majority of phonemes are consistent across languages. Important distinctions among sounds from one language may be ignored in other languages; thus, there is no absolute set of phonemes.

Phonemes in spoken form sequence of states. These can be described with Hidden Markov models.

Orality

Orality is the use and understanding of spoken language. Spoken language is quite different than written language (e.g.,[30]). Some of the differences may be due to human cognition that process written text visually and spoken text through sound. Other differences are likely due to the environments in which each medium is used. Oral material generally uses shorter sentences and is less well constructed. Communication by text-based electronic media has an oral texture perhaps because it is informal and transitory. greater shared context between speakers in spoken language and conversation, both because their participants may have more of a history (even a very brief one) than in anonymous written communication, and because body language and gesture play such a large role in our understanding of one another.

Modes of Oral Interaction Story telling. Preservation of stories and traditions.

Oratory. Poetry.

Oral histories enhanced with annotated touch screens showing information resources. Narrative (6.3.6). History (5.13.0). People are not very good at understanding and describing their own behavior. Self-attribution and delf-reports can be unreliable (5.5.1).

Autobiography (5.13.3). Story corps. Great speeches. The advantages of oral information not being recorded. Oral arguments need to compensate for that. Oral documents.

Oral cultures. Epic poetry as an oral medium. Indeed, it provides cultural memory. Polynesian navigation. Inuit maps of the coastline.

Cognitive Effects of Language Use

We have already discussed human language learning as well as computer models for text generation

Figure 11.16: Homer, who is credited with writing *The Odyssey* and *The Iliad*, was an oral poet. Campfire (center). Polynesian navigation song. (check permission).

(10.4.3). In addition, we have cognitive processes in reading (10.2.0)and writing (10.3.1). Cognitive effects of bilinugualism.

Voice Applications

Searching podcasts. Indexing radio programs.

Personalized speech synthesis.

Speech Recognition

Automatic Speech Recognition (ASR) attempts to identify speech elements — phonemes — and match them to words. This is increasingly effective but this can be difficult; for instance, compare the sounds of the spoken phrases in Fig. 11.17; the sounds are very similar, but the meanings are quite different.

I scream	Ice cream
Wreck a nice speech	recognize speech

Figure 11.17: Some passages that are particularly difficult for speech processing systems to distinguish.

In addition, people differ greatly in their speech patterns and pronunciation. An important element of speech recognition systems is the training that the people using them may require. An individual with a need for continued and prolonged use of a speech recognition program, such as a person without the ability to type, may use what is known as a "speaker-dependent system". These systems are trained to individuals. With time and training, these systems generally develop a high degree of accuracy. More general ASR systems are known as "Speaker-independent systems," and apply to all users. These require no training, but are generally not as accurate.

Recognition of conversational speech is also much more difficult than recognition of prepared speech (e.g., news broadcasts). Whereas prepared speeches are usually composed of complete sentences, whole words, and — if delivered properly — lack stutters and miscues, conversations are usually completely the opposite. It is difficult for a speech recognition program to understand many of the half-words and colloquialisms typically used in conversation.

Another dimension of speech recognition systems is the differences between isolated-word recognition schemes and continuous-word recognition schemes. Isolated-word recognition seeks to simply identify the word being spoken, usually by first identifying its phonemes. Continuous-word recognition, on the other hand, seeks to understand and identify words using not only their phonemes, but also the context in which they are used, partly basing the definition of a given word on what has come before it.

Phoneme Recognition

As the basic elements of speech, phonemes also constitute the most fundamental units that can be processed by speech recognition software (11.3.3). A high degree of phoneme recognition is necessary to both isolated- and continuous-word recognition programs. Obviously, this process begins by identifying a word's constituent phonemes, or by segmenting and categorizing them. As with speech recognition

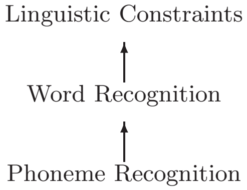


Figure 11.18: Stages in a simple a bottom-up spoken language processing system.

as a whole, phoneme recognition and segmentation processes can proceed by feature matching alone or by a combination of features-matching and templates (1.4.4), or models. The differences between these two methods are similar the differences between "bottom-up" and "top-down" processing, as used in other recognition processes (1.4.4).

When the combined feature matching and templates are used, Hidden Markov Models can prove particularly useful as they recursively calculate the identification probability as each new phoneme is recorded by the recognition program (-A.5.5). These prediction models can determine the probability that a particular phoneme matches a particular template, and revise that probability as new information (more input) is added. While there may not usually be an exact fit to the template, Hidden Markov Models help to select the match. Other methods for clarifying and increasing the accuracy of word recognition are through the integrated use of multimedia. In addition to speech many visual cues can supplement speech understanding.

Fig. 11.19 shows lip positions for different words. The lip positions are the result of producing phonemes (11.3.3). Image processing can be used to identify lip positions and these can augment the processing of the sounds. Other methods that increase the accuracy of word-recognition programs include gesture annotation (11.4.1) and social context, which analyzes gaze and pose (5.6.5).

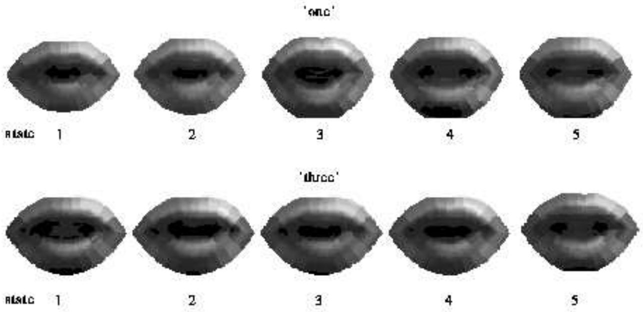


Figure 11.19: Vision can enhance speech processing. Lip positions for the word "one" in the upper row and the word "three" in the lower row[19]. (check permission)

Word Recognition

Acoustic models. A spoken word can be characterized as a sequence of phonemes. How is a program to move from recognizing a phoneme, or series of phonemes, to recognizing an entire word? From a purely numeric approach, if every phoneme were recognizable then every word could be tabulated according to its included phonemes. Certain elements of spoken language, such as alliteration, could confuse such a system.

It can be difficult, even for human listeners, to differentiate between two similar sounding words; the same is true of word and speech recognition programs. One method of differentiation is to use inflection as a type of punctuation (11.3.3) to contrast similar sounding syllables.

Developing Word Models with HMMs A spoken word may be modeled as a sequence of phonemes. Word models. Language models and word sequences. Models of words and matching them.

Markov Chains and Stochastic Finite State Machines are also weighted automata. These may be adapted as Markov models. This is helpful if we observe a process that we feel is not random, but we are not sure what the pattern is. The pattern, or model, is "hidden" and must be inferred. To do this, we chart the observations (the process) with a Markov Model and then apply that model to whatever end we need, such as speech recognition. Hidden Markov Models (HMMs) are very important for many applications such as speech and gesture recognition (-A.5.5).

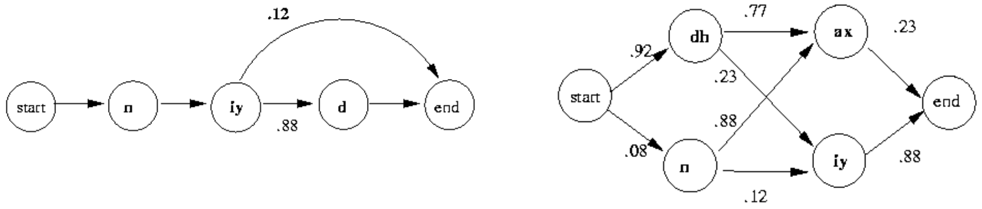


Figure 11.20: Words may be described as strings of phonemes. Word-level Hidden Markov Models for pronunciation of "need" (left) and "the" (right)[28]. Notice that phonemes may be skipped as with the "d" in need and that several related phonemes may be substituted.

11.3.4. Audio Search

11.4. Action and Behavior

11.4.1. Gestures

Motion is essentially changing position in space through time. The thing that extends video beyond the media we examined in earlier chapters is motion. Descriptions of behavior are useful for animations and multimodal inputs. Motion recognition is also integral to many types of interactivity. Motion ranges from simple and regular motions to complex motions that are semantically meaningful and that communicate intentional patterns of action. From actions to intentions. Beyond hand gestures. Facial expressions. Behavior monitoring and observation. Motion pattern analysis. Action as part of tasks. Gesture input (4.2.4).

Modeling typical human behavior and activities from large numbers of photographs on the web.

Gestures are behaviors which convey meaning directly or are used in conjunction with other types of communication. They are especially associated with speech, Fig. 11.21 illustrates the use of a metaphoric gesture. Fig. 11.22 shows one system of categories for gestures[29]. Gestures for framing a space. Gestures as rhetorical device. Expressing emotion with gestures[3].

Gestures as related to visual languages [?] indeed, they may be predecessor to spoken language. Gesture has meaning units analogous to phonemes.

Sign language (6.1.2).

Gestural interfaces. WII. Kinect. For instance, gestures can be used for musical performance [?].

Kinematic gesture corpus[4].

Generating gestures and facial expressions for conversational agents.

Gesture recognition.

Gestural play. Wii (4.2.4).

11.4.2. Formal Models for Action

11.4.3. Visual Tracking

We may want to follow an object or person as it moves through a scene [?]. Visual tracking combines aspects of motion analysis and object recognition. For this we can use tools such as spatial position,

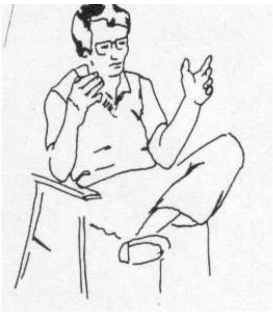


Figure 11.21: One example of a gesture is a “conduit metaphor”. The hands indicate space along which a sequence of objects is located.)[29]. (redraw-K) (check permission)

Type	Description or Example
Iconic	“OK” sign with fingers resembles the letters “OK”
Metaphoric	Abstract metaphor such as using hands to show containment
Beats	Rhythmic actions, often synchronized with speech
Deictics	Pointing
Cohesives	Indicating that ideas are tied together
Emblems	Specific actions that have acquired a meaning of their own

Figure 11.22: Categorization of gestures that are coordinated with speech (based on[29]).

sound localization and multimodal tracking. Projecting trajectories, understanding physical processes. Plan recognition (3.7.2).

11.5. Performance

A performance is an ensemble of actions. Theater. Opera. First person games. as a type of performance. Enactment.

11.5.1. Dance

Dance, like music, is highly structured. Dance shows physical aspects of emotions, social interaction, even communicative gestures. It has regularities at several levels: in the movement of an individual, in the group of individuals on stage, and across a composition. Dance can be expressive. Fig. 11.23 shows an example of Labanotation, which describe ballet movements; however, this is not a full language. Dance emotion as expressiveness and conveying meaning. Labanotation is like a musical score. Dancing may be described with high-dimensional grammars[41].

Creating a dance composition is choreography. Creating a story with dance.

Non-western dance.

11.5.2. Theater

Stage directions. Scene descriptions.

11.5.3. Cyber-Drama

Dynamic story telling. Increasingly, stroes are interactive and immersive. The concept is illustrated by the holodeck from StarTrek (Fig. 11.25). Beyond interactive theater and cinema. Procedural rhetoric. Branching story graph.

Player model.

Game generation.

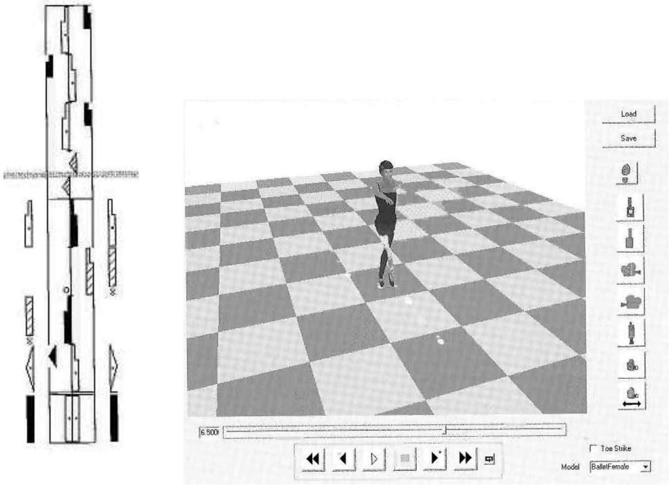


Figure 11.23: Labanotation is a representation for describing ballet movements (left) and is used to generate the animated figure on the right[7]. (check permission)

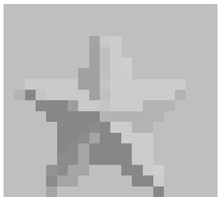


Figure 11.24: Managing theater.

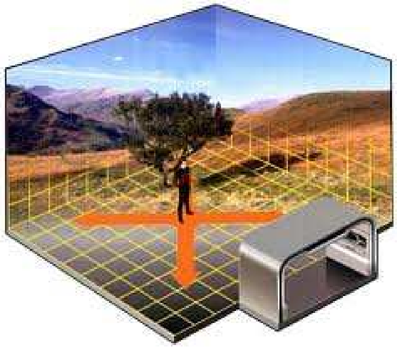


Figure 11.25: The holodeck is a fictional virtual reality environment for creating a personalized story. (check permission)

11.5.4. Drama and Narrative Theory

Aristotle again. Poetics. Poetics versus rhetoric (6.3.5). Experiencing drama as a kind of rehearsal for real life. Tragedy as catharsis.

Narrative is recounted while drama is enacted. (6.3.6). Narratology as a theory of narrative. Artistole's Poetics and drama. Hypertextual fiction. Does interactive fiction fit the Poetics? The interactive story, Facade [?] allows players to interact with and attempt reconcile the relationship between two characters in a story space.

Drama management (11.7.2). Cyber-drama (11.5.3). Experience management. Virtual theater (**??**).

Story is a lot of causal relationships.o

Figure 11.26: Facade interactive game. (check permission)

11.6. Active Visual Media

11.6.1. From Video to Active Visual Media

Active Visual Media is distinguished from static visual media, very simply, by its use of moving pictures or images. For a long time — since Louis Lumiere invented the first widely accessible and practical motion picture camera and projector in 1895 — television and film were the only common examples of active visual media. Over the years, however, a wide range of formats have been introduced, ranging from digital video recording to computer simulation.

Structured and unstructured video.

Active visual media play an increasingly large role in our lives. The continuing development of better, faster, smaller, and more reliable information transmission systems, infrastructure, and devices has had a significant impact on the shape and consciousness of our society. The ever-increasing number of video cameras in the hands of individual citizens led to a bystander being able to record police beating a suspect, thus sparking the Rodney King case (Fig. 11.27); satellite transmission of video allows news footage, and even live war coverage, to be beamed into the homes of viewers half a world away — instances and technologies such as these will continue to influence the development of our society. Now, cameras on every cellphone.

Figure 11.27: The arrest of Rodney King, a Los Angeles motorist, was captured on video by a bystander. It became the basis of a controversial court case. (YouTube example) (check permission)

Increasingly, these active visual media are being augmented by information processing to create hypermedia formats: Media environments composed of different modes, formats, and technologies. Similar to hypertexts, in which embedded links can take a user to content outside the original document, hypermedia applications allow users to do the same thing with active visual media. Interactivity disrupts the continuity of narrative in video and film production. Electronic games and such story-telling media will remain distinct.

11.6.2. Video Retrieval, Processing, Formats, and Libraries

TV Anytime, PBCore.

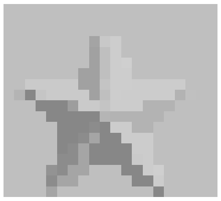


Figure 11.28: Home movies. (check permission)

Video summaries.

MPEG-7 is a standard for describing multimedia objects. Just as text documents are marked up following the DTDs of the Text Encoding Initiative (2.3.3), video can be marked up. For instance, news programs and football games are highly structured and could be marked up. Many standards are being developed for describing the contents of videos. MPEG-7 is XML-based. Descriptors and Description Schemas can be part of the mark-up; one can have a record for each scene and shot (Fig. 11.29). Video documents.

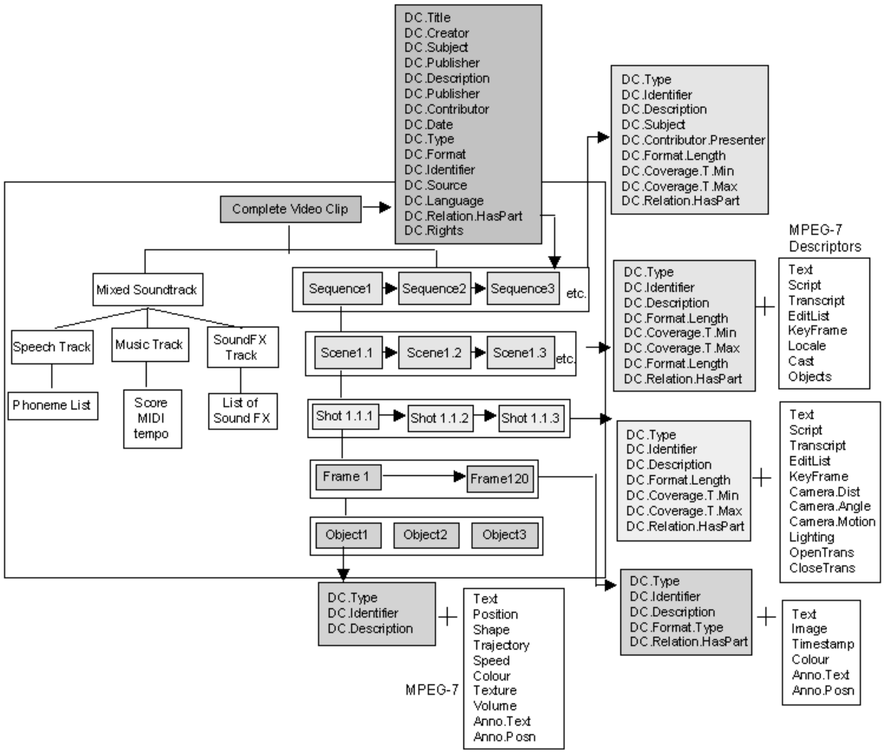


Figure 11.29: MPEG-7 Description Schema[21]. Note the separate metadata that is kept for frames, shots, scenes, and sequences. (check permission)

Video summarization.

Semantic Annotation of Video

Semantic annotation (7.8.4). For complex multimedia such as video, we may have a human observer describe the content.

Inference for enriched indexing of the video.

Video Processing

Combining motion data from several cameras.

11.6.3. Movie Animation

Animation languages. Virtual camera (POV).

Motion capture. What creates an impression of human-ness. Fig. 11.30 Uncanny valley. Quasi-human. Evolutionary hypothesis. Evolutionary caution. (**??**).

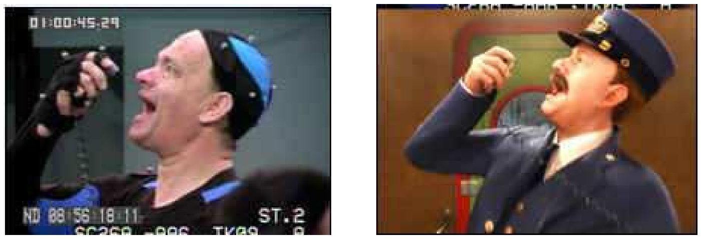


Figure 11.30: Motion capture of Tom Hanks from the movie *Polar Express*. This unsettled some viewers because it fell into the uncanny valley. (check permission)

Types of human motion: voluntary, involuntary, social. Human emulation. Avatars (11.10.3). Clothing. Hair (e.g., Fig. 11.31). Hair is difficult because each strand has a complex interaction with all the other strands.

Figure 11.31: *The Incredibles* was particularly noted for its advances in the simulation of human hair. (check permission)

Digital Cinema

Film Editing

Montage: Narrative and High-Level Structure. A movie, obviously, is more than a series of shots and scenes: it tells a story. Each part of a film is intended to work together to create a unified, final whole. Equally obvious, is that some films do this better than others. The movies that successfully create this final, unified whole have usually managed to make a multimedia presentation; it is not simply a single actor or a good story that makes them good, but the way that they integrate these and other elements to create a mood, or cohesive overall effect. Many of these things often go unnoticed by viewers at the time of viewing, like the way varying theme music may be associated with specific characters, or the way that distinct individual narrative components relate to the work as a whole (6.3.6). Through things such as "establishing shots" that provide context, pacing that matches the film's overall feeling, the director's portrayal of space, architecture and geography, and the depiction of time passing, a successful, cohesive film will monopolize a viewer's attention and create an immersion effect.

11.7. Play and Games

11.7.1. Play

Specific time and place. For fun.

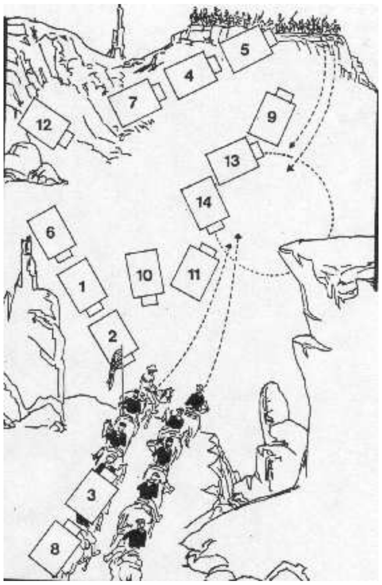


Figure 11.32: The sequence of movie shots for two groups approaching each other is designed to shift attention from one group and the other and to highlight the tension between them (from[9]). (re-draw-K) (check permissions)

11.7.2. Variety of Games

Games and constraints. Taxonomy of games. Games are interactions with structured environment which often resemble natural environments but which have consequences less extreme than the natural world. As we have observed, there are many types of games (2.1.3). Social aspects of games. Role-playing games. Comparing games to sports to play.

Because games mimic aspects of the world, they often reinforce culture.

Scoring.

Business models for games include purchasing good for participation in the game rather than subscription to the game.

Retention. Social

Virtual environments like Second Life. Adding user generated content to virtual environments and games.

Many aspects of game design beyond technical issues. Participatory design for games.

Player model for developing adaptive game engines.

Games typically have few direct real-world implications. Indeed, they may be practice for real-world tasks. One exception is the effort to develop serious games, Those which can be seen as models of complex interactions and are educational.

Games as facilitating experience and enjoyment.

Effects of games on the players. Emotionteering. Narratives (6.3.6).

Games also raise many issues we have seem about other types of information systems such as business models, security, and censorship.

As we discussed back in Section 2.1.3, there does not seem to be any simple set of attributes which define a game. Genres (6.3.7).

Indeed, the definition for games seems to be based on the lack of clear motivation. That is, a game is an activity which does not seem to accomplish a task. Presumably, games, like passive entertainment, meet other human needs such as sociability, exercise, and arousal.

Rules are known. The games are winnable.

One of the most rapidly evolving and expanding hypermedia applications is game design. There is a huge variety in the types of games.

Another type of market.

Principles of effective game design. Make it clear to what the choices are. Scaffolding keeps them involved in a game.

Massively multiplayer games. Organizational skills for developing Guilds.

Many games, however, are educational and teach necessary, or at least useful, life skills; in this sense, games are more than mere entertainment. Whatever their larger definition or significance may be, games are human activities that are structured in such a way that any actions taken within the context of the game produce only very limited real-world consequences. Fig. **??** shows some dimensions of games. At what point is a game different from an interactive simulation.

While computer, or technology-based games are the focus of this chapter, all games — from sport to simulation — have a lot in common. All games simulate reality to some degree, though often metaphorically. All games, have rules (9.5.0), though games are typically much more highly structured. Games are typically characterized by interactivity and by a lack of task.

Why are games absorbing? Physical stimulation. mental challenge. Cyberdrama is a combination of story and game. Rhetoric and procedurality. Immersion, agency, Transformation [**?**].

Games offer insights into both overall human behavior and local or regional culture. Real-world behaviors and customs are often expressed or represented in the rules or structure of games; his similarity has often been noted[35]. In addition to the explicit rules of games, there are often constructive rules or implicit rules. The various implicit and explicit rules of a game have many purposes, themselves both explicit and implicit. These may include the enhancement of the entertainment value of a game; abstract or theoretical intellectual enhancement (5.11.0) as in game theory and strategy (3.4.1); the emotional development that comes from competition, surprise, challenge, teamwork, and even violence (5.9.4); and cultural education that results from the history of a particular game or its narrative (6.3.6).

Puzzles and Riddles

Tetris (Fig. 11.33).

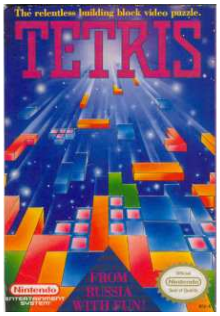

The relentless building block video puzzle.
TETRIS
Nintendo
ENTERTAINMENT SYSTEM
FROM RUSSIA WITH FUN!
Official Nintendo Seal of Quality

Figure 11.33: Tetris. (check permission)

Interactive Stories

Cyber-drama (11.5.3). Drama management shares many aspects with intelligent tutoring systems (ITS) (5.11.3) and adaptive hypertext. Experience management.

Point of view. First-person and third-person.

Team communication in games.

Stories and character development for games[36].

Augmented reality games. Overlay games on to the natural world.

Serious and Persuasive Games

Motivating aspects of games. Games can improve training. How effective are they for education? Management games. Related to education (5.11.5).

Game space and game trees (-A.3.2).

Figure 11.34: Guitar Hero. (check permission)

11.7.3. Games as Information Resources

Board games often have a set of rules and physical playing environment. Electronic games have implicit rules and control the playing environment. Either way, games can be seen as information resources. Thus, they can be indexed and organized. For instance, they can be assigned metadata.

Game genres. Designing experiences.

Video recording of a person playing the game.

What does it mean to preserve MMORGS where the interaction with other players is the key.

State graph for game states (-A.3.2).

Player modeling is related to student models (5.11.3)and other user models (4.10.2),

Graphical rendering.

Games as Cultural Works

Games as an art form. Preservation of games as cultural memory (5.9.3). Game culture evolves really fast.

Descriptions of how games play. Event capture for games.

Each game instance may be considered a distinctive work since it evolves as it is played.

Fan blogs. “Who owns the game?” The players or the game company? Socialization. Preservation of games[27]. Killing characters in MMPG In some cases, games include game mods. That is, players can develop extensions to them.

Game Users

Video game violence (5.9.4).

11.8. 3-D Images and Solid Models

11.8.1. 3-D Images

The development of representation techniques for 3-D objects is an important advancement in information systems. This allows digital objects to be represented in a way that more closely approximates how we already see the world, and how our minds understand it. This can lead to a more natural interface for humans and computers. There are many elements just on the human end of developing accurate 3-D representations, including the physical principles of vision (4.2.3) and 3-D visual perception. Depth perception in particular is critical to developing a 3-D experience. There are many applications for even simple 3-D representations, ranging from art to engineering, and many that have yet to be imagined; that the computing power to develop these objects is only recently becoming readily available is something that has surely limited the scope of the possibilities.

Extending CAD models by adding behavior and by incorporating simulations (9.5.0).

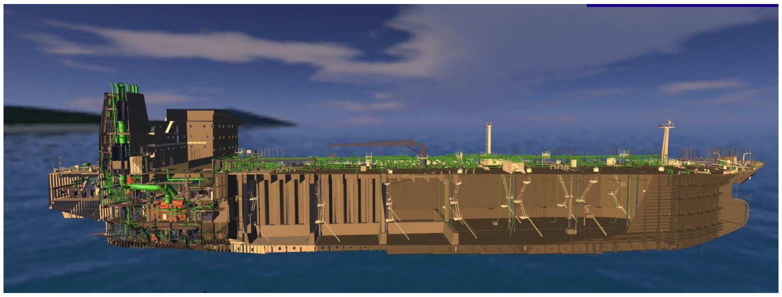


Figure 11.35: Supertanker graphic applies a multi-layer model so the viewer can interact with the model at several levels of granularity[6]. (check permission)

Design specification (3.8.4). CAD for design, for Manufacturing, for Use, for Assembly.

3-D printer technology (8.12.1).

11.8.2. Solid Models

3-D Modeling.

A voxel is the name given to a 3-D pixel.

Perhaps the most widespread and established technique for developing a 3-D representation of an object is Computer Aided Design, or CAD (3.8.0). CAD-designed objects are solid models ((sec:solidmodels)), in the sense that they have shape, size, and dimension, and can be viewed from all angles, the same as any solid object. Engineers, draftsman, designers, and architects have been using CAD for more than twenty years to model 3-dimensional structures or objects. Over that time, the power and capabilities of CAD programs have increased dramatically. Fig. 11.35 shows a CAD wire-frame model. We have already considered 3-D perception and representation for people (11.8.2).

3-D representation's are fundamental for people interacting with the world. Thus, we focus on them here (Fig. 4.9). One theory suggests that 3D perceptual representations are based on simple volume-filling shapes which are processed at the pre-attentive stage[13]. Fig. 11.36 shows how such geons can describe the shape of an airplane.

11.9. Wearables, Tangibles, and Smart Environments

11.9.1. Wearables

Embodiment.

Tracking personal health.

Personalization of fashion selections.

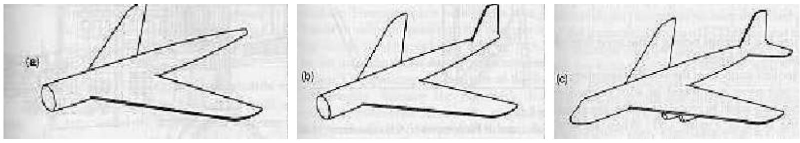


Figure 11.36: One approach to representation of complex 3D objects develops a complex representation from simple shapes called geons. Here, for instance, an airplane is built up from such simple shapes[13]. (redraw with car) (check permission)

Wearable technology. Glasses, watches.

Presentation of the self (5.5.1). Virtual fitting room.

Conductive fabric.

Monitor wearer. Sensors.

Augmented reality ((sec:AR)).

11.9.2. Tangibles

Objects in the world. Toys. Fig. 11.37. Types of tangibles. Remote telepresence with haptics. Marble answering machine with token encoding actions. Touch surfaces.

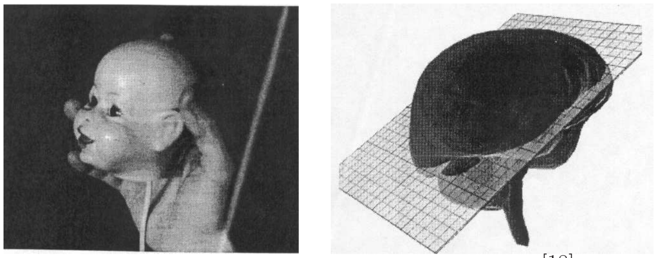


Figure 11.37: A tool for manipulating the location of a cross-section of the brain[18]. Moving the plastic head changes the position of the cutting plane and, ultimately, the view of a brain section. (check permission)

11.9.3. Ambient Design

Constructing environments with electronic displays.

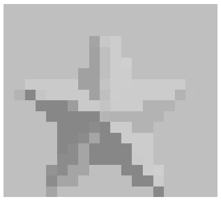


Figure 11.38: Ambient design.

11.9.4. Internet of Things

Location management. Inventories. RFID.

Supply chain (8.12.1)..

11.9.5. Active Environments

Sensors in the environment.

11.10. Reality and Beyond

11.10.1. Mixed and Augmented Realities

While a virtual reality environment can be entirely fabricated, it is more common that virtual reality elements are coordinated with the natural world. We can call these environments “mixed realities”. The uses and applications of mixed realities are practically unlimited; whole libraries of virtual objects can be assembled to facilitate the creation of augmented realities. This helps to open up many ways to augment reality. Hidden structural supports (figure). View management.

Media space ((sec:mediaspace)).

Metaverse. Sensors and GIS and lifelogging.

Audio augmented reality. Visual augmented reality. Cellphones for supporting augmented realities.

Difficulties of rendering complex environments for virtual realities with little lag.

11.10.2. Virtual Reality, Virtual Environments, and Virtual Worlds

CVEs (5.6.6).

Second Life. Project Wonderland.

Virtual economies (5.2.2). Governance of virtual worlds[16]. Validation of virtual worlds. What aspects make them seem real.

3-D Navigation (2.6.3).

Difficulty of real-time updates a complex 3D space.

Experimentation in the virtual space.

Virtual Rome (Fig. 11.39).

Figure 11.39: Virtual reconstruction of ancient Rome. (check permission)

Virtual crowds. Crowd dynamics.

Interacting with virtual worlds.

Scripting for virtual worlds.

11.10.3. Animated Characters (Avatars)

Avatars are simulations of living organisms, especially of human beings. The word “avatar” comes from the Hindi word for “spirit” or “essence”. They build on the natural social interaction that comes easily for people. Indeed, they may provide an actual social presence (5.6.5) that has kept human-computer interaction relatively simplistic up to this point. Puppets as avatars.

Properties of avatar interaction. Aura is the region withing which one avatar will interact with another. Focus is the level of awareness of others within that aura. While nimbus

11.10.4. Conversational Agents

Conversation (6.4.0). Whether textual or verbal. Human animation (11.6.3).

Potentially, conversation is an effective user interface. However, full conversational interaction is difficult and success would mean meeting the Turing Test. Distributed agents (7.7.8)

Who initiates a conversation?

Natural language workflow for mobile telephones where no keyboard is available. Yes/No question interaction. Question categorization for question routing. Question routing as a service for sending traffic to an expert (e.g., a doctor).

Large sets of search results are a competitive advantage for search engine companies because they can be used to improve the quality of earch results.

Increasingly interaction is viewed as a conversation. User initiated, system initiated, mixed initiative. Interaction as conversation. System initiative is potentially intrusive. Increasingly, sensors support input devices for interfaces and they are cheap and widely deployed. Simulating gestures in conversation.

Conversational bots.

Ultimately there would be many applications for effective agents who could understand the nuances of conversation. However, there are huge difficulties in effectively understanding meanings.

Turing test.

11.11. Robots and Cyber-Physical Systems

11.11.1. Robotics

Robots of many forms and applications. Humanoid robots. Social robots.

Robots and manufacturing. Robots and employment (8.8.2).

Cooperation and coordination among teams of robots.

These are often inspired by the nature's solution to design challenges.

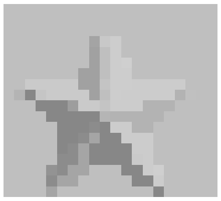


Figure 11.40: Robotic companion.

Robots who know their limitations. Many types of robots. Delivery. Symbiotic robots. Personal robotics. Laws of robotics[10] (Fig. ??).

1.	A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2.	A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
3.	A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

11.11.2. Cyber-Physical System

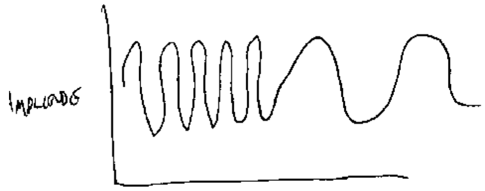
Exercises

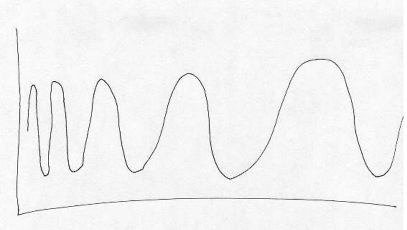
Short Definitions:

Acoustics
Active environment
Agent-based simulation
Animation
Autonomous agent
Augmented reality
Audio contours
Avatar
CAD
Coarticulation
Color depth
Deterministic model
Disfluency
Digital talking book
Dissolve (video production)
Earcon
Focus+Context
Frame differences
Game
Gazetteer (spatial)
Georeferenced data
Gesture
GIS
Information visualization
Inverse kinematics
Hypermedia
Immersive virtual reality
Kinematics
Linguistic markers
Location-aware service
Mixed reality
Modeling
Monte Carlo simulations
Morpheme (speech)
Morphing
Orality
Periodic motion
Phoneme
Pitch
Pixels
Prosody
Range query
Rendering
Route finding
Schematics
Scenes and shots
Scene graph
Semantic zooming
Solid model
Sonification
Sound rendering
Spatial cognition
Spatialization
Spectrogram
Structured Query Language (SQL)
Telepresence
Visual language
Wayfinding
Wearables
Word spotting

Review Questions:

1. What types of material are better presented in an image and better in text. (11.2.1)
2. What are some difficulties for automatic recognition of objects from photographs? (11.2.2)
3. Image processing. ((sec:simpleimageproc), -A.2.3)
4. Thumbnail images are sometimes used as a surrogate for pages in documents. Describe what properties a thumbnail should have to be useful as a surrogate. (10.7.3, (sec:thumbnails))
5. In what sense are visual language like natural languages? (11.2.4)
6. Explain the difference between "information visualization" and "data visualization". (11.2.5)
7. What are the relative advantages and disadvantages of the bifocal and fisheye lens techniques? (11.2.5)
8. Draw the spectrograms for the sound waves shown in these figures (check): (11.3.1)





,

9. What are the characteristics of speech that distinguish it from other types of audio recordings. ((sec:sounds))
10. The optimal sampling rate is twice the highest frequency. CD music is sampled at 44KHz. What is the highest frequency that can be represented on a CD? (11.3.1, (sec:musicalnotes))
11. Give an example of structure in music. (11.3.2)
12. Aside from the words themselves, how do people convey meaning with spoken language? (11.3.3)
13. Pronunciation. (11.3.3)
14. How does the sound of your speaking change if you make a conscious effort to open your mouth more fully as you talk? ((sec:speechproduction))
15. Speech application. (11.3.3)
16. Motion analysis. ((sec:motionanalysis))

17. Give an example of each of the gesture categories described in Fig. 11.22. (11.4.1)
18. Video metadata. (11.6.2)
19. How is multimedia markup different from text markup? ((sec:multimediamarkup))
20. Games. (11.7.0)
21. What are some strategies for exploring data in more than three dimensions ((sec:3-Dvis))
22. What is the value of metaphor for the design of virtual environments. (11.10.2)
23. Distinguish between first-person and third -person viewpoint. Give an example. (11.5.0)
24. Virtual environments. (11.10.2)
25. Distinguish between "virtual reality" and "mixed reality". (11.10.2, 11.10.1)
26. Give some examples of mixed realities. (11.10.1)
27. Avatars. (11.10.3)

Short-Essays and Hand-Worked Problems:

1. Choose a picture that is accompanied by text from a newspaper or magazine and explain whether you believe it is well composed. How does it relate to the text in which it appears? (11.2.1)
2. If you had to develop a system for matching pictures in the newspaper with the text of the stories with which they were associated what would you do? ((sec:simpleimageproc), -A.2.3)
3. Describe how computer graphics and digital photography are merging. Hint: Think about object-level descriptions of images. ((sec:graphics))
4. In what ways are the mechanisms of retrieval from a text collection similar to retrieval from a collection of images? (Hint: Think about representations and queries.) ((sec:sketchinginterface))
5. What types of tasks would be suitable for a sketching interface for image retrieval? ((sec:sketchinginterface))
6. Do you agree with the statement: "A picture is worth 1000 words"? Explain. (11.2.4)
7. How would you apply the Gricean maxims to visualization systems. (6.4.1, 11.2.5)
8. Why is it more difficult to control acoustics at an outdoor concert than in an indoor concert? (11.3.1)
9. Why do you think people like music? Explain your answer. (4.6.2, 11.3.2)
10. Describe some of the problems of attempting to extract the lyrics from the music of a song. (11.3.2)
11. Describe an interface you might build for helping to teach students how to write a specific type of poetry (e.g., sonnets) or song lyrics. (11.3.2, 11.3.2)
12. Describe the components you would need for an audio-only query tune-retrieval system. (11.3.2)
13. Describe some of the difficulties in using tune matching for retrieval. (11.3.2)
14. Speech does not include explicit punctuation marks as does text. What are some of the features of speech that function as punctuation? (11.3.3)
15. Why is the "signal-to-noise-ratio" an important consideration for both music and speech recognition. ((sec:musicrecognition), 11.3.3)
16. Compare template models for visual recognition with template models for phoneme recognition. (11.2.2, 11.3.3)
17. Identify the phonemes in "dog," "cat," "apple," "tree," "comb," "bomb". (11.3.3)
18. What is the phonetic representation of "the Rain in Spain"? (11.3.3)
19. What are the characteristics of radio that make it a popular and economically successful medium. ((sec:radio)).
20. Describe an audio-only interface for accessing Web documents. ((sec:spokencommands))
21. Estimate how many telephone calls are made in the United States each day. Justify your estimate.
22. Estimate how many words are spoken by the world's population in a day. Justify your estimate.
23. Describe an audio collection you would like to develop. Describe the collection management procedures your would use. (7.2.2, 11.3.2)
24. Chose a short video and describe its structure. Create a storyboard for it by sketching the main scenes. ((sec:storyboards))

25. Video metadata. (11.6.2)
26. Suppose you wanted to enter graphical queries to a video retrieval system about the ways in which objects moved across the image itself (some objects might move fast, some slow, some in a smooth motion, and some in an erratic motion, etc.). Describe (and/or sketch) an interface you might develop for entering those queries. (11.6.2)
27. How might you build an interface to examine scenes to be returned by a video retrieval system? ((sec:videosearching))
28. Animation. ((sec:animation))
29. Affine equations. (11.2.2)
30. What aspects of gestures reveal emotion? (4.6.0, 11.4.1)
31. Is a smile a gesture? Why or why not? (11.4.1)

32. Hypermedia representations. (11.1.5)
33. Synchronization of multimedia events. ((sec:finesync))
34. Preservation of interactive systems. (7.5.5, 11.1.5)
35. Characterize an online game in terms of the dimensions in Fig. **??**. (11.7.0)
36. 3-D object calculations. (11.8.1)
37. 3-D projections. ((sec:3-Denvironments))
38. Is "reality" subjective or objective? (11.10.2)
39. Suppose you were going to make a simulation of foot traffic on your campus. Describe how you could use an agent-based simulation. (9.5.0)
40. Suggest a critical question you might ask in a Turing Test. Why do you believe it would be effective? (11.10.4)
41. Describe a scenario for an interactive drama. (11.5.3)
42. Chose a character from history and describe an interaction with them. Describe what you would have to do to get a computer to produce those interactions. ((sec:syntheticinteractions))
43. Are you engaged in "mixed reality" during a telephone call? (11.10.1)
44. Suggest a novel application for wearable computers. (11.9.1)
45. List all computers in your home. Be sure to include those in hand-held games, appliance controls, and communication devices. How many of these computers are able to be directly controlled by users and how many are "embedded"? ((sec:pervasivecomputing))
46. Virtualized reality. ((sec:virtualizedreality))
47. What cues are most important for developing a convincing virtualized reality? What does it mean for people to "suspend disbelief"? ((sec:virtualizedreality))

Going Beyond:

1. Visual object recognition. (11.2.2, 11.2.2)
2. How does "compositionality" apply to visual languages? (1.1.3, 11.2.4)
3. How is visualization related to hypertext? To information needs? (11.2.5)
4. Describe a simulated audio environment you would like to develop. Describe some of the difficulties you would have and how you might overcome them. (11.3.1)
5. (a) Explain the principles of audio beam tracing. Develop a simple implementation of audio beam tracing. ((sec:beamtracing))
6. Music to complement lyrics. (11.3.2)
7. How could a thesaurus improve the recognition rate for a speech processing system? (2.2.2, 11.3.3)
8. Languages which are not written have many fewer words than languages that are written. Why is this? (6.2.1, 11.3.3)
9. What is an appropriate signal-to-noise ratio for preservation of music? ((sec:musicprocessing), -A.9.2)
10. Formants. (11.3.3)
11. Describe how music could be synthesized using a grammar. ((sec:syntheticmusic))
12. Ask two friends to read a paragraph of text into a tape recorder or speech capture system. Examine the samples to attempt to determine how the speech characteristics of your friends differ. (11.3.3)
13. Describe the similarities and differences between typing errors and speech errors. Develop a cognitive model that explains these differences. (4.3.3), 10.3.1, 11.3.3
14. Describe how a search engine for speech might be developed that would index and match the phonemes of speech without text representations. (11.3.3)
15. Propose a cognitive model for comprehension of speech. For instance, are spoken words and textual words simply converted to a common format and then processed the same way? (11.3.3)
16. Describe a business based on speech recognition technology. (11.3.3)
17. Describe some of the social changes we might expect if speech recognition is realized. (11.3.3)
18. Develop a simple HMM for processing speech. (11.3.3, -A.5.5)
19. How important is it to keep track of whether a group of songs appears together on a phonograph album as distinct from simply preserving the recording of the individual songs? (7.5.1, 11.3.2).
20. Describe the requirements for an audio editing system. (7.9.1, (sec:audioediting))
21. Describe some design principles for audio hypertext. ((sec:audiohypertext))
22. Voice mail indexing systems might use distinct speech patterns to extract the return telephone number in a voice mail message. Listen to several voice mail messages and determine some of these phrases. ((sec:voice-mailindexing))
23. Create a storyboard for TV sitcom you have recorded ((sec:storyboards))
24. Describe a language for video events and objects. (11.6.2)
25. Find an online video and develop a multimedia summary of it. (11.6.2)

26. Develop two simple wire-frame displays. Demonstrate morphing from one of the wire-frames to the other.((sec:wireframes), (sec:morphing))
27. Motion analysis. Affine equations. (11.2.2)
28. How do the gestures of television actors compare to the gestures made by ordinary speakers? Why do think there is a difference? (11.4.1)
29. Develop a temporal scripting language. (11.1.5)
30. Hypermedia applications. ((sec:hypermediaapps))
31. Why do relatively few women play video games? (4.9.1, 11.7.0)
32. How could you use hypertext and user models to develop adaptive games? (2.6.1, 4.10.2, 11.7.0)
33. Observe the discussion among the players in a multi-player game. What principles of interaction do you observe? How would this discussion be different if the game were conducted with audio-only communication? (5.6.5, 11.7.0)
34. If games are artificial environments, is it helpful to use them for education? Recall that we argued that education was best when it is situated in the environments to which it applies. (5.11.7, 11.7.0)
35. What standards might you apply to evaluate whether a game is fun? (7.10.2, 11.7.0)
36. Describe an online game based on railroad trains. (11.7.0)
37. 3-D objects, (11.8.1)
38. Virtual reality systems generally emphasize graphical interaction. They do not necessarily give a sense of the cultural assumptions or other contexts associated with an environment. How could these other factors be conveyed? ((sec:3-Denvironments))
39. The term “virtual reality” seems like an oxymoron. What might be a better term? (11.10.2)
40. Mixed and augmented reality, (11.10.1)
41. What are the attributes of effective shared virtual spaces? (11.10.2)

Related Books

- Aldrich, C. *Learning by Doing: A Comprehensive Guide to Simulations, Computer Games, and Pedagogy in e-Learning and Other Educational Experiences.* Pfeiffer (Wiley), New York, 2005.
- Arijon, D. *Grammar of the Film Language.* Silman James Press, Los Angeles, 1976.
- Bailey, T. *Interactive Spatial Data Analysis.* Prentice Hall, New York, 1995.
- Besser, H. *Introduction to Imaging.* Oxford University Press, New York, 2003.
- Bernstein, L. *The Unanswered Question: Six Talks at Harvard.* Harvard University Press, Cambridge MA, 1976.
- Bermudez, J.L., and Marcel, A.J., and Eilan, N. *The Body and The Self.* MIT Press, Cambridge MA, 1998.
- Bosselmann, P. *Representations of Places: Reality and Realism in City Design.* University of California Press, Berkeley CA, 1998.
- Breneman, L.N., and Breneman, B. *Once Upon a Time.* Nelson Hall, Chicago, 1983.
- Castanova, E. *Synthetic Worlds: The Business and Culture of Online Games.* University of Chicago Press, Chicago, 2005
- Chen, C.M. *Information Visualization: Beyond the Horizon*, 2^{nd} ed. Springer, Berlin, 2004.
- Cook, P., (Ed.) *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics.* MIT Press, Cambridge MA, 1998.
- Erard, M., *Um.... Slips, Stumbles, and Verbal Blunders and What They Mean.* Pantheon, New York, 2007.
- Grau, O. *Virtual Art: From Illusion to Immersion.* Leonardo Books, Cambridge MA, 2003.
- Horn, R.E. *Visual Language: Global Communication for the 21^{st} Century.* Macro VU Press, Bainbridge Island WA, 2000.
- Jurafsky, D., and Martin, J.H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Prentice Hall, New York, 2000.
- Kinsler, L.E. *Fundamentals of Acoustics.* Wiley, New York, 2000.
- Kolker, R.P. Prentice-Hall, New York, 1999.
- Kosko, B. *Noise.* Viking/Penguin, New York, 2006.
- Kraak M.J. *Cartography: Visualization of Spatial Data.* Addison-Wesley, Boston, 1996.
- Kress, G.R., and van Leeuwen, T. *Reading Images: The Grammar of Visual Design.* Routledge, London, 1996.
- Levitin, D. *This is Your Brain on Music: The Science of a Human Obsession.* Duton Press, Hialeah FL, 2006.
- McNeil, D. *Hand and Mind: What Gestures Reveal about Thought.* University of Chicago Press, Chicago, 1995.
- Monmonier, M.S. *How to Lie with Maps.* 2^{nd} ed. University of Chicago Press, Chicago, 1996.

- Newcombe, N.S., and Huttenlocher, J. *Making Space: The Development of Spatial Representation and Reasoning.* MIT Press, Cambridge MA, 2000.
- Rabiner, L.R., and Juang, B.H. *Fundamentals of Speech Recognition.* Prentice Hall, New York, 1993.
- Ryan, M.L., *Narratives as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media.* Johns Hopkins University Press, Baltimore, 2001.
- Salen, K., and Zimmerman, E. *Rules of Play.* MIT Press, Cambridge MA, 2004.
- Sheldon, L. *Character Development and Storytelling for Games.* Thomson, Boston, 2004.
- Smith, M.M. *Sensing the Past.* MIT Press, Cambridge MA, 2007.
- Smith, M. *Engaging Characters.* Clarendon Press, Oxford UK, 1995.
- Solso, R.L. *Cognition and the Visual Arts.* MIT Press, Cambridge MA, 1997.
- Tufte, E. *Envisioning Information.*
- Ullman, S. *High-level Vision: Object Recognition and Visual Cognition.* MIT Press, Cambridge MA, 1996.
- Wilford, J.N. *The Mapmakers.* MIT Press, Cambridge MA, 1996.
- Wolf, M.J.P. and Perron. P. (Eds.) *The Video Game Theory Reader.* Routledge, New York, 2003.