17 Multimedia User Interface Design

Alistair Sutcliffe

CONTENTS

17.1 INTRODUCTION

The distinguishing characteristics of multimedia are information-intensive applications that have a complex design space for presenting information to people. Design of multimedia interfaces currently leaves a lot to be desired. As with many technologies, it is the fascination with new devices, functions, and forms of interaction that has motivated design rather than ease of use, or even utility of practical applications. Poor usability limits the effectiveness of multimedia products, which might look good but do not deliver effective use (Scaife et al. 1997). With the growth of the web, use of media has become a vital component of attractive and engaging design.

This chapter describes a design process that starts with an information analysis then progresses to deal with issues of media selection and integration. The background to the method and its evolution with experience can be found in several publications (Sutcliffe and De Angeli 2005; Faraday and Sutcliffe 1996, 1997, 1998; Sutcliffe and Faraday 1994). A more detailed description is given in Sutcliffe (2003). The time-to-market pressure gives little incentive for systematic, principled design, so at first reading, a systematic approach may seem to be counter to the commercial drivers of development. However, I would argue that if multimedia design does not adopt a usability engineering approach, it will fail to deliver effective and usable products.

Traditional multimedia markets have been in education and training, although dialogue in many systems has been restricted to drill and quiz interaction, interactive simulations and microworlds are more effective (Rogers and Scaife 1998). Multimedia has been used extensively in task-based applications in process control and safety-critical systems (Alty 1991; Hollan, Hutchins, and Weitzman 1984). With the advent of the Web 2.0 and beyond, interactive multimedia is a continuing design challenge.

Design for multimedia user interfaces (UIs) expands conventional definitions of usability (e.g., ISO 9241 Part 11: ISO 1997) into five concerns as follows:

1. *Operational usability:* It is the conventional sense of usability that concerns design of graphical user interface (GUI) features such as menus, icons, metaphors, and navigation in hypermedia.
2. *Information delivery:* It is a prime concern for multimedia or any information-intensive application and raises issues of media selection, integration, and design for attention.
3. *Learning*: Training and education are both important markets for multimedia and hence learnability of the product and its content are key quality attributes. However, design of educational technology is a complex subject in its own right, and multimedia is only one part of the design problem.
4. *Utility*: In some applications, this will be the functionality that supports the user's task; in others, information delivery and learning will represent the value perceived by the user.
5. *Engagement and attractiveness*: The attractiveness of multimedia is now a key factor especially for websites. Multimedia interfaces have to attract users and deliver a stimulating user experience, as well as being easy to use and learn.

Multimedia design involves several specialisms, which are technical subjects in their own right. For instance, design of text is the science (or art) of calligraphy that has developed new fonts over many years; visualization design encompasses the creation of images, either drawn or captured as photographs. Design of moving images, cartoons, video, and film are further specializations, as are musical composition and design of sound effects. Multimedia design lies on an interesting cultural boundary between the creative artistic community and science-based engineering. One implication of this cultural collision is that space precludes "within media" design, that is, guidelines for design of one particular medium, being dealt with in depth in this chapter. Successful multimedia design often requires teams of specialists who contribute from their own skill sets (Kristof and Satran 1995; Mullet and Sano 1995).

17.2 DEFINITIONS AND TERMINOLOGY

Multimedia essentially extends the GUI paradigm by providing a richer means of representing information for the user by use of image, video, sound, and speech. The following definitions broadly follow those in the ISO standard 14915 on Multimedia User Interface Design (ISO 1998). The starting point is to ask about the difference between what is perceived by someone and what is stored on a machine.

Communication concepts in multimedia can be separated into the following:

- *Message*: The content of communication between a sender and receiver.
- *Medium (*plural *media)*: The means by which that content is delivered. Note that this is how the message is represented rather than the technology for storing or delivering a message. There is a distinction between perceived media and physical media such as CD-ROM and hard disk.
- *Modality*: The sense by which a message is sent or received by people or machines. This refers to the senses of vision, hearing, touch, smell, and taste.

A message is conveyed by a medium and received through a modality. A modality is the sensory channel that we use to send and receive messages to and from the world, essentially our senses. Two principal modalities are used in human–computer communication as follows:

1. Vision: All information received through our eyes, including text and image-based media
2. Hearing: All information received through our ears, as sound, music, and speech

In the future, as multimedia converges with virtual reality (VR), we will use other modalities more frequently: haptic (sense of touch), kinaesthetic (sense of body posture and balance), gustation (taste), and olfaction (smell). These issues are dealt with in Chapters 19 and 29.

Defining a medium is not simple because it depends on how it was captured in the first place, how it was designed, and how it has been stored. For example, photograph can be taken on film, developed, and then scanned into a computer as a digitized image. The same image may have been captured directly by a digital camera and sent to a computer as an e-mail file. At the physical level, media may be stored by different techniques.

Physical media storage has usability implications for the quality of image and response time in networked multimedia. A screen image with 640×480 VGA resolution using 24 bits per pixel for good color coding gives 921,600 bytes, so at 30 frames per second, 1 second needs around 25 megabytes of memory or disk space. Compression algorithms, for example, MPEG (Moving Pictures Expert Group), reduce this by a factor of 10. Improvements in disc storage have reduced file size concerns; however, physical image media constraints are still important on networks, when bandwidth limits the desired display quality. For example, the low-resolution video on mobile phones is typically transmitted at 15 frames per second at a resolution of 240×320, although high-resolution cameras 1270×780 will be available in the near future. In contrast, Internet video and film on-demand services deliver much higher picture quality, by using intelligent fetch-ahead algorithms, but they need good quality and broadband connections. Sound, in comparison, is less of a problem. Full stereo audio with a complete range of harmonic frequencies consumes only 100 kilobytes for 5 minutes, so there are few technology constraints on high-quality audio.

17.3 COGNITIVE BACKGROUND

The purpose of this section is to give a brief overview of cognitive psychology as it affects multimedia design. More details can be found in Part I, Humans in HCI.

17.3.1 Perception and Comprehension

Generally, our eyes are drawn to moving shapes, then complex, different, and colorful objects. Visual comprehension can be summarized as "what you see depends on what you look at and what you know." Multimedia designers can influence what users look at by controlling attention with display techniques such as use of movement, highlighting, and salient icons. However, designers should be aware that the information people assimilate from an image also depends on their internal motivation, what they want to find, and how well they know the domain (Treisman 1988). A novice will not see interesting plant species in a tropical jungle, whereas a trained botanist will. Selection of visual content, therefore, has to take the user's knowledge and task into account.

Because the visual sense receives information continuously, it gets overwritten in working memory (Baddeley 1986). This means that memorization of visually transmitted information is not always effective unless users are given time to view and comprehend images. Furthermore, users only extract very high-level or "gist" (general sense) information from moving images. Visual information has to be understood by using memory. In realistic images, this process is automatic; however, with nonrealistic images, we have to think carefully about the meaning, for example, to interpret a diagram. Although extraction of information from images is rapid, it does vary according to the complexity of the image and how much we know about the domain. Sound is a transient medium, so unless it is processed quickly, the message can be lost. Even though people are remarkably effective at comprehending spoken language and can interpret other sounds quickly, the audio medium is prone to interference because other sounds can compete with the principal message. Because sound is transient, information in speech will not be assimilated in detail, and so only the gist will be memorized (Gardiner and Christie 1987).

17.3.2 Selective Attention

We can only attend to a limited number of inputs at once. Although people are remarkably good at integrating information received by different senses (e.g., watching a film and listening to the soundtrack), there are limits determined by the psychology of human information processing (Wickens, Sandry, and Vidulich 1983). Our attention is selective and closely related to perception; for instance, we can overhear a conversation in a room with many people speaking (the cocktail party effect). Furthermore, selective attention differs between individuals and can be improved by learning: for example, a conductor can distinguish the different instruments in an orchestra, whereas a typical listener cannot. However, all users have cognitive resource limitations, which means that information delivered on different modalities (e.g., by vision and sound) has to compete for the same resource. For instance, both speech and printed text require a language-understanding resource, whereas video and a still image use image interpretation resources. Cognitive models of information-processing architectures (e.g., interacting cognitive subsystems: Barnard 1985) can show that certain media combinations will not result in effective comprehension because they compete for the same cognitive resources, thus creating a processing bottleneck. We have two main perceptual channels for receiving information: vision and hearing; information going into these channels has to be comprehended before it can be used. Figure 17.1 shows the cognitive architecture of human information processing and resource limitations that lead to multimedia usability problems.

Capacity overflow (1) may happen when too much information is presented in a short period, swamping the user's limited working memory, and cognitive processor's capability to comprehend, chunk, and then memorize or use the information. The connotation is to give users control over the pace of information delivery. Integration problems (2) arise when the message on two media is different, making integration in working memory difficult; this leads to the thematic congruence principle. Contention problems (3) are caused by conflicting attention between dynamic media, and when two inputs compete for the same cognitive resources. For example, speech and text require language understanding. Comprehension (4) is related to congruence; we understand the world by making sense of it with our existing long-term memory. Consequently, if multimedia content is unfamiliar, we cannot make sense of it. Finally, multitasking (5) makes further demands on our cognitive processing, so we will experience difficulty in attending to multimedia input while performing output tasks.

Making clear a theme in a multimedia presentation involves directing the user's reading and viewing sequence across different media segments. Video and speech are processed in sequence, whereas text enforces a serial reading order by the syntactic convention of language. In contrast, viewing image media is less predictable since it depends on the size and complexity of the image, the user's knowledge of the contents, task and motivation (Norman and Shallice 1986), and designed effects for salience. Attention-directing effects can increase the probability that the user will attend to an image component, although no guarantee can be given that a component will be perceived or understood.

17.3.3 Emotion and Arousal

The content of image media in particular can evoke an emotional response, which can be used to promote a more exciting and engaging user experience. These issues are dealt with more extensively in other chapters; for example, the use of human image and speech to persuade users. People treat human photographs, video, and even animated characters with similar social responses as they give to real people, so human image content can be used to increase interest and

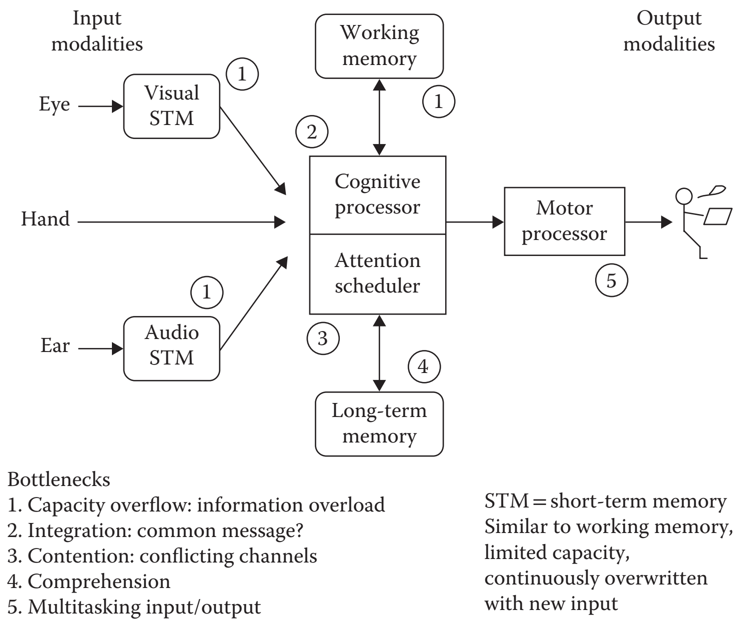

FIGURE 17.1 Approximate model of human information processing using a "human as computer system" analogy, based on the Model Human Processor. For more on cognitive models, see Chapter 2 (Proctor and Vu) and Chapter 5 from the Second Edition of the HCI Handbook (Byrne). (From Card, S. K., T. P. Moran, and A. Newell. *The Psychology of Human Computer Interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1983. With permission.)

draw attention. Emotional responses (see Chapter 4 from the Second Edition of the HCI Handbook) can be invoked not only by content but also by surprise interactive effects, for example, when a character suddenly appears to challenge the users. Surprise effects, moving image, and stimulating images we are not expecting, all affect the arousal system that broadly equates with our feeling of excitement. Designs that stimulate our arousal are more likely to be memorable and engaging.

17.3.4 Learning and Memorization

Learning is the prime objective in tutorial multimedia. In these applications, the objective is to create a rich memory schema, which can be accessed easily in the future. We learn more effectively by active problem solving or learning by doing. This approach is at the heart of constructivist learning theory (Papert 1980), which has connotations for tutorial multimedia. Interactive microworlds where users learn by interacting with simulations, or constructing and testing the simulation, give a more vivid experience that forms better memories (Rogers et al. 1998). Multiple viewpoints help to develop rich schemata by presenting different aspects of the same problem, so the whole concept can be integrated from its parts. An example might be to explain the structure of an engine, then how it operates, and finally display a causal model of why it works. Schema integration during memorization fits the separate viewpoints together.

The implications from psychology are summarized in the form of multimedia design principles (ISO, 14915, Part 3 Media Integration, ISO 1997). The principles are high-level concepts, which are useful for general guidance, but they have to be interpreted in a context to give more specific advice.

- *Thematic congruence*: Messages presented in different media should be linked together to form a coherent whole. This helps comprehension as the different parts of the message make sense by fitting together. Congruence is partly a matter of designing the content so it follows a logical theme, for example, the script or story line makes sense and does not assume too much about the user's domain knowledge; and partly a matter of attentional design to help the user follow the message thread across different media.
- *Manageable information loading*: Messages presented in multimedia should be delivered at a pace which is either under the user's control or at a rate that allows for effective assimilation of information without causing fatigue. The rate of information delivery depends on the quantity and complexity of information in the message, the effectiveness of the design in helping the user extract the message from the media, and the user's domain knowledge and motivation. Some ways of reducing information overload are to avoid excessive use of concurrent dynamic media and give the user time to assimilate complex messages.
- *Ensure compatibility with the user's understanding*: Media should be selected that convey the content in a manner compatible with the user's existing knowledge, for example, the radiation symbol and road sign icons are used to convey hazards and dangers

to users who have the appropriate knowledge and cultural background. The user's ability to understand the message is important for designed image media (diagrams, graphs) when interpretation is dependent on the user's knowledge and background.

- *Complementary viewpoints*: Similar aspects of the same subject matter should be presented on different media to create an integrated whole. Showing different aspects of the same object, for example, picture and design diagram of a ship can help memorization by developing richer schema and better memory cues.
- *Consistency*: It helps users learn an interface by making the controls, command names, and layout follow a familiar pattern. People recognize patterns automatically, so operating the interface becomes an automatic skill. Consistent use of media to deliver messages of a specific type can help by cueing users with what to expect.
- *Interaction and engagement*: They help understanding and learning by encouraging the user to problem solve. Memory is an active process. Interaction increases arousal and this make the user's experience more vivid, exciting, and memorable.
- *Reinforce messages*: Redundant communication of the same message on different media can help learning. Presentation of the same or similar aspects of a message helps memorization by the frequency effect. Exposing users to the same thing in a different modality also promotes rich memory cues.

17.4 DESIGN PROCESS

Multimedia design has to address the problems inherent in the design of any UI, namely, defining user requirements, tasks, and dialogue design; however, there are three issues that concern multimedia specifically:

1. *Matching the media to the message* by selecting and integrating media so the user comprehends the information content effectively.
2. *Managing users' attention* so key items in the content are noticed and understood, and the user follows the message thread across several media.
3. *Interaction and navigation* so the user can access, play, and interact with media in an engaging and predictable manner.

Figure 17.2 gives an overview of the design process that addresses these issues.

The method shown in the figure starts by requirements and information analysis to establish the necessary content and communication goals of the application. It then progresses to domain and user characteristic analysis to establish a profile of the user and the system environment. The output from these stages feeds into media selection and integration, which match the information requirements to available media resources. This is interleaved with interaction design unless the application is restricted to information presentation. Design then progresses to thematic integration of the user's reading/viewing sequence and design to direct the users' attention. Even though the process is described as a sequence, in practice, the stages are interleaved and iterated; however, requirements, information modeling, and media selection should be carried out, even if they are not complete, before subsequent design stages commence.

Design approaches in multimedia tend to be iterative and user-centered. *Storyboards* are a well-known means of informal modeling in multimedia design (Nielsen 1995; Sutcliffe 1999). Originating from animation and cartoon design, storyboards are a set of images that represent key steps in a design. Translated into software, storyboards depict key stages in interaction and are used for conducting walk-throughs to explain what happens at each stage. Allowing the users to edit storyboards and giving them a construction kit to build their own encourages active participation. Storyboards are followed by building concept demonstrators using multimedia authoring tools (e.g., Macromedia Director, Adobe Dreamweaver) to rapidly develop early prototypes. *Concept demonstrators* are active simulations that follow a scenario script of interaction; departure from the preset sequence is not allowed. Several variations can be run to support comparison; however, the user experience is passive. In contrast, users can test *interactive prototypes* by running different commands or functions. The degree of interactivity depends on the implementation cost, which increases as prototypes converge with a fully functional product.

17.4.1 Users, Requirements, and Domains

The starting point for multimedia, as in all applications, is requirements analysis. The difference in multimedia lies in the greater emphasis on information requirements. A variety of analytic approaches can be adopted, such as task analysis (see Chapter 43), contextual inquiry (Chapter 44), or scenario analysis (Chapter 48). Requirements are listed and categorized into information, task-related, and nonfunctional classes. These will be expanded in subsequent analyses.

It is important to get a profile of the target user population to guide media selection. There are three motivations for user analysis:

1. *Choice of modalities*: This is important for people with disabilities, but also for user preferences. Some people prefer verbal-linguistic material over image.
2. *Tuning the content*: This is presented to the level of users' existing knowledge. This is particularly important for training and educational applications.
3. *Capturing the users' expectations*: So the experience can be geared to their background, for example, different styles for younger people, older people, culture, and socioeconomic audiences.

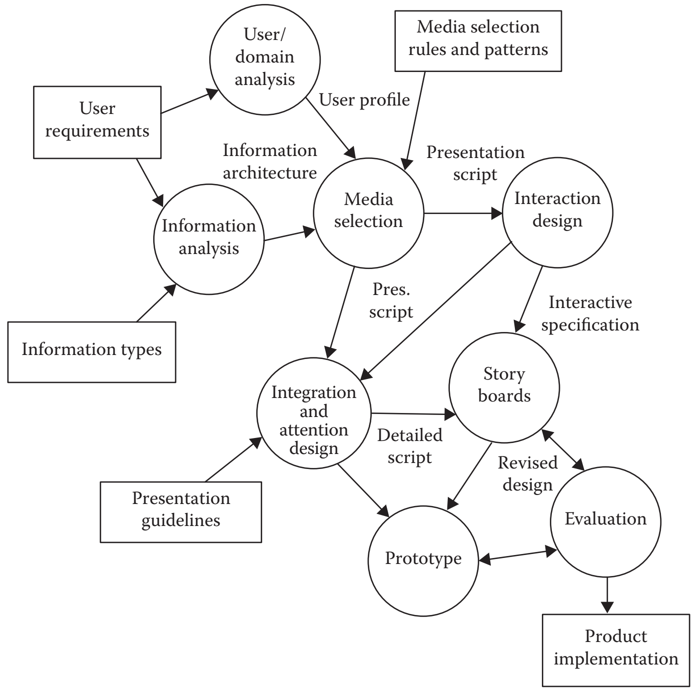

FIGURE 17.2 Overview of the multimedia design process expressed as a data flow diagram.

Acquiring information about the level of experience possessed by the potential user population is important for customization. User profiles are used to design training applications to ensure that the right level of tutorial support is provided, and to assess the users' domain knowledge so that appropriate media can be selected. This is particularly important when symbols, designed images, and diagrams may be involved. The role and background of users will have an important bearing on design. For example, marketing applications will need simple focused content and more aesthetic design, whereas tutorial systems need to deliver detailed content. Information kiosk applications need to provide information, as do task-based applications, but decision-support and persuasive systems (Fogg 1998; see also Chapter 14) also need to ensure that users comprehend and are convinced by messages. Domain knowledge, including use of conventions, symbols, and terminology in the domain, is important because less-experienced users will require more complete information to be presented.

The context and environment of a system will also have an important bearing on design. For example, tourist information systems in outdoor public areas will experience a wide range of lighting conditions, which can make image and text hard to read. High levels of ambient noise in public places or factory floors can make audio and speech useless. Hence, it is important to gather information on the location of use (office, factory floor, public/private space, and hazardous locations), pertinent environmental variables (ambient light, noise levels, and temperature), usage conditions (single user, shared use, broadcast), and expected range of locations (countries, languages, and cultures). Choice of language, icon conventions, interpretation of diagrams, and choice of content all have a bearing on design of international UIs.

As well as gathering general information about the system's context of use, domain modeling can prove useful for creating the system metaphor. Domain models are recorded as sketches of the work environment showing the layout and location of significant objects and artifacts, accompanied by lists of environmental factors. Structural metaphors for organizing information and operational metaphors for controls and devices have their origins in domain analysis.

17.4.2 Information Architecture

This activity consists of several activities that will differ according to the type of application. Some applications might have a strong task model, for instance, a multimedia process control application where the tasks are monitoring a chemical plant, diagnosing problems, and supporting the operator in controlling plant operation. In task-driven applications, information requirements are derived from the task model. In information-provision applications, such as websites with an informative role, information analysis involves categorization and the architecture generally follows a hierarchical model. In the third class of explanatory or thematic applications, analysis is concerned with the story or argument, that is, how the information should be explained or delivered. Educational multimedia and websites with persuasive missions fall into the last category.

In task-driven applications, information needs are annotated on to the task model following a walk-through asking what information the users need to complete the task subgoal, or to take a decision at this step, or to provide as input (see Sutcliffe [1997] for more detail). In information-provision applications, classification of the content according to one of more user views defines the information architecture; for example, most university departments have an information structure with upper-level categories for research, undergraduate courses, postgraduate courses, staff interests, departmental organization, mission and objectives, and so on. For explanatory applications, a theme or story line needs to be developed. This will depend on the application's objectives and the message the owner wishes to deliver. An example thematic map from a health awareness application is illustrated in Figure 17.3.

The requirement is to convince people of the dangers of heart disease. The theme is a persuasive argument that first tries to convince people of the dangers from smoking, poor diet, stressed lifestyles, and so on, then explains how to improve their lifestyle to prevent heart disease, followed by reinforcing the message with the benefits of a healthy lifestyle such as lower health insurance, saving money, longer life. Subthemes are embedded at different points so users can explore the facts behind heart disease, the statistics and their exposure, and how to get help. Information is then gathered for each node in the thematic map. How this architecture will be delivered depends on interaction design decisions: it could become an interactive story to explore different lifestyle choices, combined with a quiz. The outcome of information architecture analysis will be an information-enhanced task model, a thematic map, or a hierarchy/network to show the structure and relationships of information categories. The next step is to analyze the information content by classifying it by types.

Information types are amodal, conceptual descriptions of information components that elaborate the content definition. Information components are classified into one or more of the following:

- Physical items relating to tangible observable aspects of the world
- Spatial items relating to geography and location in the world
- Conceptual-abstract information, facts, and concepts related to language
- Static information which does not change: objects, entities, relationships, states, and attributes
- Dynamic, or time-varying information: events, actions, activities, procedures, and movements
- Descriptive information, attributes of objects and entities
- Values and numbers
- Causal explanations

More complex taxonomies elaborate concepts and linguistic information as ontologies and arguments (Mann and Thompson 1988), but additional complexity is only warranted for tools that automatically generate multimedia output (Zhou and Feiner 1998). It is important to note that one component may be classified with more than one type; for instance, instructions on how to get to the railway station may contain procedural information (the instructions <turn left, straight ahead, etc.>) and spatial or descriptive information (the station is in the corner of the square, painted blue). The information types are "tools for thought" that can be used either to classify specifications of content or to consider what content may be necessary. To illustrate, for the task "navigate to the railway station," the content may be minimally specified as "instructions how to get there," in which case the information types prompt questions in the form "what sort of information does the user need to fulfill the task/user goal?" Alternatively, the content may be specified as a scenario narrative of directions, waymarks to recognize, and description of the target. In this case, the types classify components in the narrative to elucidate the deeper structure of the content. The granularity of components is a matter of the designer's choice and will depend on

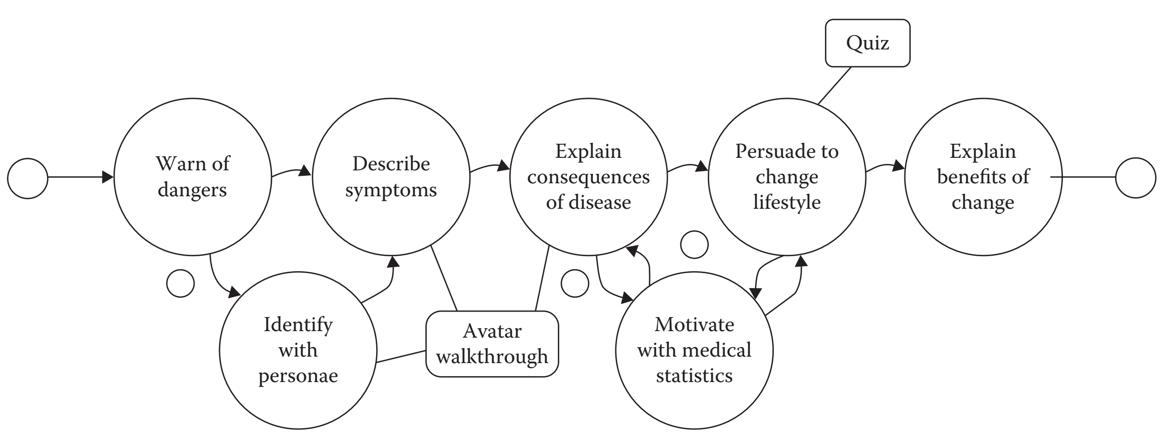

FIGURE 17.3 Thematic map for a healthcare promotion application.

the level of detail demanded by the application. To illustrate the analysis:

Communication goal: Explain how to assemble a bookshelf from ready-made parts.

Information component 1:
- Parts of the bookshelf, sides, back, shelves, connecting screws
- Mapping to information types:
  - Physical-static-descriptive; parts of the bookshelf are tangible, do not change and need to be described
  - Physical-static-spatial; dimensions of the parts, how they are organized
  - Physical-static-relationship type could also be added to describe which parts fit together

Information component 2:
- How to assemble parts instructions
- Mapping to information types:
  - Physical-dynamic-discrete action
  - Physical-dynamic-procedure
  - Physical-static-state; to show final assembled bookshelf

17.4.3 Media Selection and Combination

The information types are used to select appropriate categories of media resource(s). Media classifications focus on the psychological properties of the representations rather than the physical nature of the medium (e.g., digital or analogue encoding in video). Note that these definitions are combined to describe any specific medium, so speech is classified as an audio, linguistic medium, whereas a cartoon is classified as a nonrealistic (designed) moving image.

The definitions may be usefully considered in two dimensions of abstraction: the designer's involvement in creating the medium and rate of change. Media resources are classified using the decision tree illustrated in Figure 17.4. More fine-grained taxonomic distinctions can be made, for instance, between different signs and symbolic languages (see Bernsen [1994]), but as with information types, richer taxonomies increase specification effort.

The approach to classifying media uses a walk-through of the decision tree with the following questions that reflect the facets of the classification:

- Is the medium perceived to be realistic or not? Media resources captured directly from the real world will usually be realistic, for example, photographs of landscapes, sound recordings of bird song. In contrast, nonrealistic media are created by human action. However, the boundary case category that illustrates the dimension is a realistic painting of a landscape.
- Does the medium change over time or not? The boundary case here is the rate of change, particularly in animations where some people might judge 10 frames/s to be a video, but 5 slides in 1 minute shown by a PowerPoint presentation to be a sequence of static images.
- Which modality does the resource belong to? In this case, the categories are orthogonal, although one resource may exhibit two modalities; for example, a film with a soundtrack communicates in both visual and audio modalities.

Classification of media resources facilitates mapping of information types to media resources; however, the process may also guide the acquisition or creation of new resources,

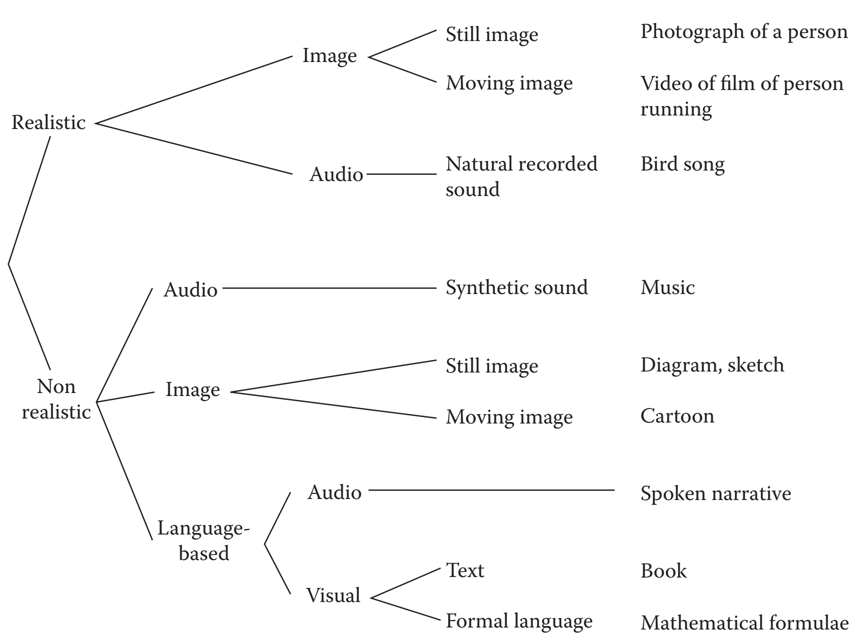

FIGURE 17.4 Decision tree for classifying media resources.

if appropriate resources are not present in the designer's media resource library. Finally, the classification provides a mechanism for indexing media resource libraries.

17.4.3.1 Media Selection

Recommendations for selecting media have to be interpreted according to the users' task and design goal. If information provision is the main design goal—for example, a tourist kiosk information system—then persistence of information and drawing attention to specific items is not necessarily as critical as in tutorial applications. Task and user characteristics influence media choice; for instance, verbal media are more appropriate to language-based and logical reasoning tasks; visual media are suitable for spatial tasks involving moving, positioning, and orienting objects. Some users may prefer visual media, whereas image is of little use for blind users. Media resources may be available for selection, or have to be purchased from elsewhere. If existing media can be edited and reused, this is usually preferable to creating new media from scratch. Graphical images can be particularly expensive to draw, whereas capture of images by scanning is usually quick and cheap. The following heuristics are supplemented by more detailed examples in Table 17.1.

- To convey detail, use static media, for example, text for language-based content, diagrams for models, or still image for physical detail of objects (Booher 1975; Faraday and Sutcliffe 1998).
- To engage the user and draw attention, use dynamic media—video for physical information, animation, or speech, for example.
- For spatial information, use diagrams, maps, with photographic images to illustrate detail, animations to indicate pathways (Bieger and Glock 1984; May and Barnard 1995).
- For values and quantitative information, use charts and graphs for overviews and trends, supplemented by tables for detail (Bertin 1983; Tufte 1997).
- Abstract concepts, relationships, and models should be illustrated with diagrams explained by text captions and speech to give supplementary information.
- Complex actions and procedures should be illustrated as a slideshow of images for each step, followed by a video of the whole sequence to integrate the steps. Text captions on the still images and speech commentary provide supplementary information (Hegarty and Just 1993). Text and bullet points summarize steps at the end, so choice trade-offs may be constrained by cost and quality considerations.
- To explain causality, still and moving image media need to be combined with text (Narayanan and Hegarty 1998). For example, the cause of a flood is explained by text describing excessive rainfall with an animation of the river level rising and overflowing its banks. Causal explanations of physical phenomena may be given by introducing the topic using linguistic media, showing the cause and effect by a combination of still image and text with speech captions for commentary; integrate the message by moving image with voice commentary and provide a bullet point text summary.

Because most components in the information architecture will have multiple information types and each information type may match several media, the selection process encourages multimedia integration. For example, when a procedure for explaining a physical task is required, first a series of realistic images will be selected, followed by video, and speech to integrate the steps, then text to summarize the key points.

The end point of media selection is media integration: one or more media will be selected for each information group to present complementary aspects of the topic. Some examples of media combination that amplify the basic selection guidelines are given in Table 17.1.

17.5 MEDIA DESIGN FOR USER ENGAGEMENT

The design process in Section 17.4 was oriented to a task-driven view of media. However, multimedia design is frequently motivated by the need to attract users' attention and to make the user experience interesting and engaging. These considerations may contradict some of the earlier guidelines because the design objective is to please the user and capture their attention rather than deliver information effectively. First, a health warning should be noted: The old saying "beauty is in the eye of the beholder" has good foundation. Judgments of aesthetic quality suffer from considerable individual differences. A person's reaction to a design is a function of their motivation (see Chapter 4 from the Second Edition of the HCI Handbook). Individual preferences, knowledge of the domain, and exposure to similar examples, to say nothing of peer opinion and "fashion." Furthermore, attractiveness is often influenced more by content than the choice of media or presentation format. The following guidelines should, therefore, be interpreted with care and their design manifestations tested with users.

17.5.1 Multimedia to Motivate and Persuade

Design of media for motivation is a complex area in its own right, and this topic is dealt with in more depth in Chapter 7 of the Second Edition (Fogg), so the treatment here will focus on media selection issues. Simple photographs or more complex interactive animations (talking heads or full body mannequins) have an attractive effect. We appear to ascribe human properties to computers when interfaces give human-like visual cues (Reeves and Nass 1996); however, the effectiveness of media representing people depends on the characters' appearance and voice; see Figure 17.5. In human–human conversation, we modify our reactions according to our knowledge, or assumptions about, the other person's role, group identification, culture, and intention (Clark 1996). For example, reactions to a military mannequin will be very different from those to the representation of a parson. Male voices tend to be treated as more authoritative than female voices.

TABLE 17.1
Media Selection Example

Information Type Media Type	Causation	Conceptual	Continuous Action	Descriptive	Discrete Action	Event	Physical	Procedure	Relationship	Spatial Information	State	Value
Realistic audio	Sound of rain and storms		Sound of skiing		Click of ON switch	Sound of the starting gun	*Noise of a tornado*			Echoes in a cave	Sound of snoring	Musical note encodes a value
Nonrealistic audio		Rising tone illustrates increasing magnetic force	Continuous tone signals progress of action	Morse code describes a ship	Tones signal open/close door	*Alarm siren*			Tones associate two objects	Sonar and Doppler effect	Continuous sound in a heartbeat monitor	
Speech	*Tell someone why El Nino happens*	Tell someone about your religious beliefs	Tell someone what a ski turn looks like	Verbal description of a person	Tell someone how to turn computer on	*Tell someone race has started*	Tell someone how it feels to be in a storm	Speak instructions on engine assembly	Tell someone Jack and Jill are related	Tell someone pathway to and location of railway station	Tell someone "Jane's asleep"	Verbal report of numbers, figures
Realistic still image	*Photograph of El Nino storms and ocean currents*	Statue of Liberty photograph represents "freedom"	Set of photographs showing snap shots of action	*Overview and detail photographs of a car*	*Photograph of computer ON switch*	Photograph of the start of a race	*Photograph of a person's face*	*Photographs showing engine assembly*		*Photograph of a landscape*	Photograph of a person sleeping	
Nonrealistic still image	*Diagrams of ocean currents and sea temp. to explain El Nino*	*Hierarchy diagram of plant taxonomy*	Diagram with arrow depicting ski turn motion	Histogram of aging population	Diagram showing where and how to press ON switch	Event symbol in a race sequence diagram		Explode parts diagram of engine with assembly numbers	*Graphs, histograms,* ER *diagrams*	*Map of the landscape*	Waiting state symbol in race sequence diagram	*Charts, graphs, scatter plots*
Text	*Describe reasons for El Nino storms*	*Explain taxonomy of animals*	Describe ski turn action	*Describe a person's appearance*	Describe how to turn computer on	Report that the race has started	Report of the storm's properties	*Bullet point steps in assembling engine*	*Describe brother and sister relationship*	Describe dimensions of a room	*Report that the person is asleep*	*Written number one, two*
Realistic moving image	*Video of El Nino storms and ocean currents*		*Movie of person turning while skiing*	Aircraft flying		*Movie of the start of a race*	*Movie of a storm*	*Video of engine assembly sequence*		Fly through landscape	Video of a person sleeping	
Nonrealistic moving image	*Animation of ocean temperature change and current reversal*	Animated diagram of force of gravity	Animated mannequin doing ski turn		Animation showing operation of ON switch	Animation of start event symbol in diagram		*Animation of parts diagram in assembly sequence*	Animation of links on ER diagram			
Language-based: formal, numeric	Equations, functions formalizing cause and effect	*Symbols denoting concepts, for example, pi*			Finite state automata	Event-based notations		Procedural logics, process algebras	Functions, equations, grammars		State-based languages, for example, Z	numeric symbols

The table summarizes the media selection and combinations for each information type. The italics denote the preferred mappings for media and information types, whereas ordinary text shows other potential media uses for the information type.

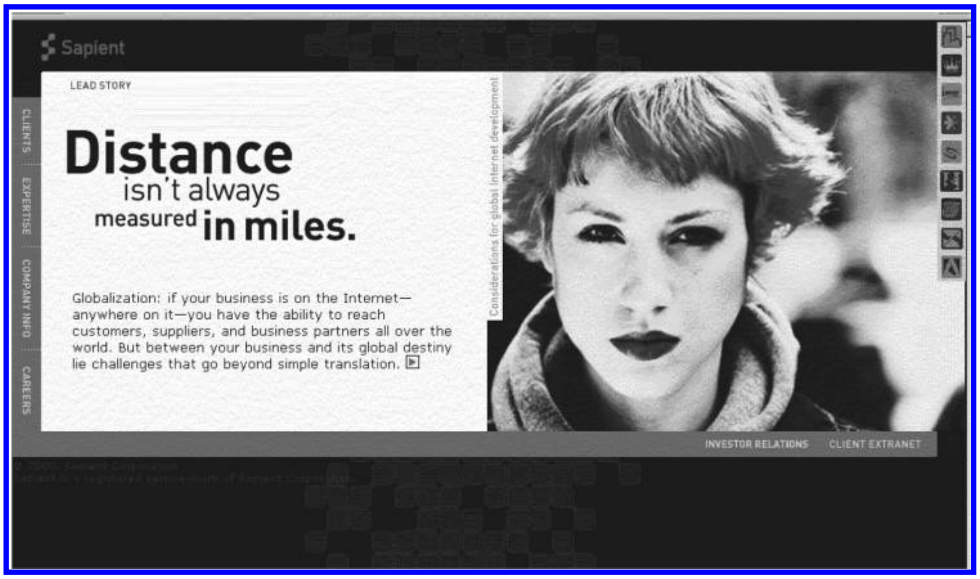

FIGURE 17.5 Effective use of human image for attraction. The picture attracted by the direction of gaze to the user as well as by the appearance of the individual.

Use of human-like forms is feasible with prerecorded video and photographs; however, the need depends on the application. Video representation of the lecturer can augment presentations, and video communication helps interactive dialogue. A good speaker holds our attention by a variety of tricks, such as maintaining eye contact, varying the voice tone, using simple and concise language, and delivering an interesting message. These general effects can be reinforced by projected personality. Friendly people are preferred over colder, more hostile individuals. TV announcers who tend to be middle-aged, confident, but avuncular characters have the attention-drawing power of a dominant yet friendly personality. Both sexes pay attention to extrovert, young personalities, whereas the male preference for beautiful young women is a particularly strong effect. These traits have been exploited by advertisers for a long time. There are lessons here for multimedia designers as the web and interactive TV converge, and when we want media to convey a persuasive message (Reeves and Nass 1996; Fogg, this book). Media selection guidelines for motivation and persuasion, adapted from Reeves and Nass (1996), can be summarized as follows:

- Human image and speech invokes the computer-as-social actor effect to facilitate motivation and persuasion.
- Photographs of people attract attention especially when the person is looking directly at the user.
- Faces that represent the norm in a population (Mr./Ms. average) and young children are more attractive. We are very susceptible to the large-eyes effect in young animals, as exploited by Disney cartoons.
- Polite praise: Use of "Please," "Thank you," and simple compliments like "That was an excellent choice" increase people's tendency to judge the computer as pleasant and enjoyable.
- Short compelling argument: Such as the well-known British World War I recruiting poster featuring General Kitchener gazing directly at the viewer with the caption "Your country needs you."

For more detailed treatment of design for persuasive technology, see Fogg, Chapter 7, Second Edition of the HCI Handbook.

17.5.2 Media for Emotional Effects

Media design for affect (emotional response and arousal) involves both choice of content and interaction. Arousal is increased by interactive applications, surprising events during interaction, use of dynamic media, and challenging images. In contrast, if the objective is to calm the users, arousal can be decreased by choice of natural images and sounds, and soothing music. The most common emotional responses that designers may want to invoke are pleasure, anxiety and fear, and surprise. Pleasure, anxiety, and fear usually depend on our memory of agents, objects, and events (Ortony, Clore, and Collins 1988), so content selection is the important determinant. Anxiety can be evoked by uncertainty in interaction and cues to hidden effects, whereas emotional response of fear or pleasure will depend on matching content to the user's previous experience. Some guidelines to consider are as follows:

- *Dynamic media*, especially video, have an arousing effect and attract attention; hence, video and animation are useful in improving the attractiveness of presentations. However, animation must be used with care, as gratuitous video which cannot be turned off quickly offends (Spool et al. 1999).

- *Speech* engages attention because we naturally listen to conversation. Choice of voice depends on the application: female voices for more restful and information effects, male voices to suggest authority and respect (Reeves and Nass 1996).
- *Image*s may be selected for mood setting, for example, to provide a restful setting for more important foreground information (Mullet and Sano 1995). Backgrounds in half shades and low-saturation color provide more depth and interest in an image.
- *Music* has an important emotive appeal, but it needs to be used with care. Classical music may be counterproductive for a younger audience, whereas older listeners will not find heavy metal pop attractive. Music can set the appropriate mood, for example, loud strident pieces will arouse and excite, romantic music calms and invokes pleasure, and so on.
- *Natural sounds* such as running water, wind in trees, bird song, and waves on a sea shore have restful properties and hence decrease arousal.
- *Dangerous and threatening episodes*, for example, being chased by a tiger, gory images (mutilated body), and erotic content all increase arousal and invoke emotions ranging from fear to anger, whereas pleasant images (e.g., flowers, sunset) tend to decrease it, that is, have calming effects and produce pleasurable emotional responses.
- *Characters* can appear threatening or benevolent depending on their appearance or dress. For example, disfigured people appear threatening and evoke emotions ranging from fear to disgust. Characters familiar from popular culture can be used for emotional effect.
- *Dialogue* is probably the most powerful tool for creating emotional responses, from threats to empathy. Emotional effects are additive so choice of character with a threatening appearance, complemented by a menacing voice tone and an aggressive dialogue, all reinforce the emotions of anxiety and fear.

Media integration rules may be broken for emotive effects. For example, use of two concurrent video streams might be arousing for a younger audience, as music TV (MTV) and pop videos indicate. Multiple audio and speech tracks can give the impression of complex, busy, and interesting environments.

17.5.3 Multimedia and Aesthetic Design

If the requirements analysis indicates that having a pleasing and attractive design is important for the user's perception, then aesthetics need to be considered in depth. However, aesthetics should be considered as a design criterion for all applications since poor appearance and interaction design may provoke adverse reaction (Norman 2004). Some studies suggest that aesthetic design is an important component of usability and overall preference (Tractinsky 1997; Tractinsky, Shoval-Katz, and Ikar 2000; Hassenzahl 2004); however, others have shown that aesthetic preferences are open to contextual effects on users' judgment (Sutcliffe and De Angeli 2005; Hartmann, Sutcliffe, and De Angeli 2008). Judging when aesthetics may be important is not easy. For example, in e-commerce applications with high-value, designer-label products, aesthetic presentation is advisable; similarly, when selling to a design-oriented audience. However, in many applications, the decision is not clear-cut.

Aesthetic design primarily concerns graphics and visual media. Evaluation questionnaires assess design on classic aesthetics, which broadly equate with conventional usability guidelines on structured and consistent layout, and expressive aesthetics that capture the more creative aspects of visual design (Lavie and Tractinsky 2004); however, these measure user reaction to general design aspects such as "original," "fascinating," "clear," and "pleasant." The following heuristics provide more design-directed guidance, but they may also be used for evaluation (Sutcliffe 2002; Sutcliffe and De Angeli 2005).

- *Judicious use of color*: Color use should be balanced and low-saturation pastel colors should be used for backgrounds. Designs should not use more than two to three fully saturated intense colors. Yellow is salient for alerting, red/green have danger/safety positive/negative associations, and blue is more effective for background. Low-saturated colors (pale shades with white) have a calming effect and are also useful for backgrounds. Color is a complex subject in its own right; for more guidance, see Travis (1991).
- *Depth of field*: Use of layers in an image stimulates interest and can attract by promoting curiosity. Use of background image with low-saturated color provides depth for foreground components. Use of layers in an image and washed-out background images stimulate curiosity and can be attractive by promoting a peaceful effect.
- *Use of shape*: Use of curved shapes conveys an attractive visual style, in contrast to blocks and rectangles which portray structure, categories, and order in a layout.
- *Symmetry*: Symmetrical layouts, for example, bilateral, radial organization that can be folded over to show the symmetrical match.
- *Simplicity and space*: Uncluttered, simple layout that uses space to separate and emphasize key components.
- *Design of unusual or challenging images* that stimulate the users' imagination and increase attraction: Unusual images often disobey normal laws of form and perspective.
- *Visual structure and organization*: Dividing an image into thirds (right, center, left; or top, middle, bottom) provides an attractive visual organization,

whereas rectangular shapes following the golden ratio (height/width = 1.618) are aesthetically pleasing. Use of grids to structure image components promotes consistency between pages.

Although guidelines provide ideas that can improve aesthetic design and the attractiveness of interfaces, they are no guarantee that these effects will be achieved. Design is often a trade-off between ease of use and aesthetic design; for instance, use of progressive disclosure to promote flow may well be perceived by others as being difficult to learn. Visual effects often show considerable individual differences and learning effects, so a well-intentioned design might not be successful. The advice, as with most design, is test ideas and preliminary designs with users to check interpretations, critique ideas, and evaluate their acceptability. There are several sources of more detailed advice on aesthetics and visual design (Kristoff and Satran 1995, Mullet and Sano 1995; Lidwell, Holden, and Butler 2003); however, advice is usually given as examples of good design rather than specific guidelines.

17.6 INTERACTION AND NAVIGATION

Although discussion of interactive multimedia has been delayed until now, in practice, dialogue and presentation design proceed hand in hand. Task analysis provides the basis for dialogue design and specification of navigation controls. Navigational and control dialogues allow flexible access to the multimedia content and enable users to control how media are played. Dialogue design may also involve specifying how users interact with tools, agents, and objects in interactive microworlds.

17.6.1 Metaphors and Interaction Design

Although task and domain analysis can provide ideas for interaction design, this is also a creative process. Interaction design is essentially a set of choices along a dimension from simple controls such as menus and buttons where the user is aware of the interface, to embodiment in which the user becomes involved as part of the action by controlling an avatar or other representation of their presence. At this end of the dimension, multimedia interaction converges with VR (see Chapter 29). Interactive metaphors occupy the middle ground.

Some interactive metaphors are generally applicable, such as timelines to move through historical information, the use of a compass to control direction of movement in an interactive space, controls based on automobiles (steering wheels) or ships (rudders). Others will be more specific, for example, selecting and interacting with different characters (young, old, male, female, overweight, fit, etc.) in a health-promotion application. Design of interaction also involves creating the microworld within which the user moves and interactive objects that can be selected and manipulated.

Interaction via characters and avatars can increase the user's sense of engagement first by selecting or even constructing the character, although some users may not have the patience to build their own avatar using a graphical paint program. In character-based interaction, the user can either see the world from an egocentric viewpoint, that is, from their character's position, or exocentric when they see their character in the graphical world. The sophistication in control of movement and interaction will depend on the hardware available (e.g., joystick, wand, or standard mouse and keyboard). Although mimicking physical interaction via data gloves and tracking requires VR technology, relatively complex interaction (e.g., actions in a football game, pass, head ball in direction north/south/east/west) can be programmed using buttons and function keys. Engagement is also promoted by surprise and unexpected effects, so as the user moves into a particular area, a new subworld opens up, or system-controlled avatars appear. These techniques are well known to games programmers; however, they are also applicable to other genres of multimedia applications. The design concepts for engagement can be summarized as follows (see Chapter 32):

- *Character-driven interaction*: This interaction provides the user with a choice of avatars or personae they can adopt as representations of themselves within the interactive virtual world; see Figure 17.6. Avatar development tools enable virtual characters to be designed and scripted with actions, and simple speech dialogues. Most sophisticated semi-intelligent "chatterbots" (e.g., Alice, Jabberwocky*) use response-planning rules to analyze user input and generate naturally sounding output; however, it is easy to fool these systems with complex natural language input.
- *Tool-based interaction*: This places tools in the world which users can pick up; the tool becomes the interface, for example, a virtual mirror magnifies, a virtual helicopter flies (Tan, Robertson, and Czerwinski 2001).
- *Collaborative characters*: In computer-mediated communication, these characters may represent other users; in other applications, system-controlled avatars appear to explain, guide, or warn the user.
- *Surprise effects*: Although conventional human–computer interaction (HCI) guidelines should encourage making the affordances and presence of interactive objects explicit, when designing for engagement, hiding, and surprise are important.

Interaction design for an explanatory/tutorial application is illustrated in Figure 17.6. This is an interactive microworld in which the user plays the role of a dinosaur character, illustrating use of the engagement concepts. A compass navigation metaphor allows the user to act as the dinosaur moving

*http://alice.pandorabots.com/ and http://www.jabberwacky.com

FIGURE 17.6 Interactive microworld: Big Al game (www.bbc.co.uk/sn/). The user plays the dinosaur role by navigating with the compass metaphor. The photograph updates with each move and the user is given a choice of attacking or avoiding other dinosaurs in the virtual world.

around the landscape illustrated in photographs. The user is given feedback on the characteristics of other predators and prey in the vicinity and has to decide whether to attack or avoid them. Other controls that might be added to such interactive microworlds could be settings to change the environment, for example, add more predators, change the weather, and so on. Engagement can be taken even further by giving the user facilities to actually design the MacWorld so the application becomes a domain-oriented design environment (Fischer et al. 2004).

17.6.2 Navigation

In information-intensive multimedia where access to content is the main design goal, hypermedia dialogues that link content segments will be appropriate. Good hypertext design is based on a sound information analysis that specifies the pathways between related items, and use of cues to show the structure of the information space to the user. In document-based hypermedia (e.g., HTML and the web), links can only access the whole media resource rather than point to components within it. The access structure of most hypermedia will be hierarchical, organized according to the information model and categorization of content, for example, information grouped by function, organization, task usage, or user preference. Navigation design transforms the user's conceptual model of an information space into a hypermedia structure. Unfortunately, individual users have different models so this may not be an easy task. Implementing too many links to satisfy each user's view will make the system too complex and increase the chance of the user getting lost. Too few links will frustrate users who cannot find the associations they want. Unfortunately, hypermedia systems assume a fixed link structure so the user is limited to the pathways provided by the designer. More open-ended hypermedia environments (e.g., Microcosm: Lowe and Hall 1998) provide more flexibility via links with query facilities which allow access to databases. Dynamic links attached to hotspots in images or nodes in text documents provide access paths to a wider variety of data.

One problem with large hypermedia systems is that users get lost in them. Navigation cues, waymarks, and mini-map overviews can help to counter the effects of disorientation. *Mini-maps* give an overview of the hypertext area and a reference context for where users are in large networks. *Filters* help to reduce complexity by showing only a subset of nodes and links that the user is interested in. Having typed links helps filtering views because the user can guess the information content from the link type, for example, reference, example, source, related work, and so on. Other navigation facilities are *visit lists* containing a history of nodes traversed in a session and *bookmarks* so users can tailor a hypermedia application with their own navigation aide-memoires (Nielsen 1995). Once the structure has been designed, access structures and link cues need to be located within media resources, so the appropriate cues need to be considered for each medium, such as the following, for example.

- *Text media*: The web convention is to underline and highlight text in a consistent color, for example, blue or purple. Text image thumbnails can be used to illustrate document and page structure, to facilitate direct pointing access, as used in the Adobe PDF Reader.
- *Images*: Link cues can be set as standalone icons or as active components in images. Icons need to be tested with users because the designer's assumed meaning can be ambiguous. Active components should signal the link's presence by captions or pop-up hover text so the user can inspect a link before deciding whether to follow it. Mosaics of image thumbnails are an alternative access path and can be organized in dimensions or layers to communicate categories and properties, such as time × location dimensions, general views to close-ups arranged in concentric circles of magnification. Slideshow presentation of images organized in navigation sequences with a stop button is another option for rapid access (De Bruijn, Spence, and Chong 2002).
- *Moving images*: Links from animation and film are difficult to design because the medium is dynamic; however, link buttons can be placed below the video window. Active components, for example, overlaid buttons within a moving image, are technically more challenging to program. Buttons may also be timed to pop-up at appropriate times during the video. Sample frames set in a mosaic and timeline structure summarize videos and enables access by pointing to segments. This can be taken further with multirunning movie thumbnails to provide overviews; however, this can produce more distraction than useful navigation, and many instances of dynamically running media overload our attention (see Section 17.3).
- *Sound and speech* links are difficult for the same reason as with moving images. One solution is to use visual cues, possibly synchronized with the sound or speech track. If speech recognition is available, then voice commands can act as links, but these commands need to be explained to the user. Visual access structures can be based on sonograms of the audio track or more usefully annotated timelines.

In many cases, controls will be provided by the media-rendering device, for example, video player for .avi files, or Quicktime movies. If controls have to be implemented from scratch, the following should be considered for each media type:

- *Static media*. Size and scale controls to zoom and pan; page access if the medium has page segmentation, as in text and diagrams; the ability to change attributes such as color, display resolution, font type, and size in text.
- *Dynamic media*. The familiar video controls of stop, start, play, pause, fast-forward, and rewind, also the ability to address a particular point or event in the media stream by a time marker or an index, for example, "go to" component/marker, and so on.

Navigation controls use standard UI components (buttons, dialogue boxes, menus, icons, sliders) and techniques (form filling, dialogue boxes, and selection menus); for more guidance, see ISO 9241, Parts 12, 14, and 17 (ISO 1997) and ISO 14915 Part 2 (ISO 1998).

17.6.3 Design for Attention

Having selected the media resources, the designer must now ensure that the user will extract the appropriate information. An important consideration of multimedia design is to link the thread of a message across several different media. This section gives recommendations on planning the user's reading/viewing sequence, and guidelines for realizing these recommendations in presentation sequences, hypermedia dialogues, and navigation controls. The essential differences are timing and user control. In a presentation design, the reading/viewing sequence and timing are set by the designer; whereas the reading/viewing sequence in hypertext implementation and interactive dialogues is under user control.

Presentation techniques help to direct the user's attention to important information and specify the desired order of reading or viewing. Thematic links between information components are specified and attention-directing techniques are selected to implement the desired effect.

The design issues are as follows:

- To plan the overall thematic thread of the message
- To draw the user's attention to important information
- To establish a clear reading/viewing sequence
- To provide clear links when the theme crosses from one medium to another

Design for attention is particularly important for images. User attention to time-varying media is determined by the medium itself, that is, we have little choice but to listen to speech or to view animations in the order in which they are presented. The reading sequence is directed by the layout of text, although this is culturally dependent, for example, western languages read left to right, Arabic in the opposite direction. However, viewing order in images is unpredictable unless the design specifically selects the user's attention.

The design problem is how to direct the user's attention to the appropriate information at the correct level of detail. Initially, users will tend to extract information from images at the scene level, that is, major objects will be identified but with very little descriptive detail (Treisman 1988). Regular layout grids help design composite images (Mullet and Sano 1995) and encourage viewing sequences in image sets. Alternatively, the window frame can be set to control which parts of an image are viewed. Larger window frames will be attended to before smaller areas. A list of the key components that the user needs to focus on and the facts that should be extracted are checked against the initial presentation design to see if the key components will attract sufficient attention or whether the user is likely to be confused by extraneous detail.

Position on screen is a key influence on attention. Eye-tracking studies have demonstrated that components on the top-left and center areas of screens receive more attention than the lower and right-hand side (Beymer, Orton, and Russell 2007; Hornof and Halverson 2003). Furthermore, large centralized images tend to dominate attention while layout structure, such as columns and blocks focus users' gaze within these structures. So layout, as well as media choice, can be used to influence attention; then within each media type, attention can be directed by applying the following highlighting techniques.

17.6.3.1 Still Image Media

Highlighting techniques for designed and natural images, organized in approximate power of their effect, are summarized in Table 17.2. A common highlighting technique will pick out spatially distributed objects, for example, change all the related objects to the same color; co-located objects can be grouped by using a common color or texture for their background or drawing a box around them. The highlighted area will set the granularity of the user's attention. Captions linked to objects in an image are another useful means of drawing attention and providing supplementary information (e.g., identity). Dynamic revealing of captions is particularly effective for directing the user's viewing sequence. Sequential highlighting is also useful for showing the pathways or navigational instructions.

17.6.3.2 Moving Image Media

Directing attention to components within moving images is difficult because of the dynamic nature of the medium. Design of film and video is an extensive subject in its own right, so treatment here will necessarily be brief. The following design advice is based on Hochberg (1986). The design objectives, as for other media, are how to draw the user's attention to key components within the video or animation.

First, the content needs to be structured into scenes that correspond to the information script. To structure animation sequences and make scene boundaries obvious, use a cut, wipe, or dissolve to emphasize that a change in the content structure has taken place. However, cuts should be used with care and continuity maintained between the two sequences if they are to be integrated. Continuity is manifest as the same viewpoint and subject matter in two contiguous shots. Change in background or action, such as an individual walking left in one clip and walking right in the next, is quickly noticed as a change. An establishing shot that shows the whole scene helps to introduce a new sequence and provide context. To provide detail of a newly introduced object or context, the object is shown filling the frame with a small amount of surrounding scene; while to imply a relationship or compare two objects a tight two shot with both objects together in the same frame is advisable.

17.6.3.3 Linguistic Media (Text and Speech)

As with moving image, the literature is extensive, so the following heuristics are a brief summary; see Levie and Lentz (1982) for more detail. Text may be structured to indicate subsections by indentation, formatting into paragraphs, columns, or segmented by background color. Bullet points or

TABLE 17.2
Attention-Directing Techniques for Different Media

	Attention-Orienting Techniques in Approximate Order of Power	Notes
Still image: designed and natural	Movement of or change in the shape/size/color of an object. Use of bold outline. Object marked with a symbol (e.g., arrow) or icon. Draw boundary, use color, shape, size, or texture to distinguish important objects.	Some effects may compromise natural images because they overlay the background image with new components (e.g., arrows, arcs, icons). Group objects by a common technique.
Moving image	Freeze frame followed by applying a still image highlight. Zoom, close-up shot of the object. Cuts, wipes, and dissolve effects.	Change in topographic motion, in which an object moves across the ground of an image, is more effective than internal movement of an object's components. Size and shape may be less effective for highlighting a moving object.
Text	Bold, font size, type, color, or underlining. To direct attention to larger segments of text, use formatting, bullet points, sub-sections, indentation.	Formatting techniques are paragraphs; headings/titles as entry points; indents to show hierarchical nesting, with bullet points and lists.
Speech/sound	Familiar voice. Silence followed by onset of sound. Different voices, or a change in voice prosody (tonality), amplitude (loudness), change, and variations in pitch (frequency), voice rate, change source direction, alarm sounds (police sirens).	Voices familiar to the user (e.g., close relatives) attract attention over nonfamiliar speech. Discourse markers "next," "because," "so," and so on draw attention to subsequent phrases.

numbered sections indicate order more formally, such as for procedures. Different voices help to structure speech while also attracting attention. If language is being used to set the context for accompanying media, it is important that the correct level of identification is set. For instance, a higher level concept, or the whole scene in an accompanying image, is described at the beginning of a script, and then lower level topics reset the user's focus. Discourse markers can make phrases and sentences more salient.

Adding attention-directing effects completes the design process; however, as with all UIs, there is no substitute for usability testing. Designs are constructed incrementally by iterations of design and evaluation that checks for usability using standard methods, with additional memory and comprehension tests for multimedia. So when testing a design, ask the user to tell you what they understood the message to be. This can be done during the presentation with a think-aloud protocol to check that users did attend to key items, and afterwards by a memory test. If key components in the message are not being remembered, then the design may need to be improved.

17.7 CONCLUSIONS

Multimedia still poses many issues for further research. The design method described in this chapter coupled with user-centered design can improve quality; however, there is still a need for experts to create specific media resources, for example, film/video, audio experts. Furthermore, considerable research is still necessary before we fully understand the psychology of multimedia interaction. Design for motivation and attractiveness is still poorly understood, and personality effects in media may not be robust when usability errors impede communication. The process by which people extract information from complex images still requires extensive research, although the increasing number of eye-tracking studies is beginning to throw some light on this topic. In the future, language and multimodal communication will change our conception of multimedia from its current CD-ROM or web-based form into interfaces that are conversational and multisensory. Multimedia will become part of wearable and ubiquitous UIs where the media is part of our everyday environment. Design for multisensory communication will treat media and artifacts (e.g., our desks, clothes, walls in our homes) as a continuum, whereas managing the diverse inputs to multimedia from creative design, technology, and usability engineering will be one of the many interesting future challenges.

REFERENCES

Alty, J. L. 1991. Multimedia: What is it and how do we exploit it? In *Proceedings of HCI '91: People and Computers VI*, ed. D. Diaper and N. V. Hammond, 31–41. Cambridge: Cambridge University Press.

Baddeley, A. D. 1986. *Working Memory*. Oxford: Oxford University Press.

Barnard, P. 1985. Interacting cognitive subsystems: A psycholinguistic approach to short term memory. In *Progress in Psychology of Language*, ed. A. Ellis, Vol. 2, 197–258. London: LEA.

Bernsen, N. O. 1994. Foundations of multimodal representations: A taxonomy of representational modalities. *Interact Comput* 6(4):347–71.

Bertin, J. 1983. *Semiology of Graphics*. Madison, WI: University of Wisconsin Press.

Beymer, D., P. Z. Orton, and D. M. Russell. 2007. An eye tracking study of how pictures influence online reading. In *Proceedings of Interact 2007*, 456–60. Berlin, Germany: Springer.

Bieger, G. R., and M. D. Glock. 1984. The information content of picture-text instructions. *J Exp Educ* 53:68–76.

Booher, H. R. 1975. Relative comprehensibility of pictorial information and printed word in proceduralized instructions. *Hum Factors* 17(3):266–77.

Card, S. K., T. P. Moran, and A. Newell. 1983. *The Psychology of Human Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.

De Bruijn, O., R. Spence, and M. Y. Chong. 2002. RSVP Browser: Web browsing on small screen devices. *Pers Ubiquitous Comput* 6(4):245–52.

Faraday, P., and A. G. Sutcliffe. 1996. An empirical study of attending and comprehending multimedia presentations. In *Proceedings ACM Multimedia '96: 4th Multimedia Conference, Boston, MA 18–22 November 1996*, 265–75. New York: ACM Press.

Faraday, P., and A. G. Sutcliffe. 1997. Multimedia: Design for the moment. In *Proceedings: Fifth ACM International Multimedia Conference*. Seattle, WA 9–13 November 1997, 183–92. New York: ACM Press.

Faraday, P., and A. G. Sutcliffe. 1998. Making contact points between text and images. In *Proceedings ACM Multimedia '98 of 6th ACM International Multimedia Conference*, 29–37. New York: ACM Press.

Fischer, G., E. Giaccardi, Y. Ye, A. G. Sutcliffe, and N. Mehandjiev. 2004. A framework for end-user development: Socio-technical perspectives and meta-design. *Commun ACM* 47(9):33–9.

Fogg, B. J. 1998. Persuasive computer: Perspectives and research directions. In *Proceedings: Human Factors in Computing Systems: CHI '98*, Los Angeles CA 18–23 April 1998, 225–32. New York: ACM Press.

Gardiner, M., and B. Christie. 1987. *Applying Cognitive Psychology to User Interface Design*. Chichester, UK: Wiley.

Hartmann, J., A. G. Sutcliffe, and A. De Angeli. 2008. Towards a theory of user judgment of aesthetics and user interface quality. *ACM Trans Comput Hum Interact* 15(4):15–30.

Hassenzahl, M. 2004. The interplay of beauty, goodness and usability in interactive products. *Hum Comput Interact* 19(4):319–49.

Hegarty, M., and M. A. Just. 1993. Constructing mental models of text and diagrams. *J Mem Lang* 32:717–42.

Hochberg, J. 1986. Presentation of motion and space in video and cinematic displays. In *Handbook of Perception and Human Performance, 1: Sensory Processes and Perception*, ed. K. R. Boff, L. Kaufman, and J. P. Thomas. New York: Wiley.

Hollan, J. D., E. L. Hutchins, and L. Weitzman. 1984. Steamer: An interactive inspectable simulation-based training system. *AI Mag* 5(2):15–27.

Hornof, A., and T. Halverson. 2003. Cognitive strategies and eye movements for searching hierarchical computer displays. In *CHI 2003 Conference Proceedings of Conference on Human Factors in Computing Systems*. New York: ACM Press.

ISO. 1997. *ISO 9241: Ergonomic Requirements for Office Systems with Visual Display Terminals (VDTs)*. Geneva, Switzerland: International Standards Organisation.

ISO. 1998. *ISO 14915 Multimedia User Interface Design Software Ergonomic Requirements, Part 1: Introduction and Framework; Part 3: Media Combination and Selection*. International Standards Organisation. Geneva, Switzerland.

Kristof, R., and A. Satran. 1995. *Interactivity by Design: Creating and Communicating with New Media*. Mountain View, CA: Adobe Press.

Lavie, T., and N. Tractinsky. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *Int J Hum Comput Stud* 60(3):269–98.

Levie, W. H., and R. Lentz. 1982. Effects of text illustrations: A review of research. *Educ Comput Technol J* 30(4):159–232.

Lidwell, W., K. Holden, and J. Butler. 2003. *Universal Principles of Design*. Gloucester, MA: Rockport.

Lowe, D., and W. Hall. 1998. *Hypermedia and the Web*. Chichester, West Sussex: John Wiley.

Mann, W. C., and S. A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organisation. *Text* 8(3):243–81.

May, J., and P. Barnard. 1995. Cinematography and interface design. In *Proceedings: Fifth IFIP TC 13 International Conference on Human-Computer Interaction, Lillehammer, Norway 27–29 June 1995*, ed. K. Nordbyn, P. H. Helmersen, D. J. Gilmore, and S. A. Arnesen, 26–31. London: Chapman & Hall.

Mullet, K., and D. Sano. 1995. *Designing Visual Interfaces: Communication Oriented Techniques*. Englewood Cliffs, NJ: SunSoft Press.

Narayanan, N. H., and M. Hegarty. 1998. On designing comprehensible interactive hypermedia manuals. *Int J Hum Comput Stud* 48:267–301.

Nielsen, J. 1995. *Multimedia and Hypertext: The Internet and Beyond*. Boston, MA: AP Professional.

Norman, D. A. 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.

Norman, D. A., and T. Shallice. 1986. Attention to action: Willed and automatic control of behaviour. In *Consciousness and Self-Regulation*, ed. G. E. Davidson and G. E. Schwartz, Vol. 4, 1–18. New York: Plenum.

Ortony, A., G. L. Clore, and A. Collins. 1988. *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

Papert, S. 1980. *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books.

Reeves, B., and C. Nass. 1996. *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Stanford CA/Cambridge: CLSI/Cambridge University Press.

Rogers, Y., and M. Scaife. 1998. How can interactive multimedia facilitate learning? In *Intelligence and Multimodality in Multimedia Interfaces: Research and Applications*, ed. J. Lee. Menlo Park, CA: AAAI Press.

Scaife, M., Y. Rogers, F. Aldrich, and M. Davies. 1997. Designing for or designing with? Informant design for interactive learning environments. In *Proceedings: Human Factors in Computing Systems CHI '97*, Atlanta GA 22–27 May 1997, ed. S. Pemberton, 343–50. New York: ACM Press.

Spool, J. M., T. Scanlon, C. Snyder, W. Schroeder, and T. de Angelo. 1999. *Web Site Usability: A Designer's Guide*. San Francisco, CA: Morgan Kaufmann.

Sutcliffe, A. G. 1997. Task-related information analysis. *Int J Hum Comput Stud* 47(2):223–57.

Sutcliffe, A. G. 1999. User-centered design for multimedia applications. In *Proceedings Vol. 1: IEEE Conference on Multimedia Computing and Systems, Florence*, 116–23. Los Alamitos, CA: IEEE Computer Society Press.

Sutcliffe, A. G. 2002. Assessing the reliability of heuristic evaluation for website attractiveness and usability. In *Proceedings HICSS-35: Hawaii International Conference on System Sciences*, Hawaii 7–10 January 2002, 1838–47. Los Alamitos, CA: IEEE Computer Society Press.

Sutcliffe, A. G. 2003. *Multimedia and Virtual Reality: Designing Multisensory User Interfaces*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sutcliffe, A. G. 2009. Designing for user engagement: Aesthetic and attractive user interfaces. In *Synthesis Lectures on Human Centered Informatics*, ed. J. M. Carroll. San Rafael, CA: Morgan Claypool.

Sutcliffe, A. G., and A. De Angeli. 2005. Assessing interaction styles in web user interfaces. In *Proceedings of Human Computer Interaction—Interact 2005*, 405–17. Berlin, Germany: Springer.

Sutcliffe, A. G., and P. Faraday. 1994. Designing presentation in multimedia interfaces. In *CHI '94 Conference Proceedings: Human Factors in Computing Systems 'Celebrating Interdependence,'* Boston, MA, April 24–28, 1994, ed. B. Adelson, S. Dumais, and J. Olson, 92–8. New York: ACM Press.

Tan, D. S., G. R. Robertson, and M. Czerwinski. 2001. Exploring 3D navigation: Combining speed coupled flying with orbiting. In *CHI 2001 Conference Proceedings: Conference on Human Factors in Computing Systems*, Seattle, March 31–April 5, 2001, ed. J. A. Jacko, A. Sears, M. Beaudouin-Lafon, and R. J. K. Jacob, 418–25. New York: ACM Press.

Tractinsky, N. 1997. Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. In *Human Factors in Computing Systems: CHI '97 Conference Proceedings*, Atlanta GA May 22–27, 1997, ed. S. Pemberton, 115–22. New York: ACM Press.

Tractinsky, N., A. Shoval-Katz, and D. Ikar. 2000. What is beautiful is usable. *Interact Comput* 13(2):127–45.

Travis, D. 1991. *Effective Colour Displays: Theory and Practice*. Boston, MA: Academic Press.

Treisman, A. 1988. Features and objects: Fourteenth Bartlett memorial lecture. *Q J Exp Psychol* 40A(2):201–37.

Tufte, E. R. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CN: Graphics Press.

Wickens, C. D., D. Sandry, and M. Vidulich. 1983. Compatibility and resource competition between modalities of input, output and central processing. *Hum Factors* 25:227–48.

Zhou, M. X., and S. K. Feiner. 1998. Visual task characterization for automated visual discourse synthesis. In *Human Factors in Computing Systems of CHI '98 Conference Proceedings*. New York: ACM Press.