MULTIMODAL PRESENTATION OF BIOMEDICAL DATA

Zeljko Obrenović
Dušan Starčević
Emil Jovanov
University of Alabama in
    Huntsville
Huntsville, Alabama

1. INTRODUCTION

Existing biomedical presentation is mostly based on a graphical human-computer interaction paradigm that has not been changed fundamentally for nearly two decades. As the computer power improves exponentially, the human-computer interface has become a bottleneck for many applications. Efficient presentation of data in biomedicine is even more important, as increasing complexity of biomedical data makes use and analysis of the data very difficult. Physicians and scientists have to handle very large, multidimensional, multivariate biomedical datasets, sometimes in quite limited time. In addition, some medical analyses are made under stress, which may lead to errors, sometimes with fatal consequences. The Institute of Medicine report estimated that almost 100,000 people die each year from medical errors in hospitals (1). Although most of those errors are not a direct consequence of incorrect biomedical data analysis, the improvement in this field can have direct effects on the reliability of medical procedures.

Multimodal user interfaces can significantly improve the presentation of biomedical data. Multimodal presentation "uses different modalities, like visual presentation, audio, and tactile feedback, to engage human perceptual, cognitive, and communication skills in understanding what is being presented" (2). Multimodal user interfaces are part of ongoing research in the field of human-computer interaction (HCI) aimed at better understanding of how people interact with each other and with their environment. These new interaction techniques, called perceptual user interfaces, combine the understanding of natural human capabilities (such as communication, motor, cognitive, and perceptual skills) with computer input/output (I/O) devices and machine perception and reasoning. This approach allows integration of the human mind's exploration abilities with the enormous processing power of computers, to form a powerful knowledge discovery environment that capitalizes on the best of both worlds (3). Multimodal user interfaces explore means to improve human communication with a computer by emphasizing human communication skills (4). The multidimensional nature and complexity of biomedical data presents a rich field where multimodal presentation feats occur naturally (5).

In the next section, we define some basic terms, followed by history and description of basic problems in the presentation of biomedical data. In subsequent sections, we introduce presentation modalities and various forms of multimodal integration, we describe available tools and platforms for development of multimodal interfaces, and, finally, we present some examples of multimodal presentations in biomedical applications.

2. DEFINING TERMS

The distinction between some of the terms in the HCI field, such as multimedia and multimodal interfaces, is subtle and can be misleading. These and other technical terms are often used carelessly, without complete understanding of what they mean (6). Therefore, in this section, we describe some of the basic terms important for multimodal interfaces.

Media conveys information, and it generally includes anything that can serve for that purpose: paper, paint, video, CD-ROM, or computer screen.

Modality has an even more ambiguous meaning. When talking about multimodal HCI, we refer to the human sensory modalities of vision, hearing, touch, smell, and taste. However, computer scientists often use the term mode as a synonym for state. We can also speak of *modes of interaction* or *interaction styles*. For example, direct manipulation and natural language are interaction styles or modalities.

User interface mediates in communications between humans and computers. Every user interface must include some physical I/O devices. User interface defines some form of a communication language that defines the semantic of the messages to be exchanged between computers and humans. These messages define an interactive language (i.e., a series of organized dialogue patterns determining how and which messages can be exchanged between users and systems) (7). The language also defines a set of symbols, signals, and conventions used in a consistent way for communication.

Multimodal interface is the user interface that engages multiple human sensory modalities in HCI. The modalities are integral and essential components of the interface language. Rather than focusing on the media, which is the main topic of multimedia interfaces, multimodal interfaces focus on the advantages and limitations of human perceptual channels.

Perceptualization is an extension of visualization to include other perceptual senses and enable human sensory fusion, in which the information received over one sense is accompanied by a perception in another sense. Multisensory perception could improve understanding of complex phenomena by giving other clues or triggering different associations. For example, an acoustic channel could facilitate new information channels to visualization without information overload (5,8).

3. PRESENTATION OF BIOMEDICAL DATA

Presentation of biomedical data dates back to the first biomedical measuring devices. In this section, we describe early biomedical signal detection, computer-based signal detection and presentation, as well as contemporary types of biomedical data.

2   MULTIMODAL PRESENTATION OF BIOMEDICAL DATA

3.1. Early Biomedical Signal Detection and Presentation

The earliest and still widely used ways of detecting biomedical information is direct perceptual observation. For example, in addition to electrocardiograph (ECG) signal monitoring, an ear placed on the chest enables hearing of rhythmic beatings of the heart, and fingertips positioned at an artery make it possible to feel the pulsed blood flow. It is important to note that *this kind of examination is in its nature multimodal*: Physicians combine visual inspection, hearing, touch, and smell to better examine a patient.

However, most biomedical signals are not directly perceptible by human sensory organs. Therefore, specialized devices were developed to amplify biomedical signals and transform them to a form more suitable for human inspection. For example, stethoscopes are used in auscultation to detect sounds in the chest or other parts of the body, and microscopes are used for inspecting objects too small to be seen distinctly by the unaided eye. Growth of electronic technology in the twentieth century enabled development of various electronic sensing devices (9). In 1921, Herbert Gasser used a primitive three-stage triode amplifier, later coupled to a rudimentary cathode ray oscilloscope, to observe the structure and functions of the individual nerve fibers in the frog's compound sciatic nerve. Gasser also used the electron microscope to disclose the structure and function of the olfactory nerves. In 1924, Dutch physiologist Willem Eithoven won the Nobel prize for inventing the galvanometer, used to detect and display electrical signals produced by the heart muscle (10). One of the first completely electronic biomedical sensing and presentation devices was the toposcope, developed in the 1920s at the Burden Neurological Institute in Bristol, England.

Biomedical signals have been traditionally presented as polygraphic traces. This method of representation is a legacy of paper-based display techniques used with the first sensing devices. The use of polygraphic display mechanisms has also influenced the names of some sensing devices, for example, an electroencephalograph (EEG), an electromyography (EMG), an electrocardiograph (ECG), or an electrooculograph (EOG). Beyond polygraphic traces, early sensing devices used various gauges with a graduated scale or meter for analog signal display of signal magnitude. Some stand-alone electronic units also incorporated digital displays combining graphical and numerical presentation. Other devices added sound and musical tones to indicate the current magnitude of detected physiological signals (11).

3.2. Computer-Based Signal Detection and Presentation

Computer-based biomedical devices made possible the use of many sophisticated processing and presentation techniques. As the functionality of a computer sensing and presentation system depend very little on dedicated hardware, functionality and appearance of the instrument may be completely changed.

Computer-based presentation of biomedical data is widely used today. As user interfaces are shaped and changed more easier than conventional instrument's user interfaces, it is possible to employ more presentation effects and to customize the interface for each user.

The first computer programs had character-oriented terminal user interfaces, which was necessary as earlier general-purpose computers were not capable of presenting complex graphics. Graphical user interfaces (GUIs) enabled more intuitive HCI (12). Simplicity of interaction and high intuitiveness of GUI operations made possible the creation of user-friendly virtual instruments (Fig. 1). In addition, improvements in presentation capabilities of personal computers allowed development of various sophisticated 2-D and 3-D medical imaging technologies (13).

GUIs extended presentation in the visualization domain. However, contemporary personal computers are capable of presenting other modalities such as sonification
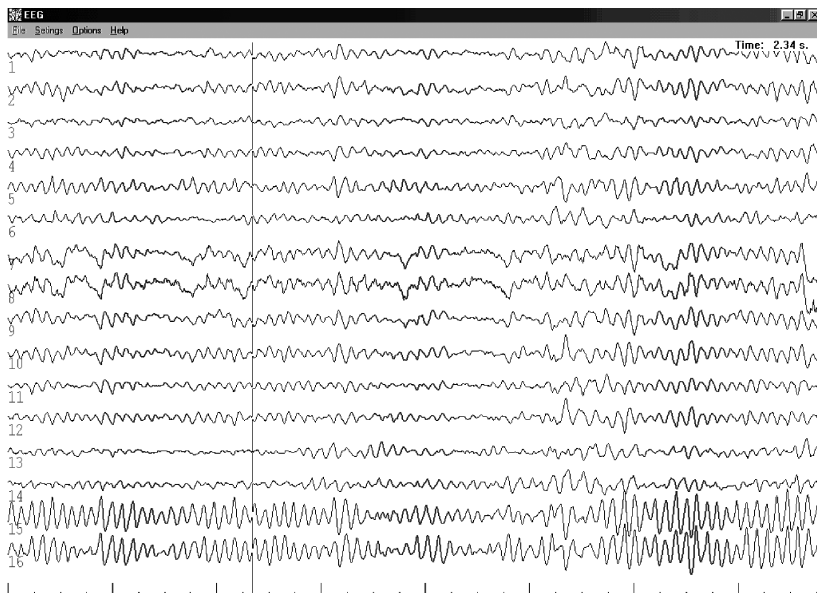

Figure 1. EEG recording with paroxysmal epileptic discharge of 3 Hz polyspike spike and wave complexes.

or haptic rendering in real-time, improving the perceptual quality of user interfaces (14).

Almost all sensing and presentation of biomedical data is computerized today. Improvements and new developments of bio-instrumentation have generated a wide variety of biomedical data formats (15). Starting from the early ECG, EEG, or EMG, biomedicine today uses many other data types captured with various devices, such as ultrasound, computerized tomography, magnetic resonance imaging, or infrared imaging (9,16). In addition, biomedical data include various biosequence data such as protein structures, protein-protein interactions, or gene expression levels.

Until recently, biomedical data have mainly had visual presentation. It affected the names of some of the biomedical data formats, for example, magnetic source *imaging* or magnetic resonance *imaging*. Although the visual presentation remains the main computer presentation channel, other presentation alternatives are now feasible. In the following sections, we will describe each of the computer presentation modalities and multimodal integration.

4. MULTIMODAL PRESENTATION

Multimodal interfaces are a part of the broader class of user interfaces called *perceptual user interfaces* (2). Perceptual interfaces introduce additional communication modalities to facilitate interaction with users via more and different sensory channels. Therefore, development of perceptual and multimodal interfaces is an interdisciplinary field that requires knowledge from many fields such as computer sciences, cognitive sciences, and biological sciences.

Multimodal interfaces apply features of everyday human-human interaction, where humans in face-to-face communication use multiple modalities, such as speech and gestures, in order to have rich and more effective conversational interaction. Multimodal communication is bidirectional. Many research projects in the HCI field have focused on computer input, for example, combining speech with pen-based gestures (4). We are primarily concerned with multimodal output that uses different modalities, such as visual display, audio, and tactile feedback, to engage human perceptual, cognitive, and communication skills in *understanding what is being presented* (3,17). In multimodal user interfaces, various modalities are some-

times used independently and sometimes simultaneously or tightly coupled.

The main benefits of using multimodal user interfaces are:

- Multimodal communication is a *natural* feature of human communication. Users have a strong preference to interact multimodally rather than unimodally (18).
- Using multiple modalities, the user can *choose* preferred modalities and can personalize presentation.
- Multimodal communication is more *robust* and less error prone (19).
- Multimodal presentation has a *higher bandwidth*. Overload on one sense modality causes tiredness and reduced attention. A balance between various modalities can reduce the cognitive load.

Figure 2 shows the abstract structure of multimodal presentation. A computer program presents some data using available presentation modalities. The program has to make multimodal integration, providing synchronized rendering of data. Computer presentation modalities use some presentation devices to modulate physical media creating signals such as light, sound, force, or scent. On the other side, humans perceive presented data over various sensory modalities, which are consciously integrated and correlated.

In the following sections, we present in more detail each of the presentation modalities, multimodal integration mechanisms, as well as some approaches to modeling of human information processing of interest to multimodal interfaces.

4.1. Computing Presentation Modalities

We can classify computer presentation modalities into four groups:

- *Visualization*, which exploits human vision;
- *Audio presentation*, which makes use of human sound perception;
- *Haptic rendering*, which includes human touch and kinesthetic senses; and
- *Olfactory presentation*, which makes possible limited use of human olfactory sense (smell).
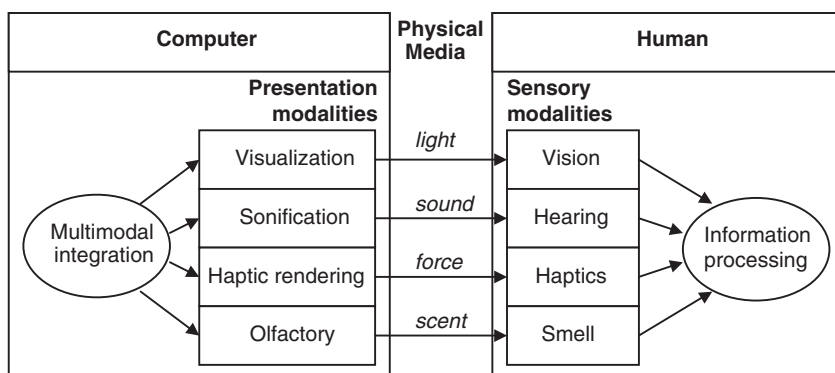

Figure 2. Multimodal presentation.

4   MULTIMODAL PRESENTATION OF BIOMEDICAL DATA

4.1.1. Visualization. In multimodal research, visualization primarily explores *perceptual effects* of the visual presentation. We present some of the recent visualization approaches that can be valuable for presentation of biomedical data such as color, motion, and three-dimensional presentation.

4.1.1.1. Color. Color is often used in user interfaces to code information, to attract user's attention to some elements, to warn a user, and to enhance display aesthetics. Many medical applications were also improved by real or false coloring. By introducing principles of color theory and design to interface tools, the design of user interfaces can be further improved. However, using color effectively and tastefully is often beyond the abilities of application programmers because the study of color crosses many disciplines. Careless use of color may lead to bad design. For example, although the addition of color stimulates and excites, using too many colors may turn into a cacophony of colors that distracts the user. In addition, many aspects of human color vision are not completely understood. For example, the number of colors the human visual system can distinguish is still not clearly determined. Hochberg states 350,000 distinguishable colors exist (20), whereas Tufte gives a range of 20,000 for the average viewer to 1,000,000 colors for trained subjects doing pair-wise comparisons (21).

One of the most difficult issues in using color in visualization is avoiding fake color-induced interpretations of the data. In human vision, colors appear as interrelated sensations that cannot be predicted from the response generated from viewing colors in isolation. For example, some colors stand out brightly over some background, whereas other colors recede. In this way, some dataset may appear to be more important, whether or not that was the designer's intent. Color is therefore often claimed to be the most relative medium (22).

Beyond perception, many other factors, such as culture, affect the meaning of color presentation. The designer of any sort of graphical representation needs to be aware of the ways that colors can influence the meaning. In addition, because most users are not sophisticated designers, they should not be required to pick individual colors, but should be able to choose among color suites suggested for their perceptual qualities and cultural meaning (23).

However, although it is possible that perception of color may differ from one user to the other, experimental evidence suggests that the relationships between colors are, in many respects, universal, and thus relatively free from individual and cultural influences. Therefore, it is possible to define a model of color experience based on the types of interactions among colors. Guidelines for usage of these kinds of color models can be found in Jacobson and Bender (24).

Contemporary presentation devices, like monitors, can produce millions of colors (16,777,216 for a 24-bit palette). Therefore, there have been many approaches to organize the range of displayable colors, creating so-called *color models*. Various color models have been proposed and used. In computing, the RGB (red, green, blue), and HSV (hue, saturation, value) models are the two most widely used color models. Both color models organize colors into a three-dimensional space, but the RGB color model is more hardware-oriented, whereas the HSV color model is based on more intuitive appeal of the artist's tint, shade, and tone (25). User interaction with this huge color space is usually achieved with specialized color picker user interfaces. These interfaces usually have a visible screen representation, an input method, and the underlying conceptual organization of the color model. Various issues, such as visual feedback and design of the color picker interface, may be important factors in improving the usability of a color selection interface.

At the end, it is important to note that there has been much research on the design of user interfaces that can be efficiently visible by users with color-deficient vision. For example, unlike normal trichromats, who use three-color receptors to blend the wavelengths corresponding to red, green, and blue light, dichromats have only two receptors functioning. By following some basic guidelines, it is possible to ensure that user interfaces do not put users with color-deficient vision at a disadvantage (26).

4.1.1.2. Motion and Animation. Overuse of static graphical representations exceeds the human's perceptual capacity to efficiently interpret information. Therefore, many attempts have been made to apply a dynamic visual presentation. Motion and animation hold promise as a perceptually rich and efficient display modality that may increase perceptual bandwidth of visualization (27). In a visually crowded environment, the moving objects clearly stand out. Low-level preattentive vision processes perceive motion very quickly. Several parts of the visual system exist that are especially motion sensitive, and specialized neural circuitry for motion perception exists. Motion is one of the strongest visual appeals to attention, especially if the motion is directed toward us. Motion perceptually awakens people and prepares them for possible action. This orienting response to motion occurs automatically. Furthermore, we perceive objects that are moving in the same direction as a group, so one can focus on the grouping formed by things moving in unison. This useful cognitive ability is thought to be the product of evolution that enabled us to detect a predator moving behind a cover of rustling leaves (28).

However, animation effects should be used carefully. Animation attracts attention, but if it is used too much, it can shut down other mental processes (14). For example, animated banners are very common on Web pages, but experimental evaluations indicate that animation does not enhance user memory (29). In addition, generating *realistic* visual scenes with animation, especially in virtual environments, is not an easy task as the human visual system is extremely sensitive to any anomalies in perceived imagery. The smallest, almost imperceptible anomaly becomes rather obvious when motion is introduced into a virtual scene (30).

4.1.1.3. Three-Dimensional Presentation. Increased performance of graphical hardware has enabled the development of a wide variety of techniques to render spatial characteristics of computer-generated objects in three-di-

mensional (3-D) space, which is particularly important for virtual reality (VR) systems. Recognition of spatial relationships based on simple 3-D depth cues is our preattentive perceptual ability, which allows us to easily understand spatial relationships among objects without cognitive load. Some human 3-D cues include motion parallax, stereovision, linear perspective (converting lines), relative size, shadow, familiar size, interposition, relative height, and horizon (31,32). There are an increasing number of research projects using 3-D effects in desktop computing. For example, 3-D document management systems proposed in Robertson et al. (33) showed statistically reliable advantages over the Microsoft Internet Explorer mechanisms for managing documents, which may be important for physicians who have to work with a great number of documents. 3-D presentation is also commonplace in medical imaging (13).

However, the use of 3-D presentation is not always justified, especially if we have the computational costs of it in mind. Some recent studies showed that some 3-D interfaces are not more efficient that traditional 2-D counterparts, although users did show a significant preference for the 3-D interface (34).

4.1.2. Audio Presentation. Natural language and audio channel are very important modalities of human-to-human communication. In HCI, audio presentation can be effectively used for many purposes. Voice and sound guide the attention and give additional value to the content through intonation. The sound dimension offers an alternative to reading text from the computer screen, which can be especially important for visually impaired users. Although the use of audio modes in user interfaces is still rather limited, with basic sound hardware now available for most computing systems, extending user interfaces by using the audio channel is becoming common.

The use of auditory modes is not always straightforward, as audio presentation has strengths and weaknesses that need to be understood. Therefore, in this section, we first talk about sound, sound rendering, and the perception of sound. After that, we describe two main groups of audio presentation: *nonspeech audio* and *speech audio*.

4.1.2.1. Sound, Sound Rendering, and Perception of Sound. Anderson and Casey identified three types of sound sources of interest to virtual environments (35):

- *Live speech*, which is produced by participants in a virtual environment;
- *Foley sounds*, which are associated with some events in the environments, such as collision of objects; and
- *Ambient sounds*, which are continuously repeated sounds, such as music, that define an atmosphere of a place, improving the sense of immersion.

The increasing processor power allows real-time sound manipulation, but sound rendering is a difficult process as it is based on complex physical laws of sound propagation and reflection (36). Several specificities of sound simulations exist that differentiate sound simulations from visual 3-D simulations (37):

- The wavelengths of audible sound (20 Hz–20 KHz) range between 0.02 meters and 17 meters, which means that small objects have little effect on the sound diffraction, and consequently, acoustics simulations use 3-D models with less geometric details. However, they must consider the effects that different obstacles have on a range of wavelengths.
- Sound propagation delays are perceptible to humans, and consequently, acoustic models must compute exact time/frequency distribution of the propagation paths.
- Sound is a coherent wave phenomenon, and interference between waves can be significant. Accordingly, acoustical simulations must consider phase when summing the cumulative contribution of many propagation paths to a receiver.
- The human ear is sensitive to five orders of magnitude difference in sound amplitude, and arrival time differences make some high-order reflections audible. Therefore, acoustical simulations usually aim to compute several times more reflections.

In addition to the knowledge of physical laws, efficient use of sound in user interfaces has to take into account the properties of the human auditory system and simplify audio presentation without degrading perceptual quality of sound (38,39). Perception of sound in space or audio localization, which assists listeners in distinguishing separate sound sources, is one of the most important properties of human auditory systems. The human ear can locate a sound source even in the presence of strong conflicting echoes by rejecting the unwanted sounds (40). The human ear can also isolate a particular sound source from among a collection of others all originating from different locations (41). In order to effectively develop aural displays, this ability of listeners to track and focus in on a particular auditory source (i.e., the cocktail party effect) needs to be better understood. Audio localization is based on many perceptual cues (42,43), including:

- Time delay between our two ears, also called the interaural time (or phase) difference, which is the strongest 3-D audio cue. Humans can distinguish interaural time delay cues of only few microseconds (44).
- Amplitude difference between our two ears, also called the interaural intensity (or level) difference.
- Shadowing effect of the head and shoulders as well as complex filtering function related to the twists and turns of the pinnae (outer ear). These direction-based changes in frequency distribution caused by the pinnae, along with others caused by head shadowing and reflections, are now collectively referred as head-related transfer function (HRTF).
- The echo.
- Attenuation of high frequencies for very distant objects.

Audio localization is also affected by the presence of other sounds and the direction from which these sounds origi-

nate. The relative frequency, intensity, and location of sound sources affect the degree to which one sound masks another.

However, all modes of data presentation are not perceptually acceptable. Applying sonification, one must be aware of the following difficulties of acoustic rendering (45):

- Difficult perception of precise quantities and absolute values.
- Limited spatial distribution.
- Some sound parameters are not independent (pitch depends on loudness)
- Interference with other sound sources (like speech).
- Absence of persistence.
- Dependence on individual user perception.

4.1.2.2. Nonspeech Audio. Nonspeech audio is a rich mode of interaction in many aspects of our lives. We use nonspeech audio on a daily basis, for example, when crossing the street, answering the phone, diagnosing problems with our car engine, and applauding in a theater. Nonspeech has also been used in HCI. The audio messages in contemporary user interfaces generally fall into one of three categories (46,47):

- alarms and warning systems, such as beeps on a PC;
- status and monitoring indicators; and
- encoded messages and data (e.g., sonification or data auralization).

According to how audio messages are generated and perceived, Gaver classified nonspeech audio cues in two groups (48):

- Musical listening, and
- Everyday listening.

In musical listening, the "message" is derived from the relationships among the acoustical components of the sound, such as *pitch*, *timing*, *timber*, etc. Sonification is achieved by mapping data to some of the acoustical components:

- *Pitch* is the subjective perception of frequency and represents the primary basis for traditional melody. Logarithmic changes in frequency are perceived as linear pitch change.
- *Rhythm* is the relative change in the timing of successive sound events.
- *Tempo* is the relative rate of movement.
- *Dynamics* is the variation and gradation in the volume of musical sound.
- *Timber* is the difference of spectral content and energy over time of instrument generating sounds that distinguishes it from other sounds from other instruments of the same pitch and volume. The same tone played on different instruments will be perceived differently.

- *Location* of the sound.

In everyday listening, however, sound is not perceived in this analytical way. What one hears are not mere acoustic attributes of the sound, but the higher-level perceptual properties of the sound source (46). For example, if one hears an object fall to the ground, in everyday listening, one usually pays attention to the type of the object, how heavy it was, or what material it was made of (paper, glass, or metal).

Nonspeech sound has been successfully used in various applications. For example, nonspeech sound was used to create sound graphs (49) and acoustical textures (50). Structured nonspeech audio messages called earcons were used to provide *navigational cues* in a menu hierarchy (51). Sound has also been used to provide serendipitous information, via background auditory cues, that is tied to people's physical actions in the workplace (52).

Although sound represents a valuable channel of communication, sound-based systems must be carefully designed and implemented. For example, a study on the audibility and identification of auditory alarms in the operating room and intensive care unit of a hospital identified many drawbacks of bad design (53). Medical staff had identified only 10 to 15 of the 26 alarms found in the operating room, while nurses had identified 9 to 14 of the 23 alarms found in the intensive care unit. This study indicated that too many alarms existed in critical areas, they were hard to distinguish from each other, and some alarms were "masked" (inaudible) by the noise of machinery used in normal practice and by other simultaneously occurring alarms. In addition, incorrect auditory cues can lead to negative training (37).

4.1.2.3. Speech and Natural Language Interfaces. Speech is increasingly used in contemporary HCIs as computer-mediated human-human communication, computer synthesized speech, or speech recognition. In computer-mediated communication, speech is stored in computer memory and later retrieved for playback or further processing. For example, a teleconference system records speech of participants and distributes it to other participants. This use of speech is quite straightforward because it is a human, not a machine, for whom the message is intended, so there is typically no need for speech recognition.

Speech synthesis is an important part of ubiquitous computing. New operating systems, such as Windows XP, incorporate this technology as their standard component. Speech synthesis technology takes ASCII or eXtensible Markup Language (XML) text as input, and outputs speech as represented in Fig. 3. So-called text-to-speech (TTS) technology is efficient and rather effective, as it is usually made by a set of production rules.

Natural language generation goes beyond mere voice synthesis, adding more complex semantics into HCI, such as grammar. Applications of natural languages are usually classified in one of three ways (46): *automatic translation*, which enables translation of text from one language to the other; *database query*, which allows communication

```
<voice required="Gender=Female;Age!=Child">
   <volume level="100">
      This text should be spoken
      by female child
      at volume level hundred.
   </volume>
</voice>
```

Figure 3. XML fragment for Microsoft Speech API TTS engine.

with databases in natural languages; and *information retrieval*.

4.1.3. Haptic Rendering. Haptic rendering makes use of the sense of touch. Haptic feedback has been found to significantly improve performance in virtual environments (54). For example, the problem-solving capability of chemists was significantly improved by including haptic feedback into synthetic molecular modeling systems (55).

The ability of humans to detect various mechanical properties of different objects by tactual exploration is closely related to both kinesthetic perception and cutaneous perception, and haptic displays should be designed to address both of these perceptual channels. Therefore, the human haptic sensory system is typically considered as consisting of two parts (56):

- *tactile* (*or cutaneous*) *sensing*, which refers to consciousness of stimulation to the outer surface of the body (the smoothness of the skin), and
- *kinesthetic sensing* (*or proprioception*), which refers to awareness of limb position and movement (for example, an ability to touch your nose with your eyes closed), as well as muscle tension (for example, estimation of object weights).

The haptic system is bidirectional as it includes simultaneous sensing and manipulation, and many activities, such as doing surgical operation or reading of Braille text, require the use of both the sensing aspect and the manipulation aspect of the human haptic system.

Haptic is a crucial sense for medical staff, as many medical procedures, especially in surgery, depends on motor skills of physicians. Many virtual reality medical simulators exist that incorporate haptic feedback (57). In the past two decades, force-feedback devices have played an important role in telesurgery and virtual reality systems by improving an operator's task performance and by enhancing a user's sense of telepresence. Haptic interface is also important for blind and visually impaired users.

4.1.4. Olfactory Presentation. Olfactory presentation has not been used much in HCI, but is potentially a very interesting presentation modality. Olfactory receptors provide a rich source of information to humans. The ambient smell of the physical environment can have a great importance on the creation of a sense of presence in a virtual environment. Without appropriate scents, a user may have a reduced virtual experience (58). For example, the user might be in a virtual world with flowers, but received ambient clues that could reinforce laboratory presence.

Although some commercially available solutions exist, olfactory hardware is currently limited and not widely available. However, some research has been done, including practical uses of olfactory stimulation in virtual environments for training in the medical diagnosis, firefighter training, handling of hazardous materials, entertainment, and the presentation of processes such as chemical reactions. Some work has been done in studying the propagation of and the requirements for odors in the virtual environment (59).

Olfactory presentation also has a great potential for medical applications. Medical applications, such as surgical simulations, need to "provide the proper olfactory stimuli at the appropriate moments during the procedure" and "the training of emergency medical personnel operating in the field should bring them into contact with the odors that would make the simulated environment seem more real and which might provide diagnostic information about the injuries that simulated casualty is supposed to have incurred" (60).

4.2. Multimodal Integration

Seamless integration of modalities is crucial for wider use of multimodal user interfaces. The selection of used modalities, choice of presentation modality for a selected dataset, as well as presentation synchronization and redundancy is not trivial and straightforward. In this section, we present some of integration mechanisms as well as some of the common multimodal profiles.

4.2.1. Integration Mechanisms. Blattner classifies multimodal integration mechanisms into three categories (6):

- *Frames*. A frame is usually defined as a net of nodes, which represents objects and relations among them. Using frames, for example, each mode can have its own data parser or generator that generates frame structures that are usually time-stamped for synchronization purposes. Frames were often used with early multimodal systems.
- *Neural networks*. Neural networks consist of four parts: a group of processing units, a method of interconnecting the units, a propagation rule, and a learning method. A neural network is fault-tolerant, error-correcting, and has the ability to learn by example and find rules. Neural networks are especially popular tools for neural user interfaces and brain-computer interfaces (61).
- *Agents*. An agent is an encapsulated computer system situated in some environment and capable of flexible, autonomous action in that environment in order to meet design objectives (62). Agent-oriented approaches are well-suited for developing complex, distributed systems (63). Agents can react dynamically in critical situations, increasing robustness and fault-tolerance. Agents may also be used to manage the dialogue between a computer and a user and to model the user (64).

Bernsen defines a multimodal representation as combinations of two or more *unimodal* representational modalities (65). Taxonomy of unimodal presentation modalities is initially created, classifying modalities according to whether they are linguistic or nonlinguistic, analog or nonanalog, arbitrary or nonarbitrary, and static or dynamic. He further divides modalities according to the three presentation media: *graphics*, *sound*, and *touch*. Multimodal integration is based on mapping rules that should constitute an optimal solution to the representation and exchange of information. Bernsen also suggests a combination of natural language expressions and analog representations based on their complementary properties. The natural language expressions are *focused* but *lack specificity*, whereas analog graphical representations are *specific* but *lack focus*, so their combination often has a superior presentation effect than the use of only one of these modes (17). For example, most of the figures in this text contain some pictures (analog graphics) and textual annotations (natural language expressions).

4.2.2. Common Multimodal Combinations. Successful multimodal presentation combines complementary presentation modalities, engaging different human perceptual capabilities. In this section, we describe some common multimodal combinations that tend to create a more effective multimodal presentation by combining complementary modalities. First, we talk about multimodal combination of visualization and sonification. Then, we describe multimodal combinations of visualization and haptic rendering. After that, we describe multimodal combination of haptic rendering and sonification.

4.2.2.1. Visualization and Sonification. Sound and graphics naturally complement each other. Visual and aural perceptions have their strengths and weaknesses. For example, although it is faster to speak than it is to write, it is faster to read than it is to listen to speech (46). Some guidelines about the use of audio and visual displays, according to the type of the message being presented, may be found in Buxton (46) and Deatherage (66). According to these guidelines, it is better to use auditory presentation if:

- The message is simple.
- The message is short.
- The message will not be referred to later.
- The message deals with events in time.
- The message calls for immediate action.
- The visual system of the person is overloaded.
- The receiving location is too bright or dark so adaptation integrity is necessary.
- The user's job requires continual movement.

On the contrary, it is better to use visual presentation if:

- The message is complex.
- The message is long.
- The message will be referred to later.

- The message deals with a location in space.
- The message does not call for immediate action.
- The auditory system of the person is overburdened.
- The receiving location is too noisy.
- The user's job allows a static position.

Humans are naturally skilled to perceive visual and sound clues simultaneously. For example, listening to speech is often complemented with naturally evolved lip-reading skills, which allows more efficient understanding of conversation in a noisy environment. Experiments also suggest that even simple nonspeech audio cues can improve unimodal visual environments (67).

4.2.2.2. Visualization and Haptic Feedback. The relation between haptics and vision is fundamental. Visual feedback is a helpful cue in many human motor activities. Exact models of human movement and performance are still under development. Fitts' law, a psychological model of human movement, explores the visual and haptic feedback in aimed hand movement tasks (68). Fitts found logarithmic dependency between task difficulty and the time required to complete a movement task. Fitts' law has proven one of the most robust and widely adopted models, and has been applied in diverse settings—from microscope to under water activities.

Visualization and haptic rendering are often used together in simulators and virtual reality trainers (69). Surgical simulations are probably the most common form of this multimodal combination in medicine (57).

4.2.2.3. Sonification and Haptic Feedback (Tactical Audio). The use of visualization with haptic rendering, although effective, is not always practically feasible or even possible. When the visual field is overwhelmed, audio feedback and synchronized auditory and haptic stimuli may be extremely useful (70). Tactical audio concerns the use of audio feedback for facilitating the precise and accurate positioning of an object with respect to another object (71). This multimodal combination has valuable applications in the field of computer-assisted surgery. Just as musicians use aural feedback to position their hands, surgeons could position instruments according to a preplanned trajectory, preplaced tags or cues, or anatomical models (72). In the course of the typical diagnostic surgical procedure, numerous needle placement errors occur, especially with regard to insertion depth, which may cause, for example, missing the tumor in a biopsy. Although ultrasound and other imaging modalities attempt to alleviate this problem, the nature and configuration of the equipment requires the surgeon to take his/her eyes off the patient. The use of tactile audio feedback enables the surgeon to affect a precise placement by enhancing his/her comprehension of the 3-D position of a surgical implement (72).

4.3. Human Information Processing

Effective multimodal presentation has to consider many features of human perceptual and cognitive apparatus. Many features of human perception have already been

exploited to improve or simplify presentation of data. For example, as the human vision has limited spatial and temporal resolution, frame rate and level-of-details (LOD) degradation can be used to reduce computer load, color, or spatial visual complexity in various tasks without significantly reducing human performance and perception (73,74). In audition, effects such as frequency or time masking were used to simplify rendering and storage of audio data.

However, it is also important to know how humans make higher-level cognitive processing. Cognitive sciences, linguistics, psychology, and philosophy examine multimodality from different viewpoints. For example, Gestalt psychology suggests that the global view differs greatly from perception of the individual parts. The analysis of how humans rationally experience the environment is known as phenomenology. Factors influencing our experience of the world are memory, attention, emotions, curiosity, awareness, inner thoughts, and human will.

There have been various approaches to explain how the mind works. Different structures were proposed to describe the form of information in brain, such as modules, levels, and hierarchies (75,76). There were also many more pragmatic approaches in definition of *cognitive architectures* that introduced various theories of how human cognition works. Some examples include:

- The Soar (States, Operators, and Reasoning) architecture (77),
- The ACT-R (Adaptive Control of Thought, Rational) (78),
- The EPIC (Executive-Process Interactive Control) architecture (79),
- The capacity theory (CAPS) architecture (80), and
- Cortical architecture and self-referential control for brain-like computation (81).

Many problems related to human information processing must still be resolved to enable better design of user interfaces. For example, spatial memory can be effectively used in order to improve memorizing of presented data in 2-D and 3-D user interfaces (33). Curiosity can be exploited to stimulate the user to explore an application and discover new functionality of software. Providing computers with the ability to recognize, express, and even "have" emotions creates a qualitatively new field of "affective computing" (82).

All these issues are even more important in critical applications such as those used in medical or military settings. For example, the DARPA launched the *Augmented Cognition* project (83) with the mission "to extend, by an order of magnitude or more, the information management capacity of the human-computer war fighting integral by developing and demonstrating quantifiable enhancements to human cognitive ability in diverse, stressful, operational environments." This project includes many various undertakings such us multimodal support for augmented cognition that aims to augment cognition via multiple sensory modalities.

5. TOOLS AND PLATFORMS

Many hardware and software solutions are available that can be used for multimodal presentation. In this section, we describe presentation hardware, software libraries, as well as generic platforms that may be used to develop multimodal solutions.

5.1. Presentation Hardware

Presentation hardware may be viewed as a stimulus source, modulating physical media in order to excite human sensors. According to the human sense they attack, presentation devices fall into four categories: visualization devices, audio devices, haptic feedback devices, and olfactory presentation devices.

5.1.1. Visualization Devices. Many types of visualization devices exist. The very common visualization devices are raster cathode ray tube (CRT) displays. These displays were first used with dedicated measurement devices, such as oscilloscopes. Today, CRT displays are primarily used in raster scan mode, for example, in computer monitors or television, visually laying out data on a panel with some resolution (for example, 800×600 or 1024×768 pixels). The size of these devices is usually given as the length of a diagonal, for example, 14", 15", 17", 19", or 21". Liquid crystal displays (LCDs) are increasingly used as CRT display replacements. As their depth decreases, they are produced in various sizes and are used with many pervasive devices such as cellular phones or personal digital assistants.

The virtual reality systems introduced many new visualization devices, such as the head-mounted displays (HMDs). An HMD is a helmet with one or two wide-angle screens placed in front of the eyes. HMDs usually have stereo speakers placed over the ears to immerse the user in computer generated virtual scene. HMDs also extended the user's field of view to more that 70 degrees per eye (84). For some medical applications, the isolating helmet has been replaced with 3-D glasses or helmets that allow the wearer to peer at real objects by looking down rather than only at the screens in the helmet. An interesting alternative to HMDs is the Cave Automatic Virtual Environment (CAVE). The CAVE system uses stereoscopic video projectors to display images on three surrounding walls and on the floor, and the participants wear glasses with LCD shutters to view the 3-D images (85). For example, CAVE was used for viewing geometric protein structures (86). Another interesting display technology is a virtual display. Virtual displays create the illusion of a large display when it is held close to the eye, even though the presentation device itself is quite small (87). For example, Motorola's Virtuo-Vue creates an illusion of a virtual 17-inch image that appears to be about 4 feet away, using a panel with dimensions of $5.9\,\mathrm{mm} \times 3.8\,\mathrm{mm}$. This technology can be especially useful in various small devices that cannot contain large displays, such us portable PCs, smart phones, and pagers.

5.1.2. Sound Presentation Devices. Basic sound hardware is now available for most computer systems. Sound cards on ordinary PCs are capable of producing high-quality sound. As sound presentation devices, such as headphones and speakers, are very common, we will not describe them extensively. The first sound systems used only one speaker. Better quality sound was achieved by using two speakers: a *tweeter* for high frequencies and a *woofer* for low frequencies. Initial stereo systems used two speakers (left and right channel) to facilitate sound localization.

Various other approaches, such as audio systems with 4 speakers (Quad system) and 5.1 speakers (Left, Center, Right, Surround Left, Surround Right, and Subwoofer) have also been used. However, using only two speakers, with additional signal processing, it is possible to create equally realistic sounds in so-called virtual surround sound systems (88). These systems mimic actual surround sound by exploiting the way that the auditory system perceives the direction from which a sound is coming. They can create a very realistic surround sound field from only two speakers placed in front, to the left and right, of the listener.

5.1.3. Haptic Feedback Devices. Many force-feedback devices are currently available in the open market. These devices include various force-feedback mice, gloves, styluses, as well as various force-feedback gaming peripherals such as joysticks, steering wheels, and flight yokes. Some more sophisticated examples include the Impulse Engine^{TM} (Immersion Corp., San Jose, CA) and the popular Phantom (SensAble Technologies Inc., Cambridge, MA) (56). The Phantom haptic interface is especially popular in virtual reality surgical simulators. The device has three or six degrees of freedom and uses actuators to relay resistance at about 1000 Hz. A good and comprehensive list of off-the-shelf force and feedback devices can be found in William Buxton's directory of sources for input technologies (89).

Beyond these off-the-shelf components, some specialized medical haptic feedback devices have been developed. For example, HT Medical Systems, Inc. developed a virtual sigmoidoscope to train physicians to exercise the flexible probes used to view the colon. Beyond video imaging synchronized to the position of the probe, the system incorporates haptic feedback. The simulation software warns the physician when injury to the "patient" is about to occur, and it can also rate the physician's performance (57).

5.1.4. Olfactory Presentation Devices. Although olfactory devices are not so widely available, several companies, such as Ferris Production Inc. from Arizona, and Digi-Scent Inc. from California, are active in this market. Digi-Scents developed a device called iSmell, a "personal scent synthesizer." The iSmell attaches to the serial or USB port of a computer and uses standard electrical power. The iSmell emits naturally based vapors into the user's personal space. The device is triggered either by user activation, such as keystroke, or by a timed response. The device works by using small cartridges with various scented

chemicals, which could be replaced. DigiScents also developed a ScentWare Developers Kit, which includes the tools required to create scent-enabled content and media.

5.2. Software Support

Development of multimodal interfaces requires the use of many libraries and software toolkits. Here, we present some of the often-used software platforms for development of various multimodal presentation effects, focusing on multiplatform and generic solutions.

Every operating system, such as Windows or Linux, has an extensive user interface library. For example, Windows operating systems, in addition to the standard I/O support, provides the DirectX as a set of application programming interfaces (APIs) for high-level or low-level development of user applications. DirectX has various parts dedicated to user interactions such as DirectDraw, DirectSound, and Direct3D addressing 2-D and 3-D graphics and sound as well as many other features.

OpenGL is the environment for developing portable, interactive 2-D and 3D graphics applications (90). The OpenGL was introduced in 1992, and it has been the industry's most widely used and supported 2-D and 3D graphics API. The OpenGL incorporates a broad set of rendering, texture mapping, special effects, and other visualization functions.

The Virtual Reality Modeling Language (VRML) is a general-purpose high-level platform for modeling of 3-D audio-visual scenes (91). The VRML is simple language for describing 3-D shapes and interactive environment. VRML is also intended to be a universal interchange format for integrated 3-D graphics and multimedia. VRML browsers and authoring tools for the creation of VRML files are widely available for many different platforms. In addition to 3-D visualization, the VRML allows creation of virtual worlds with 3-D audio sources, which improves immersion and realism of virtual environments. More realistic perception of sound in the VRML is achieved by implementation of simplified Head-Related Transfer Function (HRTF). Sonification is supported with standard VRML *Sound* and *AudioClip* nodes. The Web3D Consortium guides the development of the VRML and has various working groups that aim to extend the VRML in various application domains.

An increasing number of systems use the eXtensible Markup Language (XML). XML is a subset of the Standard Generalized Markup Language (SGML), a complex standard for describing document structure and content. XML is a meta-language, and it allows customized markup languages for different document classes. Some of the XML-based standards of interest to data presentation include XHTML, Scalable Vector Graphics (SVG), Synchronized Multimedia Integration Language (SMIL), and VoiceXML (92,93). Recently, the World Wide Web Consortium, which led the development of the XML, launched the Multimodal Interaction Activity to extend the Web user interface to include multiple modes of interaction such as GUI, speech, vision, pen, gestures, or haptic interfaces.

Java has been a very popular implementation environment for various presentation solutions. The main reasons

for Java's rapid acceptance are its architecture and its platform independence (94). Java supports development of multimodal presentation interfaces, including libraries such as Java Media Framework (JMF), Java2D, Java3D, and Java Sound for work with multimedia, 2-D and 3-D graphics, and sound. Java is also used as a script language for Virtual Reality Language (VRML) virtual environments (91).

6. MULTIMODAL PRESENTATION IN BIOMEDICAL APPLICATIONS

Multimodal presentation used in biomedical applications falls into one of four categories:

- *Medical diagnosis*, where multimodal presentation is used to enable more efficient exploration of biomedical datasets in order to simplify diagnosing various illnesses;
- *Medical education and training*, where multimodal presentation is used to simulate multimodal nature of standard medical procedures;
- *Surgical procedures*, where multimodal interfaces are user in surgery preparation and planning, as well as in some surgical operations; and
- *Rehabilitation and training*, where multimodal interfaces are presented to patients in order to improve their recovery.

6.1. Medical Diagnostics

Multimodal presentation could be used to improve the insight into complex biomedical phenomena. Santarelli et al. proposed a system for medical image processing, which allows interactive real-time multimodal 3-D visualization (95). The system supports medical specialists in the diagnosis of internal organs, such as the heart during the cardiac cycle, allowing them to compare information on perfusion/contraction match as a basis for diagnosis of important cardiovascular diseases. Meyer et al. implemented 3-D registration of multimodal datasets for fusion of clinical anatomic and functional imaging modalities (96). Valentino et al. proposed a procedure for combining and visualizing complementary structural and functional information from magnetic resonance imaging (MRI) and positron emission tomography (PET) (97). The techniques provide a means of presenting vast amounts of multidimensional data in a form that is easily understood, and the resulting images are essential to an understanding of the normal and pathologic states of the human brain. Beltrame et al. proposed an integrated view of structural and functional multidimensional images of the same body area of a patient (98). The integrated view provides physicians with a better understanding of the structure-to-function relationships of a given organ.

A combination of visualization and sonification has also been used to improve diagnostic procedures. Krishnan et al. proposed a technique for sonification of knee-joint vibration signals (99). They developed a tool that performs computer-assisted auscultation of knee joints by auditory display of the vibration signals emitted during active movement of the leg. Sonification algorithms transform the instantaneous mean frequency and envelope of the signals into sound in order to improve the potential diagnostic quality of detected signals.

Fast visual review of long medical records, such as a 24-hour EEG, is an extremely time-consuming and fatiguing process (100). Sonification can be used to improve seizure detection (101). Some examples of EEG sonification can be found in Ref. 102. Several commercial companies already provide sonification of EEG for biofeedback, such as Thought Technology, Ltd. (103) and Elixa MindSet (104).

Jovanov et al. designed the environment for more complex multimodal presentation of brain electrical activity (105,106). The environment, called the multimodal viewer (*mmViewer*), consists of the VRML head model and the Java applets that controls multimodal presentation (Fig. 4). The environment uses various visualization and sonification modalities, facilitating efficient perceptualization of data (5). Visualization is based on animated topographic maps projected onto the scalp of a 3-D model of the head, employing several graphical modalities, including 3-D presentation, animation, and color. As brain topographic maps used in the environment contain exact values only on electrode positions, all the other points must be spatially interpolated using score values calculated on electrode positions. Therefore, an additional visualization problem addressed in this environment is interpolation of a limited set of electrode values to much the more dense 3-D head model. The electrode setting is usually predefined (like International 10–20 standard), but as a higher number of electrodes results in more reliable topographic mapping, some experiments used custom electrode settings to increase spatial resolution over the selected brain region. Sonification is implemented as modulation of natural sound patterns to reflect certain features of processed data and create pleasant acoustic environment. This feature is particularly important for prolonged system use. The authors applied sonification to emphasize the temporal dimension of the selected visualized scores. As the topographic map by itself represents a large amount of visual information, the authors sonified global parameters of brain electrical activity, such as the global index of left/right hemisphere symmetry. The index of symmetry (IS) is calculated as:

$$IS = (P_1 - P_2)/(P_1 + P_2),$$

where P_1 and P_2 represent power of symmetrical EEG channels, like O1 and O2, for example. This parameter is sonified by changing the position of the sound source in the VRML world. Therefore, activation of a hemisphere could be perceived as movement of sound source toward that hemisphere.

6.2. Medical Education and Training

Traditional methods of medical education relied on a master-trainee system of learning, using patients and carcasses to perform procedures (107). However, the multimodal possibilities of contemporary computers and virtual real-
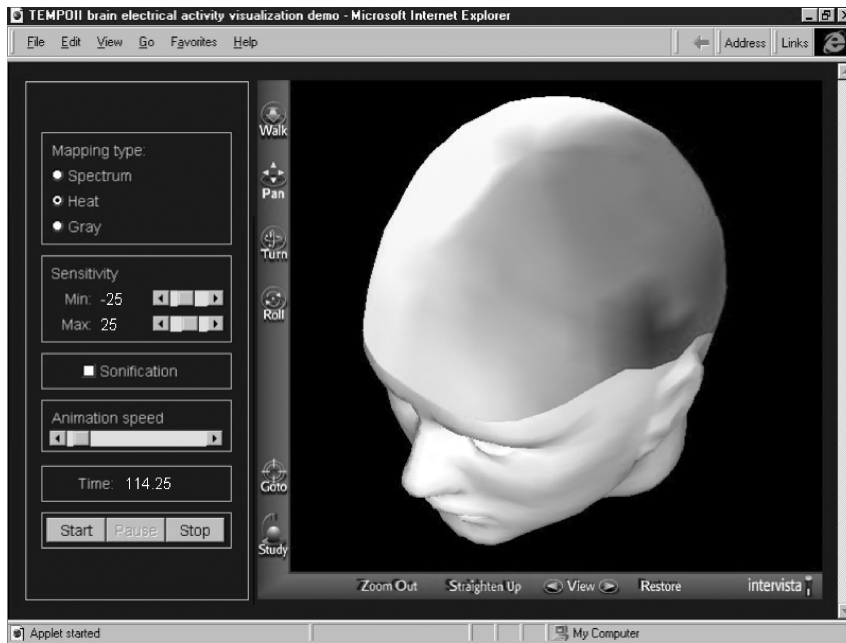
Figure 4. A multimodal viewer (mmViewer), with the VRML head model (left) and the Java applet that controls multimodal presentation (right).

ity technologies allow medical education without putting patients at risk, as complications can be safely simulated. With novel haptics interface devices, computer-based training will enable novice physicians to learn procedures in situations that closely resemble the real conditions (108). New learning environments can also give objective scores of a student's ability. Studies showed that computer-based training simulations are at least as good as standard training methods (57).

The most highly developed multimodal training environments are surgical simulations. The types of simulations range from "needle-based" procedures, such as standard intravenous insertion, central venous placement catheter, and chest-tube insertion, to more sophisticated simulations of full surgical procedures like laparoscopic cholecystectomy or hysteroscopic resection of interuterine myoma (109). For example, Basdogan et al. develop a computer-based training system to simulate laparoscopic procedures in virtual environments for medical training (110). The system used a computer monitor to display visual interactions between 3-D virtual models of organs and instruments, and, with a pair of force-feedback devices, interfaced with laparoscopic instruments to simulate haptic interactions. The environment simulates a surgical procedure that involves inserting a catheter into the cystic duct using a pair of laparoscopic forceps. With the proposed system, the user can be trained to grasp and insert a flexible and freely moving catheter into the deformable cystic duct in virtual environments. Burdea et al. proposed a virtual reality-based training system for the diagnosis of prostate cancer (111). The simulator consists of a phantom haptic interface that provides feedback to the trainee's index finger, a motion restricting board, and graphical workstation, which renders the patient's anatomy. The system provides several models of prostates: normal, enlarged with no tumor, incipient malignancy (single tumor), and advanced malignancy (tumor cluster).

6.3. Surgical Procedures

Multimodal presentation has been used in various stages of surgical procedures. Peters et al. proposed the use of multimodal imaging as an aid to the planning and guidance of neurosurgical procedures (112). They discuss the integration of anatomical, vascular, and functional data for presentation to the surgeon during surgery. Wong et al. designed a neurodiagnostic workstation for epilepsy surgery planning by combining biomedical information, from multiple noninvasive image modalities, such as functional PET, MRS, and MEG information, with structural MRI anatomy (113). Sclabassi et al. proposed a neurophysiological monitoring system that supports remote performance through real-time multimodal data processing and multimedia network communication. The system combines real-time data sources, including all physiological monitoring functions, with non real-time functions and online databases (114).

Another promising application of multimodal interfaces, especially haptic feedback, lies in a minimally invasive surgery and telesurgery. Here, a surgeon is not able to directly observe the surgery, but has to rely on some kind of a user interface. However, the main problem here is that these new user interfaces create visual and haptic distortion when compared with conventional surgery. Some research has occurred that addressed this problem. For example, Rosen et al. developed a computerized force-feedback endoscopic surgical grasper (FREG) in order to regain the tactile and kinesthetic information that is lost (115). The system uses standard unmodified grasper shafts and tips. The FREG can control grasping forces either by surgeon teleoperation control or under software control. The FREG may provide the basis for application in telesurgery, clinical endoscopic surgery, surgical training, and research. Bicchi et al. investigated the possibility of substituting detailed tactile information for softness

discrimination, with information on the rate of spread of the contact area between the finger and the specimen as the contact force increases. They also developed a practical application to a mini-invasive surgery tool (116).

6.4. Rehabilitation and Medical Treatment

Multimodal presentation has been frequently applied in rehabilitation and medical treatment as biofeedback. Biofeedback systems are closed-loop systems that detect biomedical changes and present them back to the patient in real-time. Interfaces in existing biofeedback applications range from interactive 2-D graphical tasks—in which muscle signals, for example, are amplified and transformed into control tasks such as lifting a virtual object or typing—to real-world physical tasks such as manipulating radio-controlled toys (117). The interfaces to physical rehabilitation biofeedback systems may amplify weak muscle signals, encouraging patients to persist when a physical response to therapy occurs that is generally not visible without magnification (11). For example, Popescu et al. developed a PC-based orthopedic home rehabilitation system using a personal computer, a Polhemus tracker, and a multipurpose haptic control interface that provides resistive forces using the Rutgers Master II (RMII) glove (118). The system has a library of virtual rehabilitation routines that allows a patient to monitor its progress and repeat evaluations over time. Other haptic interfaces currently under development in the same group include devices for elbow and knee rehabilitation connected to the same system. Bardorfer et al. proposed an objective test for evaluating the functional studies of the upper limbs in patients with neurological diseases (119). They employed a method that allows assessment of kinematic and dynamic motor abilities of upper limbs. They created a virtual environment using a computer display for visual information and a phantom haptic interface. The haptic interface is used as a kinematic measuring device and for providing tactile feedback to the patient. By moving the haptic interface control stick, the patient was able to move the pointer (a ball) through the labyrinth in three dimensions and to feel the reactive forces of the wall. It has been applied to healthy subjects and patients with various forms of neurological diseases, such us Friedreich Ataxia, Parkinson's disease, and Multiple Sclerosis.

Health-care providers are increasingly using brainwave biofeedback or neurofeedback as part of the treatment of a growing range of psychophysiological disorders such as attention deficit/hyperactivity disorder (ADHD), post-traumatic stress disorder, addictions, anxiety, and depression. In these applications, surface-mounted electrodes detect the brain's electrical activity, and the resulting electroencephalogram (EEG) is presented in real-time as abstract images and sounds. Using this data in reward/response-based control tasks generates increased or reduced activity in different parts of the EEG spectrum to help ameliorate these psychophysiological disorders (11).

7. CONCLUSION

The multidimensional nature and complexity of biomedical data require innovative HCIs. Multimodal presentation holds the promise of increasing the perceptual bandwidth of human-computer communications. Information overload on one sense causes tiredness and reduced attention, so the balance between various modalities can reduce the cognitive load. In addition, multimodal communication is a *natural* feature of human communication, and users have a strong preference to interact multimodally rather than unimodally. Using multiple modalities, a user can *choose* preferred modalities and customize presentation. Multimodal communication is also more *robust* and less error-prone. However, interrelated factors affect the multimodal interface development because of the nature of high-level cognitive processing. Recent experimental evaluations of multimodal interfaces for typical biomedical data collection tasks have shown that perceptual structure of a task must be taken into consideration when designing a multimodal computer interface (120).

Although many studies have shown that computer-based multimodal training simulations are at least as good as standard training methods, it was estimated that only 1% of medical students in the United States receive some type of computerized or virtual reality training (57). However, with the increasing processing and presentation capabilities of ordinary computers, this situation is changing, and we can also expect wider use of multimodal presentation techniques in various biomedical applications.

BIBLIOGRAPHY

1. L. Kohn, J. Corrigan, and M. Donaldson, eds., To Err Is Human: Building a Safer Health System. Washington, DC: Committee on Quality of Health Care in America, Institute of Medicine, National Academy Press, 1999.

2. M. Turk and G. Robertson, Perceptual user interfaces (introduction). *Communications of the ACM* 2000; **43**(3):33–35.

3. P. C. Wong, Visual data mining. *IEEE Comp. Graph. Applicat*. 1999; **20**(5):20–21.

4. S. Oviatt and P. Cohen, Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM* 2000; **43**(3):45–53.

5. E. Jovanov, D. Starcevic, and V. Radivojevic, Perceptualization of biomedical data. In: M. Akay and A. Marsh, eds., *Information Technologies in Medicine, Volume I: Medical Simulation and Education*. New York: John Wiley and Sons, 2001, pp. 189–204.

6. M. M. Blattner and E. P. Glinter, Multimodal integration. *IEEE Multimedia* 1996; **3**(4):14–24.

7. R. Prates, C. de Souza, and S. Barbosa, A method for evaluating the communicability of user interfaces. *Interactions* 2000; **7**(1):31–38.

8. E. Jovanov, D. Starcevic, V. Radivojevic, A. Samardzic, and V. Simeunovic, Perceptualization of biomedical data. *IEEE Eng. Med. Biol. Mag*. 1999; **18**(1):50–55.

9. F. Nebeker, Golden accomplishments in biomedical engineering. *IEEE Eng. Med. Biol. Mag*. 2002; **21**(3):17–47.

10. R. L. Schoenfeld, From Einthoven's galvanometer to single-channel recording. *IEEE Eng. Med. Biol. Mag.* 2002; **21**(3):90–96.

11. J. Allanson, Electrophysiologically interactive computer systems. *Computer* 2002; **35**(3):60–65.

12. J. Hassett, A *Primer of Psychophysiology*. New York: W.H. Freeman and Company, 1978.

13. R. A. Robb, *Biomedical Imaging, Visualization, and Analysis*. New York: John Wiley & Sons, 1999.

14. B. Reeves and C. Nass, Perceptual user interfaces: perceptual bandwidth. *Communications of the ACM* 2000; **43**(3):65–70.

15. S. T. C. Wong and H. K. Huang, Networked multimedia for medical imaging. *IEEE Multimedia* 1997; **4**(2):24–35.

16. J. L. Paul and J. C. Lupo, From tanks to tumors. *IEEE Eng. Med. Biol. Mag.* 2002; **21**(6):34–35.

17. N. O. Bernsen, Foundations of multimodal representations: a taxonomy of representational modalities. *Interacting with Computers* 1994; **6**:347–371.

18. S. Oviatt, Ten myths of multimodal interaction. *Communications of the ACM* 1999; **42**(11):74–81.

19. S. Oviatt, Taming recognition errors with a multimodal interface. *Communications of the ACM* 2000; **43**(9):45–51.

20. J. Hochberg, *Perception*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1978.

21. E. R. Tufte, *Envisioning Information*. Cheshire, CT: Graphics Press, 1990.

22. J. Albers, *The Interaction of Color*, revised ed. New Haven, CT: Yale University Press, 1975.

23. J. S. Donath, Inhabiting the virtual city: the design of social environments for electronic communities, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1997.

24. N. Jacobson and W. Bender, Color as a determined communication. *IBM Syst. J.* 1996; **35**:526–538.

25. S. A. Douglas and A. E. Kirkpatrick, Model and representation: the effect of visual feedback on human performance in a color picker interface. *ACM Trans. Graph.* 1999; **18**(2):96–127.

26. C. Rigden, Now you see it, now you don't. *Computer* 2002; **25**(7):104–105.

27. L. Bartram, Perceptual and interpretive properties of motion for information visualization. Proceedings of the workshop on New paradigms in information visualization and manipulation, ACM Press, 1998: 3–7.

28. P. McLeod, J. Driver, Z. Dienes, and J. Crisp, Filtering by movement in visual search. *J. Exper. Psychol. Human Percept. Perform.* 1991; **17**(1):55–64.

29. M. Bayles, Designing online banner advertisements: should we animate? Proceedings of the Conference on Human Factors and Computing Systems—CHI 2002, ACM Press, 2002:363–368.

30. R. S. Kalawsky, *The Science of Virtual Reality and Virtual Environments*. Wokingham, UK: Addison-Wesley, 1993.

31. G. S. Hubona, P. N. Wheeler, G. W. Shirah, and M. Brandt, The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization. *ACM Trans. Computer-Human Interact.* 1999; **6**(3):214–242.

32. J. C. Rodger and R. A. Browse, Choosing rendering parameters for effective communication of 3D shape. *IEEE Comp. Graph. Appl.* 2000; **20**(2):20–28.

33. G. Robertson, M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, and M. van Dantzich, Data mountain: using spatial memory for document management. Symposium on User Interface Software and Technology—UIST 1998, ACM Press, 1998: 153–162.

34. A. Cockburn and B. McKenzie, 3D or not 3D?: evaluating the effect of the third dimension in a document management system. Conference on Human Factors and Computing Systems—CHI 2001, ACM Press, 2001: 434–441.

35. D. B. Anderson and M. A. Casey, The sound dimension. *IEEE Spectrum* 1997; **34**(3):46–51.

36. T. Lokki, L. Savioja, R. Väänänen, J. Huopaniemi, and T. Takala, Creating interactive virtual auditory environments. *IEEE Comp. Graph. Appl.* 2002; **22**(4):49–57.

37. N. Tsingos, I. Carlbom, G. Elko, R. Kubli, and T. Funkhouser, Validating acoustical simulations in the Bell Labs box. *IEEE Comp. Graph. Appl.* 2002; **22**(4):28–37.

38. S. Flinn and K. S. Booth, The creation, presentation and implications of selected auditory illusions. Computer Science Technical Report, TR-95-15, The University of British Columbia, 1995.

39. T. Painter and A. Spanias, Perceptual coding of digital audio. *Proc. IEEE* 2000; **88**(4):451–515.

40. F. Biocca, Virtual reality technology: a tutorial. *J. Commun.* 1992; **42**(4):23–72.

41. W. Koenig, Subjective effects in binaural hearing. *J. Acoust. Soc. Am.* 1950; **22**(1):61.

42. J. C. Middlebrooks and D. M. Green, Sound localization by human listeners. *Ann. Rev. Psychol.* 1991; **42**:135–159.

43. D. R. Begault and E. M. Wenzel, Headphone localization of speech. *Human Factors* 1993; **35**(2):361–376.

44. P. R. Cook, Sound production and modeling. *IEEE Comp. Graph. Appl.* 2002; **22**(4):23–27.

45. G. Kramer, ed., *Auditory Display, Sonification, Audification and Auditory Interfaces*. Reading, MA: Addison-Wesley, 1994.

46. W. Buxton, Speech, language & audition. In: R. M. Baecker, J. Grudin, W. Buxton, and S. Greenberg, eds., *Readings in Human Computer Interaction: Toward the Year 2000*. San Francisco, CA: Morgan Kaufmann Publishers, 1995.

47. C. Demarey and P. Plénacoste, User, sound context and use context: what are their roles in 3D sound metaphors design? Proc. 2001 International Conference on Auditory Display, 2001: 136–140.

48. W. Gaver, Auditory icons: using sound in computer interfaces. *Human Computer Interact.* 1986; **2**(2):167–177.

49. T. L. Bonebright, M. A. Nees, T. T. Connerley, and G. R. McCain, Testing the effectiveness of sonified graphs for education: a programmatic research project. Proc. 2001 International Conference on Auditory Display, Espoo, Finland, 2001: 62–66.

50. S. Dubnov, Z. Bar-Joseph, R. El-Yaniv, D. Lischinski, and M. Werman, Synthesizing sound textures through wavelet tree learning. *IEEE Comp. Graph. Appl.* 2002; **22**(4):38–48.

51. S. A. Brewster, Using nonspeech sounds to provide navigation cues. *ACM Trans. Computer-Human Interact.* 1998; **5**(3):224–259.

52. E. D. Mynatt, M. Back, R. Want, M. Baer, and J. B. Ellis, Designing audio aura. Conference on Human Factors and Computing Systems—CHI 1998 , ACM Press and Addison-Wesley Publishing, 1998: 566–573.

53. K. Momtahan, R. Hétu, and B. Tansley, Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics* 1993; **36**(10):1159–1176.

54. G. C. Burdea, *Force & Touch Feedback for Virtual Reality*. New York: John Wiley & Sons, 1996.

55. F. Brooks, M. Ouh-Young, J. Batter, and P. Kilpatrick, Project GROPE—haptic displays for scientific visualization. *ACM Comp. Graph*. 1990; **24**(3):177–185.

56. H. Z. Tan, Perceptual user interfaces: haptic interfaces. *Communications of the ACM* 2000; **43**(3):40–41.

57. D. Sorid and S. K. Moore, The virtual surgeon. *IEEE Spectrum* 2000; **37**(7):26–31.

58. J. P. Cater, The nose have it! Letters to the editor. *Presence* 1992; **1**(4):493–494.

59. W. Barfield and E. Danas, Comments on the use of olfactory displays for virtual environments. *Presence* 1995; **5**(1):109–121.

60. M. W. Krueger, Olfactory stimuli in virtual reality for medical applications. In: R. M. Satava, K. Morgan, H. B. Sieburg, R. Mattheus, and J. P. Christensen, eds., *Interactive Technology and the New Paradigm for Healthcare*. Washington DC: IOS Press, 1995, pp. 180–181.

61. E. Haselsteiner and G. Pfurtscheller, Using time-dependent neural networks for EEG classification. *IEEE Trans. Rehab. Eng*. 2000; **8**(4):457–463.

62. D. B. Moran, A. J. Cheyer, L. E. Julia, D. L. Martin, and S. Park, Multimodal user interfaces in the open agent architecture. Proc. Intelligent User Interfaces–IUI 97, ACM Press, 1997: 61–68.

63. Z. Obrenovic, D. Starcevic, E. Jovanov, and V. Radivojevic, An agent based framework for virtual medical devices. Proc. First International Joint Conference on Autonomous Agents & Multiagent Systems—AAMAS 2002, ACM Press, 2002: 659–660.

64. H. Lieberman, Autonomous interface agents. Conference on Human Factors and Computing Systems—CHI 1997, ACM Press, 1997: 67–74.

65. N. O. Bernsen, Why are analogue graphics and natural language both needed in HCI? In: F. Paterno, ed., *Design*, *Specification and Verification of Interactive Systems*. New York: Springer-Verlag, 1994, pp. 165–179.

66. B. H. Deatherage, Auditory and other sensory forms of information presentation. In: H. P. Van Cott and R. G. Kinkade, eds., *Human Engineering Guide to Equipment Design*, rev. ed. Washington, DC: U.S. Government Printing Office, 1972.

67. Z. Obrenovic, D. Starcevic, and E. Jovanov, Toward optimization of multimodal user interfaces for tactile audio applications. In: N. Carbonell and C. Stephanidis, eds., *Universal Access Theoretical Perspectives, Practice, and Experience-Lecture Notes in Computer Sciences 2615*. New York: Springer, 2003, pp. 287–298.

68. I. S. MacKenzie, Fitts' law as a research and design tool in human computer interaction. *Human-Computer Interact*. 1992; **7**:91–139.

69. M. Srinivasan and C. Basdogan, Haptics in virtual environments: taxonomy, research status, and challenges. *Computers & Graphics* 1997; **21**(4):393–404.

70. D. DiFilippo and D. K. Pai, The AHI: an audio and haptic interface for contact interactions. Symposium on User Interface Software and Technology, ACM Press, 2000: 149–158.

71. K. Wegner and D. Karron, Surgical navigation using audio feedback. In: K. S. Morgan, et al., eds., *Medicine Meets Virtual Reality: Global Healthcare Grid*. Washington, DC: IOS Press, Ohmsha, 1997, pp. 450–458.

72. E. Jovanov, Z. Obrenovic, D. Starcevic, and D. B. Karron, A virtual reality training system for tactical audio applications. SouthEastern Simulation Conference SESC'99, Huntsville, Alabama, 1999.

73. M. Reddy, Perceptually modulated level of detail for virtual environments, Ph.D. thesis, University of Edinburgh, Edinburgh, Scotland, 1997.

74. B. Watson, N. Walker, L. F. Hodges, and A. Worden, Managing level of detail through peripheral degradation: effects on search performance with a head-mounted display. *ACM Trans. Computer-Human Interact*. 1997; **4**(4):323–346.

75. J. A. Fodor, *The Modularity of Mind*. Cambridge, MA: MIT Press, 1983.

76. R. Jackendoff, *Consciousness and the Computationals Mind*. Cambridge, MA: MIT Press, 1989.

77. J. E. Laird, A. Newell, and P. S. Rosenbloom, Soar: an architecture for general intelligence. *Artific. Intell*. 1987; **33**:1–64.

78. J. R. Anderson, *Rules of the Mind*. New York: Hillsdale, Lawrence Erlbaum, 1993.

79. D. E. Kieras, S. D. Wood, and D. E. Meyer, Predictive engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task. *ACM Trans. Computer-Human Interact*. 1997; **4**(3):230–275.

80. M. A. Just and P. A. Carpenter, A capacity theory of comprehension: individual differences in working memory. *Psychological Rev*. 1992; **1**:122–149.

81. E. Korner and G. Matsumoto, Cortical architecture and self-referential control for brain-like computation. *IEEE Eng. Med. Biol*. 2002; **21**(5):121–133.

82. R. W. Picard, *Affective Computing*. Cambridge, MA: MIT Press, 1997.

83. DARPA Information Technology Office (2003), *DARPA Augmented Cognition Project Web Site* (online). Available: *http://www.darpa.mil/ito/research/ac/index.html*.

84. K. M. Stanney, R. R. Mourant, and R. S. Kennedy, Human factors issues in virtual environments: a review of the literature. *Presence* 1998; **7**(4):327–351.

85. C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart, The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM* 1992; **35**(6):65–72.

86. N. Akkiraju, H. Edelsbrunner, P. Fu, and J. Qian, Viewing geometric protein structures from inside a CAVE. *IEEE Comp. Graph. Appl*. 1996; **16**(4):58–61.

87. J. Edwards, New interfaces: making computers more accessible. *Computer* 1997; **30**(12):12–14.

88. A. Kraemer, Two speakers are better than 5.1. *IEEE Spectrum* 2001; **38**(5):71–74.

89. W. Buxton (2004), *A Directory of Sources for Input Technologies* (online). Available: *http://www.billbuxton.com/InputSources.html*.

90. D. Shreiner, ed., *Opengl Architecture Review Board, OpenGL Reference Manual*, 3rd ed. Reading, MA: Addison-Wesley, 1999.

91. The VRML Consortium Incorporated, Information technology—Computer graphics and image processing—The Virtual Reality Modeling Language (VRML)—Part 1: Functional specification and UTF-8 encoding. International Standard ISO/IEC 14772-1:1997, 1997. Available: *http://www.vrml.org/Specifications/VRML97/index.html*.

16   **MULTIMODAL PRESENTATION OF BIOMEDICAL DATA**

92. H. W. Lie and J. Saarela, Multipurpose Web publishing using HTML, XML, and CSS. *Communications of the ACM* 1999; **42**(10):95–101.

93. B. Lucas, VoiceXML for Web-based distributed conversational applications. *Communications of the ACM* 2000; **43**(9):53–57.

94. T. Jepsen, Don't know beans? Try Java. *IT Professional* 2000; **2**(2):29–35.

95. M. F. Santarelli, V. Positano, and L. Landini, Real-time multimodal medical image processing: a dynamic volume-rendering application. *IEEE Trans. Information Technol. Biomed*. 1997; **1**(3):171–178.

96. C. R. Meyer, G. S. Leichtman, J. A. Brunberg, R. L. Wahl, and L. E. Quint, Simultaneous usage of homologous points, lines, and planes for optimal, 3-D, linear registration of multimodality imaging data. *IEEE Trans. Med. Imag*. 1995; **14**(1):1–11.

97. D. J. Valentino, J. C. Mazziotta, and H. K. Huang, Volume rendering of multimodal images: application to MRI and PET imaging of the human brain. *IEEE Trans. Med. Imag*. 1991; **10**(4):554–562.

98. F. Beltrame, G. Marcenaro, and F. Bonadonna, Integrated imaging for neurosurgery. *IEEE Eng. Med. Biol. Mag*. 1992; **11**(1):51–56.

99. S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank, Sonification of knee-joint vibration signals. *Proc. 22nd, Annual International Conference of the IEEE* 2000; **3**:1995–1998.

100. F. Pauri, F. Pierelli, G. E. Chatrian, and W. W. Erdly, Long-term EEG-video-audio monitoring: computer detection of focal EEG seizure patterns. *Electroencephalogram Clin. Neurophysiol*. 1992; **82**(1):1–9.

101. T. Hermann, P. Meinicke, H. Bekel, H. Ritter, H. M. Muller, and S. Weiss, Sonification for EEG data analysis. Proc. 2002 International Conference on Auditory Display, Kyoto, Japan, July 2002: 2–5.

102. Examples of acoustic remapping of data: Sonification of EEG data (2004) (online). Available: *http://ccrma.stanford.edu/groups/soni/eeg.html*.

103. Thought Technology, Ltd (2004) (online). Available: *http://www.thoughttechnology.com/*.

104. MindSite 16 Channel EEG Neurofeedback (2004) (online). Available: *http://www.elixa.com/mental/Mindset-Specs.htm*.

105. E. Jovanov, D. Starčević, A. Marsh, A. Samardžić, Z. Obrenović, and V. Radivojević, Multi modal viewer for telemedical applications. 20th Annual Int'l Conf. IEEE Engineering in Medicine and Biology, Hong Kong, 1998.

106. E. Jovanov, D. Starčević, A. Samardžić, A. Marsh, and Z. Obrenović, EEG analysis in a telemedical virtual world. *Future Gen. Comp. Syst*. 1999; **15**:255–263.

107. S. L. Dawson and J. A. Kaufman, The imperative for medical simulation. *Proc. IEEE* 1998; **86**(3):479–483.

108. M. Akay and A. Marsh, eds., *Information Technologies in Medicine, Volume I: Medical Simulation and Education*. New York: John Wiley & Sons, 2001.

109. R. M. Satava and S. B. Jones, Current and future applications of virtual reality for medicine. *Proc. IEEE* 1998; **86**(3):484–489.

110. C. Basdogan, C.-H. Ho, and M. A. Srinivasan, Virtual environments for medical training: graphical and haptic simulation of laparoscopic common bile duct exploration. *IEEE/ASME Trans. Mechatron*. 2001; **6**(3):269–285.

111. G. Burdea, G. Patounakis, V. Popescu, and R. E. Weiss, Virtual reality-based training for the diagnosis of prostate cancer. *IEEE Trans. Biomed. Eng*. 1999; **46**(10):1253–1260.

112. T. Peters, B. Davey, P. Munger, R. Comeau, A. Evans, and A. Olivier, Three-dimensional multimodal image-guidance for neurosurgery. *IEEE Trans. Med. Imag*. 1996; **15**(2):121–128.

113. S. T. C. Wong, R. C. Knowlton, R. A. Hawkins, and K. D. Laxer, Multimodal image fusion for noninvasive epilepsy surgery planning. *IEEE Comp. Graph. Appl*. 1996; **16**(1):30–38.

114. R. J. Sclabassi, D. Krieger, R. Simon, R. Lofink, G. Gross, and D. M. DeLauder, NeuroNet: collaborative intraoperative guidance and control. *IEEE Comp. Graph. Appl*. 1996; **16**(1):39–45.

115. J. Rosen, B. Hannaford, M. P. MacFarlane, and M. N. Sinanan, Force controlled and teleoperated endoscopic grasper for minimally invasive surgery-experimental performance evaluation. *IEEE Trans. Biomed. Eng*. 1999; **46**(10):1212–1221.

116. A. Bicchi, E. P. Scilingo, and D. De Rossi, Haptic discrimination of softness in teleoperation: the role of the contact area spread rate. *IEEE Trans. Robot. Automat*. 2000; **16**(5):496–504.

117. J. Charles, Neural interfaces link the mind and the machine. *Computer* 1999; **32**(1):16–18.

118. V. G. Popescu, G. C. Burdea, M. Bouzit, and V. R. Hentz, A virtual-reality-based telerehabilitation system with force feedback. *IEEE Trans. Information Technol. Biomed*. 2000; **4**(1):45–51.

119. A. Bardorfer, M. Munih, A. Zupan, and A. Primozic, Upper limb motion analysis using haptic interface. *IEEE/ASME Trans. Mechatron*. 2001; **6**(3):253–260.

120. M. A. Grasso, D. S. Ebert, and T. W. Finin, The integrality of speech in multimodal interfaces. *ACM Trans. Computer-Human Interact*. 1998; **5**(4):303–325.

KEYWORDS

multimodal presentation, human-computer interaction, user interfaces, perceptualization, sonification, haptic, olfactory

ABSTRACT

Insights into multidimensional phenomena, characteristic for biomedical applications, necessitates human-computer interfaces beyond standard graphical user interfaces (GUIs). Information overload in one sense, such as multiple dynamically changing windows in GUIs, may cause fatigue and reduced attention. Multimodal presentations use different sensory modalities in addition to visual presentation, such as audio and tactile feedback, to further engage human perceptual and cognitive skills. Multimodal presentation is a promising new direction toward increased perceptual bandwidth of human-computer interaction. Moreover, multimodal communication is natural feature of human communication. Therefore, most users prefer multimodal computer interaction. However, design of multimodal systems is a complex issue, and system designers must keep in mind that the nature of human perception and sensory interactions could also create false perceptions. In this chapter, we present a survey of the main presentation modalities and their characteristics, outline system design using available tools and platforms, and introduce examples of multimodal biomedical applications.

Queries and/or remarks

No Author Query