An Agenda for Human-Computer Interaction Research: Interaction Styles and Input/Output Devices

Robert J.K. Jacob

Human-Computer Interaction Lab
Naval Research Laboratory
Washington, D.C.

John J. Leggett

Department of Computer Science
Texas A&M University
College Station, Tex.

Brad A. Myers

School of Computer Science
Carnegie Mellon University
Pittsburgh, Pa.

Randy Pausch

Computer Science Department
University of Virginia
Charlottesville, Va.

INTRODUCTION

The bottleneck in improving the usefulness of interactive systems increasingly lies not in performing the processing task itself but in communicating requests and results between the system and its user. The best leverage for progress in this area therefore now lies at the user interface, rather than the system internals. Faster, more natural, and more convenient means for users and computers to exchange information are needed. On the user's side, interactive system technology is constrained by the nature of human communication organs and abilities; on the computer side, it is constrained only by input/output devices and methods that we can invent. The challenge before us is to design new devices and types of dialogues that better fit and exploit the communication-relevant characteristics of humans.

The problem of human-computer interaction can be viewed as two powerful information processors (human and computer) attempting to communicate with each other via a narrow-bandwidth, highly constrained interface (Tufte, 1989). Research in this area attempts to increase the useful bandwidth across that interface. Faster, more natural––and particularly less sequential, more parallel––modes of user-computer communication will help remove this bottleneck.

Current technology has been stronger in the computer-to-user direction (*output* devices in computer science terms, which we will use hereinafter, but *input* devices in psychological terms) than user-to-computer, so user-computer dialogues are typically one-sided, with the bandwidth from the computer to the user far greater than that from user to computer. We are especially interested in input media that can help redress this imbalance by obtaining data from the user conveniently and rapidly. Input is a neglected field relative to output, particularly considering the great strides made in computer graphics, but for that reason it is also an area that provides the greatest opportunity for research and progress. Interestingly, however, input is the source of perhaps the single biggest success story for theoretically-based research in HCI––the mouse (Card, 1978).

A basic research program in interactive systems provides the opportunity to go beyond the immediate concerns of developing the user interface for each particular new system and allows the study of HCI problems that pervade many future systems, and thus such research can have far greater leverage in the long run. The relationship between basic research and application development in this area ideally forms a circular chain: specific interface problems encountered in applications are generalized and then solved in basic research by inventing new interaction modes or techniques; and these general approaches can then be applied to the development of specific user interfaces. Finally, it is important to remember that modifying the user interface of a system, particularly the characteristics of its input/output media is not simply tinkering with the superficial appearance of the system. It changes the users' fundamental perception of a system (Ledgard, 1980). Interaction research goes to the heart of how people comprehend and work with computer systems, not simply the surface.

RESEARCH OVERVIEW

A Framework for Research in Input/Output Devices and Interaction Styles

Research in the field of input/output for human-computer interaction centers around the two ends of the communication channel:

- the devices and techniques computers can use for communicating with people, and
- the perceptual abilities, processes, and organs people can use for communicating with computers.

Attempts are then made to find the common ground through which the two can be related. The basic approach is to study new modes of communication that could be used for human-computer communication and simultaneously to develop devices and techniques to use such modes. The research paradigm is to invent new interaction techniques, implement them in hardware and software, and then study them experimentally to see their effect. In an ideal world, the development of new input/output devices would be motivated or guided by the studies of human perceptual facilities and effectors as well as the needs uncovered in studies of existing interfaces. More often, though, the hardware developments come first, people simply attempt to build "whatever can be built," and then HCI researchers try to find uses for the resulting artifacts.

We propose a "demand pull" rather than "technology push" approach to input/output devices−−where the "demand" is the range of human abilities. We wish to move from a "gadget-driven" paradigm to a "human-driven" one. The study of human characteristics in a particular user-communication setting should lead researchers to say, "we need a device that could transduce *this* particular human action to computer input," and that kind of information should guide the deployment of efforts in input device research. It must of course be tempered by informed judgments as to which such requests are reasonable and which are pure science fiction. The goal is for research in input and output devices to start with studies of the characteristics of human communication channels and skills and then develop devices that communicate effectively to and from those channels.

As with other areas of HCI research and design, it is helpful to build on the equipment and skills humans have acquired through evolution and experience and search for ways to apply them to communicating with a computer. Direct manipulation interfaces have enjoyed great success, particularly with new users, largely because they draw on analogies to existing human skills (pointing, grabbing, moving objects in space), rather than trained behaviors. As another example, research on eye movement-based computer input at NRL has tried to make use of natural eye movements (Jacob, 1990). Because eye movements are so different from conventional computer inputs, the overall approach in designing interaction techniques was, wherever possible, to obtain information from a user's natural eye movements while viewing the screen rather than requiring the user to make specific trained eye movements to actuate the system. This work began by studying the characteristics of natural eye movements and then attempted to recognize appropriate patterns in the raw data obtainable from an oculometer, turn them into tokens with higher-level meaning, and design interaction techniques for them around the known characteristics of eye movements.

Interaction techniques provide a useful focus for this type of research because they are specific, yet not bound to a single application. An interaction technique is a way of using a physical input/output device to perform a generic task in a human-computer dialogue (Foley, 1990). It represents an abstraction of some common class of interactive task, for example, choosing one of several objects shown on a display screen. Research in this area studies the primitive elements of human-computer dialogues, which apply across a wide variety of individual applications. The goal is to add new, high-bandwidth methods to the available store of input/output devices, interaction techniques, and generic dialogue components. Mockups of such techniques are then studied by measuring their properties, and attempts are made to determine their composition rules.

The Next Interaction Style

Beyond this general framework, and beyond improvements in particular device technologies, a trend toward a new class of input/output device and human-computer interaction style is emerging. Its effect on the field of input devices specifically is to move from providing objects for the user to *actuate* through specific commands to simply *sensing* the user's body. Nielsen describes this next generation interaction style as non-command-based:

> "The fifth generation user interface paradigm seems to be centered around non-command-based dialogues. This term is a somewhat negative way of characterizing a new form of interaction but so far, the unifying concept does seem to be exactly the abandonment of the principle underlying all earlier paradigms: That a dialogue has to be controlled by specific and precise commands issued by the user and processed and replied to by the computer. The new interfaces are often not even dialogues in the traditional meaning of the word, even though they obviously can be analyzed as having some

> dialogue content at some level since they do involve the exchange of information between a user and a computer. The principles shown at CHI'90 which I am summarizing as being non-command-based interaction are eye tracking interfaces, artificial realities, play-along music accompaniment, and agents." (Nielsen, 1990)

This new style of input will require new devices, new types of interaction techniques, and new software approaches to deal with them. Unlike traditional inputs, such as keyboards and mice, the new inputs represent less intentional actuation of a device or issuance of a command, but are more like passive monitoring of the user. For that reason the inputs tend to occur concurrently with one another, as different aspects of the user are monitored. From the point of view of research in input devices, this style requires a change from conventional devices to passive equipment that senses the user. Examples of such devices are unobtrusive three-dimensional trackers, hand-measuring devices, remote cameras (plus appropriate pattern recognition), range cameras, eye movement monitors, and even physiological monitors. The "natural" qualities of these input devices are well matched to new output technologies such as stereoscopic displays, head-coupled displays, and directional, non-speech audio.

Common Characteristics of the Emerging Interaction Styles

Before discussing specific areas of research, we will consider the overall environment that is on the horizon. The new non-command-based style can be characterized as based on observation of the user rather than user actuated. This section enumerates some of the common characteristics of these new interfaces and the requirements they impose on their software, hardware, and input/output devices (Green, 1991).

These interfaces are more highly interactive than their predecessors. That is, they attempt to avoid restricting the user's set of valid inputs, no matter what state the application is in. This is more significant that just moving to a less moded interaction style: the base vocabulary of what can be considered a command increased dramatically. They also typically have many degrees of freedom in both the underlying application and the interaction devices. For example a standard DataGlove has 16 degrees of freedom, each of which can be sampled up to 60 times per second. In addition, the user may typically operate several input or output devices simultaneously. These interfaces thus require large amounts of input and output bandwidth, which leads to a need for devoting significant hardware resources to processing high bit rates of data as well as to data reduction and recognition algorithms needed to handle voluminous, but imprecise inputs.

New interface styles will also be characterized by devices that are more closely matched to the user's characteristics than the computer's. Rather than training a user to operate a keyboard, he might be allowed to use his existing, natural communicative abilities (such as gesture, speech, handwriting, or even eye movements), while the computer will be programmed to perceive the user's actions and interpret them in a way meaningful to the dialogue. This interpretation will require an increasing amount of processing power devoted to user interaction as well as techniques for handling "probabilistic inputs." Unlike traditional mice and keyboards, the input/output devices used in these interfaces, such as speech or gesture recognizers, may produce a probability vector rather than a single token for a user input. Emerging pen-based and speech-based interfaces share some of these qualities, though they are still fundamentally command-based.

The new interface styles also place stringent requirements on real-time response to the user's actions. For example, with a head-mounted display, there must be very little delay between the time when the user moves his or her head and a new set of images is generated or the illusion of interacting with a real physical environment will be lost. Better support for the management of time within the user interface specification and its implementation will thus be another requirement of these new interfaces. The traditional approaches to real-time computing used in the operating systems and embedded systems communities are not powerful enough to solve the general problem of developing elegant, real-time interfaces for arbitrary applications.

SPECIFIC RESEARCH DIRECTIONS

The following sections describe the state of the art and current research opportunities in human-computer interaction styles and devices. While our overall goal is to develop a human "demand-pull" rather than "technology push" approach to this field, we cannot ignore the constraints of the underlying technologies. These sections describe current realities and opportunities of such technologies.

Research in Interaction Styles

The general framework in which to view research into styles of human-computer interaction is a search for both higher bandwidth communication between human and computer and better "fit" between human and computer (that is, reductions in cognitive load on the user, training required, and effort devoted to non-task concerns). While natural speech already provides appropriate bandwidth for verbal communication, there is no such counterpart for symbolic, non-verbal styles of interaction, and that is where the search for increased bandwidth is of greatest interest, as are abstractions and encodings that increase effective bandwidth.

The principal interaction styles currently in use and emerging are command languages, menus, natural language, direct manipulation, and virtual reality. Command languages are concise and unambiguous but typically difficult to learn and use. Their continuing advantage in the face of newer interaction styles is that they are most amenable to abstraction, or writing programs or scripts of user input commands. Menus require the user to recognize the desired entry rather than recall it, reducing the load on long-term memory. Menus can exist in many forms. although their traditional implementation has the user click with a mouse over the item to be selected, the user could just as well respond via voice command, just as real menus are used in restaurants. Natural language is still a misnomer, since, given the state of the art, working natural language interfaces must still be restricted to a subset of natural language, and the subset must be chosen carefully––in vocabulary, range of syntactic constructs, and domain. Although future research in natural language offers the hope of human-computer communication that is so natural it is "just like talking to a person," such conversation may not always be the most effective way of commanding a machine (Small, 1983). Direct manipulation interfaces present a set of objects on a screen and provide the user a repertoire of manipulations that can be performed on any of them (Shneiderman, 1983). The result is a computer interface consisting of a set of objects and operations that closely resemble their counterparts in the physical world. Finally, research in virtual reality carries the user's illusion of manipulating real objects still further, allowing interaction with computer-generated images as if they were real objects located in space surrounding the user. Some aspects of virtual reality can be approached as "three-dimensional direct manipulation plus voice input/output," but not all virtual reality interactions can be characterized that simply (Fisher, 1986, Fisher, 1988, Foley, 1987).

We can see that increasing verisimilitude in interaction styles has yielded powerful steps from command line to direct manipulation and virtual reality styles, but it can also be a millstone. Some tasks are inherently abstract or non-physical, and for such, we may want to explore new interaction styles that are not constrained by a need to mimic operations that occur in the physical world.

With respect to current interaction styles, the goal of research is to make them more usable. Better design guidance for screen layout, menu organization, or color usage are needed here. Such guidance should, where possible, be motivated by powerful and general underlying scientific principles, rather than by ad-hoc judgment. With a sufficiently powerful underlying framework in place, guidance of this sort can also be conveniently embodied in automated tools to aid designers.

With respect to future interaction styles, the goal of research is to invent and study new paradigms for interaction that provide increased human-computer communication bandwidth in media and styles that are well matched to and driven by human capabilities. Such styles may be new ways of using existing input/output devices or they may be based on new devices. In addition, new developments in interaction styles should facilitate improvements in such areas as end user customization, adaptive interfaces, the use of intelligent "agents" in interfaces, and interfaces for multi-user cooperative work. They should also accommodate features, discussed in other sections of this report, such as command history (redo, undo, extended to non-linear sequences and graphical commands), user programming by demonstration, on-line help, generation of relevant documentation, and the use of evaluation techniques

Research in Input Devices

Three-dimensional pointing and manipulation: A magnetic tracker (e.g., Polhemus or Ascension), can provide the three-dimensional analogue of the mouse or data tablet. Other technologies, such as ultrasonic ranging or video tracking, can also be used in this way. All of these are still limited in one way or another−−in latency, precision, stability, proximity to CRTs, or number of available samples per second. Better approaches to three-dimensional tracking thus constitute a useful research area. A good technology for this purpose not only addresses the problems noted and is cheap and reliable, but also should not encumber the user. This suggests that the ideal solution may lie in sensors that observe the user, without requiring him to hold or wear anything. Camera-based locator devices are being studied, but are still limited. A single-camera system is limited to its line of sight; more cameras can be added but full coverage of an area may require many cameras and a way to switch among them smoothly. This approach requires some type of image processing to interpret the picture of the user and extract the desired hand or body position.

General gesture input: The three-dimensional mouse can be taken a step further. Rather than simply designating a location in three-space, it can allow a user to make natural, continuous gestures in space. Progress in this area requires not only a non-encumbering three-dimensional tracking technology but also a way to recognize human gestures occurring dynamically. These are typically made with poor precision and repeatability, so a useful input device would have to tolerate such imprecision and still glean the user’s intended action. Two-dimensional gestures on a surface share these properties, and are of particular interest in emerging pen-based interfaces. Gesture-based input is currently becoming a very active research area (Schmandt, 1983, Sturman, 1989, Ware, 1988, Ware, 1989).

Simultaneous two-hand input: Aside from touch typing, computer input scenarios typically assume the use of only one hand at a time, though people are quite good at manipulating both hands in coordinated or separated tasks (Buxton, 1986). This area requires hardware that can be replicated (certain tracking technologies permit only one tracked object at a time and are thus unsuitable) as well as research into interaction techniques and styles that make appropriate and advantageous use of both hands. In addition, current user interface software architectures assume a single input stream; new approaches are needed for multiple, parallel inputs (Hill, 1986).

Stereo display in conjunction with three-dimensional manipulation: A natural extension of three-dimensional tracking and gesturing is to allow the direct manipulation of virtual, displayed three-dimensional objects. This suggests stereoscopic display, but any perceivable display of depth might be adequate. The key is linking the three-dimensional input to the three-dimensional output in a faithful and convincing way.

Virtual input devices: The notion of three-dimensional object manipulation can be extended to manipulating virtual tools, where the user first obtains a tool (by three-dimensional direct manipulation) and then applies it to a three-dimensional virtual object. A virtual tool, can of course, metamorphose as needed for the job at hand and otherwise improve upon the properties of non-virtual tools.

Speech: Speech input has been a long standing area of research. While progress is being made, it is slower than optimists originally predicted, and further work remains in this field. Although the goal of continuous speech recognition remains elusive, unnatural, isolated-word speech recognition is appropriate for some tasks and even natural for communicating with a computer, rather than another human. Research is needed not only in the actual speech recognition technology but also in how to use speech in an interface (Kamel, 1990). The notion that the ideal computer is one that behaves and communicates just like a personal assistant is a naive one: people expect computers to behave like tools, not like other people; computers have thus far proven more adept at the former than the latter; and furthermore the computer-as-person approach ultimately limits the usefulness of the computer to that of the person being emulated (Small, 1983).

Speech, in coordination with other modes, such as manual input: Speech is often most useful in conjunction with other input media, providing an additional channel when the user is already occupied. (Driving a car and conducting a conversation is an everyday example.) Research into appropriate ways to combine modes and allocate subtasks to them would be useful. In the long run, however, the most interesting cases are those where the use of speech is not obviously mandated. If the user's hands, feet, and eyes are busy, speech may be the only reasonable choice for some input. The technical issues in making this choice are trivial since there are no alternatives. The more interesting case is to begin with a collection of tasks in a user interface and then allocate them to the range of the user's communication modes; a comprehensive, principled framework for so doing is still lacking. Another use for multiple modes is to combine otherwise ambiguous inputs from several modes (such as pointing and speaking) to yield an unambiguous interpretation of the user's input (Bolt, 1980, Schmandt, 1982).

Better eye input technology: One specific area of user input technology that has recently come under investigation is eye movements. Eye movement-based input, properly used, can provide an unusually rapid and natural means of communication as well as a good example of the passive-sensing quality of the non-command interaction style. However, eye tracking technology has progressed slowly, and is still not adequate for use in practical applications for non-disabled users. The accuracy of an eye tracker that is useful in a real-time interface (as opposed to the more stringent requirements for eye movement research) is limited by the size of the fovea. The best current eye trackers approach this limit. However, stability and repeatability of the measurements leave much to be desired. In a research study it is acceptable if the eye tracker fails very briefly from time to time; it may require that an experimental trial be discarded, but the user need not be aware of the problem. In an interactive interface, though, as soon as it begins to fail, the user can no longer rely on the fact that the computer dialogue is influenced by where his eye is pointing and will thus soon be tempted to retreat permanently to whatever backup input modes are available. While eye trackers have dropped somewhat in price, their performance in this regard has not improved. Performance does not appear to be constrained by fundamental limits, but simply by lack of effort in this area (Young, 1975).

Better use of eye input: The other half of eye movement input is to make wise and effective use of eye movements, ideally in a non-command-based style. Eye movements, like other passive, non-command inputs (e.g., gesture, conversational speech) are often non-intentional or not conscious, so they must be interpreted carefully to avoid annoying the user with unwanted responses to his actions. In eye movements, we call this the "Midas Touch" problem. The problem with a simple implementation of an eye tracker interface is that people are not accustomed to operating devices simply by moving their eyes. They expect to be able to look at an item without having the look cause an action to occur. At first it is helpful to be able simply to look at what you want and have it occur without further action; soon, though, it becomes like the Midas Touch. Everywhere you look, another command is activated; you cannot look anywhere without issuing a command. Eye movements are an example of how most of the non-command, passive inputs will require either careful interface design to avoid this problem or some form of "clutch" to engage and disengage the monitoring (Bolt, 1981, Bolt, 1982, Jacob, 1990, Jacob, 1991, Jacob, 1993, Starker, 1990, Ware, 1987).

Techniques for communication and manipulation of multidimensional data: Typical current visualization work is rather passive. At most, the user might be able to change viewpoint or rotate a displayed object, but the object itself remains the same. Even so, user-controlled motion (kinesthetic feedback) has been found to add dramatically to perception of complex three-dimensional displays. This framework can be extended to consider interacting with the visualization in a more substantive way. This raises questions of what interaction techniques are appropriate for directly manipulating a visualization display, particularly in several dimensions. The choice of appropriate interaction techniques will be influenced by the output representations selected. The goal is to find input methods that minimize the user's cognitive effort in translating between the metaphor or representation used in the display and the action he must make. Three-dimensional displays (note that this refers to a three-dimensional display image, not three-dimensional underlying data) suggest the use of three-dimensional input devices and techniques.

Passive monitoring of user attitude: A user-computer dialogue could be improved if the computer knew the answer to such simple questions as, Is the user still sitting in his chair? Is he facing toward the computer? Is he using the telephone? Is another person present in the room? Real-world objects can do useful things when they detect even simple events like user arrival and departure. Ironically, bathroom sinks in modern airports (which use sensors) are more sophisticated than computer terminal screen savers (which rely on a time-out strategy).

Further physiological measurements: In addition to three-dimensional position tracking and eye tracking, a variety of other physiological characteristics of the user might be monitored and the information used to modify the computer's dialogue with its user. Pupil diameter and galvanic skin response are examples of measurements that are relatively easy and comfortable to make, although their accurate instantaneous interpretation is an open question. A more difficult measure is an electroencephalogram, although progress has been made in identifying specific evoked potential signals in real time (Wickens, 1983). The most accurate results are currently obtained with a somewhat unwieldy superconducting detector (Lewis, 1987), rather than the conventional electrodes, but improvements in this technology can be envisioned, particularly if high-temperature superconductivity becomes practical.

Direct connect: Perhaps the final frontier in user input and output devices is to measure and stimulate neurons directly, rather than relying on the body's transducers. This is unrealistic at present, but it may someday be a primary mode of high-performance user-computer interaction.

Ergonomic considerations: In a different vein from the above, research into input devices and styles for disabled users is another fruitful area. Ideally, at least some of the accommodations required for disabled users will be helpful to all users, in unexpected ways. An analogy is made to curb cuts, placed on sidewalks to accommodate wheelchairs, which have proven valuable to a variety of non-disabled users, such as bicyclists, stroller-pushers and skateboard-riders.

Device interfaces: A mundane but nagging problem in this field is connecting new devices to a computer. New devices often introduce new, slightly different hardware connections and software protocols for communication. A standard physical interface and communication protocol, along with operating system-level support for it, would go a long way toward alleviating the essential, but fundamentally trivial work required of researchers to begin using a new device. The communication requirements of many input devices are sufficiently similar and undemanding that this is not a serious technical problem nor should a standard levy an unreasonable performance penalty. The MIDI standard interface for both physical connection and simple logical protocol for keyboard-oriented musical instruments provides an example of the technical approach needed; and the dramatic success of MIDI in expanding the usefulness of electronic musical instruments suggests the payoff available.

Device taxonomy: Even without adding new devices to the range of available choices, there is considerable confusion among current position- and gesture-sensing devices, touch tablets, joysticks, and mice. For example, sophisticated forms of joysticks are available for controlling six degrees of freedom, both isometric, isotonic, and in combinations. Software for using such input devices must typically be modified for each device. A more useful approach would be to start with a taxonomy of these devices and design user interfaces with respect to the taxonomy rather than specifically for each individual device. The graphics-standard approach of logical devices is probably too limiting here, we acknowledge instead that the different input devices may require different user interface designs. The goal is simply to be able to switch among these designs at runtime. The first step is an appropriate taxonomy or model of input device characteristics, such as (Bleser, 1990, Card, 1990). Each device could identify itself within the taxonomy to its host software or it might simply give its name as an index to a table of device descriptions kept in the host.

Automatic interface design: A further extension of this approach would allow the user interface software to synthesize a new interface for the application at runtime based on the characteristics of the available devices and on rules about human factors and interface design.

Research in Output Devices

Better graphics display packaging: The quest for a large, flat, high-resolution display screen continues, with limited success. Head-mounted displays with wide-angle optics can also provide some of the same benefits. We do anticipate rapid advances in small head-mounted displays over the next few years. Major challenges will be in the physics and optics of getting correct field of view and stereoscopy parameters, as well as just getting a large number of color pixels into a small package.

Better graphics resolution: While great strides have been made in graphical output resolution, the demands of high-performance user interaction are far from satisfied. People can routinely make effective use of densely-packed documents such as a printed road map or navigational chart. Their high resolution is well matched to human perceptual abilities, yet it is well beyond the current state of practice in display technology.

...And lots more of it: Similarly, people can make effective use of considerably more high-resolution display surface than current technology allows. A conventional office desk, covered with paper, provides perhaps ten square feet of visible surface, with a resolution of 300-1000 dots per inch. Users can typically make good use of that entire area, "panning" and "zooming" in and out as needed. The usefulness of a display of at least comparable richness is thus already demonstrated.

Autostereoscopic displays: There is a variety of methods currently available for displaying stereo images, but most are sufficiently specialized that they would not be used for conventional day-to-day user-computer interaction. Similarly, graphic displays were once so exotic that they too were only used for specialized purposes. Today they are used for routine data file manipulation. A high-quality, readily convincing, easily fusible, high-resolution autostereoscopic (no glasses) display would make three-dimensional images similarly accessible to user interface designers. Perhaps the best current technology is that using alternating CRT frames with a synchronized shutter; it is quite readily fusible, but it does require glasses. Varifocal mirror displays use a moving mirror to trace a convincing-looking three-dimensional wire-frame "object" in space out of a light beam; they require no glasses, but are somewhat limited in resolution, size, and color. A refinement of this approach uses a laser beam projected onto a spinning translucent helix; this should permit color objects viewable from any angle. Both the varifocal mirror and helix approaches produce real images in three dimensional space, but both suffer in that objects in the scene cannot occult one another. Holograms can display objects that appear entirely three-dimensional. Digital holography, that is, holograms constructed by computer calculation rather than optical imaging, is thus a highly promising approach, although it is currently extremely computation-intensive. Much as happened with ray tracing, improved algorithms and hardware will likely solve this problem in the future.

Touchable three-dimensional displays: A further improvement in the realism of a three-dimensional display would be to permit the user to reach out and feel the displayed object. Force and resistance feedback are discussed below; they derive particular power when combined with a convincing stereoscopic display.

Techniques for representing complex abstract data: Visualization for abstract or non-geometrical data requires an additional step compared to visualization of concrete physical objects. An appropriate concrete geometrical object or other such representation must first be designed for the data, one which effectively communicates information about the underlying non-geometrical data. Then, the representation must be produced on a display, as with conventional visualization work. Visualization for non-geometrical data is a more significant problem, since it is necessary first to devise a good way to represent the non-geometrical information in a perceptible way and only then to develop the machinery to display the new representation to a user. The research issue here is in what image to draw to communicate the underlying information to the user and on how to exploit human perceptual channels in appropriate combinations. Because of the difficulty of designing and understanding appropriate concrete representations for abstract or other non-geometric data, inputs from perceptual theory are important here.

Techniques for representing high-dimensional data: Data of more than three or four dimensions also share the above property. Since the data to be displayed have more dimensions than can be represented in the real world (even with three-dimensional displays or volume visualization), the general approach for representing such data is to map it into symbols or other images in two or three dimensions (including objects which might then be rendered by volume visualization).

Non-speech audio output for 'visualizing' data: It is not necessary to restrict the notion of "visualization" to visual displays. The basic idea of visualization is to put data into some representation perceptible to the user. This representation could be entirely visual or visual plus, for example, audio or not visual at all (tactile plus audio). For example, a useful means for encoding information is in sounds. Typical computer systems use only a few simple beeps for alerts, but humans can glean considerably more meaning from simple sounds, both natural sound effects and artificially-conceived tones (Bly, 1982, Buxton, 1989, Gaver, 1986). Andrews curves have been used to represent many-dimensional points graphically, as a Fourier series; such curves would be a clear candidate for presenting a point in a many-dimensional space as a single complex tone (Andrews, 1972). In addition, sounds can be added to visual representations to increase the amount of data that can be communicated to the user by exploiting an additional channel. This is especially useful when added dimensions must be displayed.

When displaying multi-dimensional data, issues of interactions between dimensions also arise: For example, Does the value of one display parameter (e.g., color) affect perception of the value of another (e.g., size)? How much, and in what way? How, then, should several data values be presented together? Some of these questions have already been answered by the perceptual psychology community, but others are still open research issues. A particular case of interest appears when one of the display parameters drives a visual output and another drives an audio output––how can information be communicated in both modes simultaneously, and how does what is communicated to the user relate to the underlying data? Devising representations that will help the user integrate the displayed information together requires answers to these questions.

Directional audio: It is possible to create the effect of a sound that comes from a specific location in three-space. One approach is simply to fill a room with independently addressible speakers, but recent research has led to a "software" method of achieving the same effect (Wenzel, 1988). This approach is still in its infancy, but improvements in robustness, convenience, and user-independence will make this (or variations on it) a useful medium.

Speech: Natural, continuous speech output is difficult to achieve, but simple, isolated utterances may still be reasonable when talking with a computer. Much of the discussion under speech for input, above, applies here too, and, again, research in how to use speech and integrate it into a multi-mode interface is particularly important.

Force and tactile feedback: Better force and tactile feedback would be useful in making computer-generated objects more realistic and thereby allowing people to use their everyday skills in dealing with them. This is currently an interesting area of research, but is much in need of new ideas and technologies. It is an area in which no single technical approach has shown dominance (unlike, for example, output, where CRT technology is far ahead of its competitors). Now is a time at which conceptual research could pay off handsomely in this field. The ideal here is to create fully general purpose feedback, so that one could display "virtual tools" or other objects and allow a user to grasp and finely manipulate them, using all of his senses. Current research has focussed on specific forms of feedback, however, such as force or texture (Minsky, 1990). A continuing problem is to apply sufficient force to simulate the effect of hitting a solid surface, and yet not endanger the user with the risk that such force could be applied mistakenly. One solution here is to provide feedback in the form of resistance to motion, rather than force, although this requires more attachments to the user.

Smell: Olfactory output has not been used much, other than some erratic attempts in movie theaters (or perhaps the smell of burning circuit boards, the ultimate computer error message). Although it is a powerful means of communicating with people, olfactory output suffers from the technical difficulty of controlling it precisely over time, that is, displaying a single odor at a single point in time and then "clearing" the display. More fundamentally, knowledge of the components of odors and the mechanisms of olfaction are still somewhat primitive. Odors are still classified in terms of many seemingly arbitrary components, rather than a single, universal basis. It is not yet possible to synthesize an odor to represent something from some universal basis in the way one can synthesize a convincing graphical image of an arbitrary object by modulating beams of red, green, and blue light.

Environmental controls: A simple, but useful form of computer output is control over aspects of the user's environment, such as room temperature, lighting, sounds, chair posture. This is a relatively low-bandwidth channel, but it could interact usefully with some of the passive physiological measurement inputs discussed above. It also provides a further step toward making the computerized world in which the user does much of his work richer and more varied, more like the real world, and less like the fixed glass screen and keyboard.

Input-output: For completeness, note that some of the input categories discussed above are really combination input-output media, such as speech, virtual tools, and even direct connect.

Ergonomic considerations: In a different direction, less tiring output media will become increasingly important as users spend more of their lives receiving computer output. CRT displays continue to elicit complaints of muscle fatigue, eye fatigue, and the like. While some of these seem to be caused by factors outside the display device, such as poor job design and poor seating posture, some may also be caused by the actual characteristics of the device. Better alternatives would be worthwhile.

OVERALL CONCERNS

Impediments to Interaction Research

What are some of the stumbling blocks for doing research in these areas?

One problem is that the best research in interaction techniques and devices is highly interdisciplinary. It does not fall directly within the current disciplinary structure of Computer Science or Psychology, but right at the boundary between the two, and includes disciplines such as Sociology, Anthropology, Graphic Design, and Electrical and Mechanical Engineering. Institutional pressures tend to lead researchers toward the centers of their disciplines, not the edges. For example, it is surely easier to achieve tenure in a department by doing work that is close to the center of its discipline. Funding is also typically tied to traditional disciplines. However, funding can serve to lead institutions to alter their perspective and, ultimately, their discipline structure. An aggressive funding program in interactive systems can, first, legitimize this hybrid discipline and, ultimately, lead universities and laboratories to organize around it in order to attract such funding.

Another problem is that interaction technique research is an area in which progress in artifacts currently leads that in theory. The field is in its infancy, and the majority of new ideas are still coming from builders of devices and interfaces. The underlying theory is not yet well established, and relatively few advances have derived from theoretical work thus far. This need not always be the case, but for the present it does give research in the field the flavor of ad-hoc "hacking," rather than science. It therefore does not fit the traditional scientific research mold as well as other fields and is often less academically respectable. This is partly the result of the newness of the field, and also partly the result of its interdisciplinary nature (it is neither "pure" Psychology, nor pure engineering). Here, too, a coherent and well-reasoned funding program could lead the way in establishing this field.

Finally, work in new interaction techniques often requires expensive equipment. Prices of equipment in this area tend to drop over time, but, to be timely, the research must be done early on, when they are still expensive. Three-dimensional tracking, virtual reality, eye movements, and directional audio are all good examples. University funding, however, tends to favor labor-intensive rather than equipment-intensive research; both funding agencies and universities prefer investment in human resources than in equipment, since the former provides additional educational benefits to society. This is a continuing problem for research in this area. It is somewhat mitigated in large research laboratories, which do not also have the same educational mission. However, companies that actually manufacture much of the equipment in question are often very small and thus not able to pursue the long-term basic research that is needed.

One possible way to mitigate this last problem is through planned sharing, supported by a new infrastructure. Sharing a new input/output device among researchers is considerably more difficult than sharing software, since the physical device must be reconstructed. A valuable contribution to the infrastructure needed to support this research would therefore be a central "job-shop" that can produce hardware used for interaction. If such devices could be produced centrally, it would be relatively easy for the central shop to make duplicates of a new device for other researchers, since the bulk of the effort typically goes into design not fabrication. In addition, the skills needed for actually building such devices more often reside in mechanical or electrical engineering than in HCI, yet it is important that such work be guided by insight into human-computer communication issues and needs. The central shop could allow economical sharing of the relevant expertise from other fields. Such an organization could also be a force for developing and promulgating the kind of standard device interface discussed above.

Benefits

Research in new input/output devices and interaction styles for interactive systems can lead to two kinds of benefits. The first is progress in our scientific understanding of human-computer communication: it can improve our understanding of new modes by which people can communicate with machines and lead to understanding of the characteristics of the devices, interaction techniques, and dialogue styles that allow us to exploit these modes. The second is in application of these results: it can lead to faster and more natural communication with interactive systems, enable better quality and efficiency in the operation of such systems, and improve the working conditions of their users by providing them with richer and more natural means of communication.

References

Andrews,.
D.F. Andrews, ‘‘Plots of High-dimensional Data,’’ *Biometrics*, vol. 28, p. 125, 1972.

Bleser,.
T.W. Bleser and J.L. Sibert, ‘‘Toto: A Tool for Selecting Interaction Techniques,’’ *Proc. ACM UIST’90 Symposium on User Interface Software and Technology*, pp. 135-142, Addison-Wesley/ACM Press, Snowbird, Utah, 1990.

Bly,. S. Bly, ‘‘Presenting Information in Sound,’’ *Proc. ACM Human Factors in Computer Systems Conference*, pp. 371-375, 1982.

Bolt,.R.A. Bolt, ‘‘Put-That-There: Voice and Gesture at the Graphics Interface,’’ *Computer Graphics*, vol. 14, no. 3, pp. 262-270, Seattle, 1980.

Bolt,.R.A. Bolt, ‘‘Gaze-Orchestrated Dynamic Windows,’’ *Computer Graphics*, vol. 15, no. 3, pp. 109-119, August 1981.

Bolt,.R.A. Bolt, ‘‘Eyes at the Interface,’’ *Proc. ACM Human Factors in Computer Systems Conference*, pp. 360-362, 1982.

Buxton,.
W. Buxton and B.A. Myers, ‘‘A Study in Two-Handed Input,’’ *Proc. ACM CHI’86 Human Factors in Computing Systems Conference*, pp. 321-326, 1986.

Buxton,.
W. Buxton, ‘‘Introduction to This Special Issue on Nonspeech Audio,’’ *Human-computer Interaction*, vol. 4, pp. 1-9, 1989.

Card,.S.K. Card, W.K. English, and B.J. Burr, ‘‘Evaluation of Mouse, Rate-controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT,’’ *Ergonomics*, vol. 21, no. 8, pp. 601-613, 1978.

Card,.S.K. Card, J.D. Mackinlay, and G.G. Robertson, ‘‘The Design Space of Input Devices,’’ *Proc. ACM CHI’90 Human Factors in Computing Systems Conference*, pp. 117-124, Addison-Wesley/ACM Press, 1990.

Fisher,.

S.S. Fisher, M. McGreevy, J. Humphries, and W. Robinett, ‘‘Virtual Environment Display System,’’ *Proc. Workshop on 3D Interactive Computer Graphics*, pp. 77-87, Chapel Hill, N.C., 1986.

Fisher,.

S.S. Fisher, E.J. Wenzel, C. Coler, and M.W. McGreevy, ‘‘Virtual Interface Environment Workstations,’’ *Proc. 32nd Annual Meeting of the Human Factors Society*, pp. 91-95, Anaheim, Calif., 1988.

Foley,.

J.D. Foley, ‘‘Interfaces for Advanced Computing,’’ *Scientific American*, vol. 257, no. 4, pp. 127-135, October 1987.

Foley,.

J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes, *Computer Graphics: Principles and Practice*, Addison-Wesley, Reading, Mass., 1990.

Gaver,.

W.W. Gaver, ‘‘Auditory Icons: Using Sound in Computer Interfaces,’’ *Human-Computer Interaction*, vol. 2, pp. 167-177, 1986.

Green,.

M. Green and R.J.K. Jacob, ‘‘Software Architectures and Metaphors for Non-WIMP User Interfaces,’’ *Computer Graphics*, vol. 25, no. 3, pp. 229-235, July 1991.

Hill,.R.D. Hill, ‘‘Supporting Concurrency, Communication and Synchronization in Human-Computer Interaction–The Sassafras UIMS,’’ *ACM Transactions on Graphics*, vol. 5, no. 3, pp. 179-210, 1986.

Jacob,.

R.J.K. Jacob, ‘‘What You Look At is What You Get: Eye Movement-Based Interaction Techniques,’’ *Proc. ACM CHI’90 Human Factors in Computing Systems Conference*, pp. 11-18, Addison-Wesley/ACM Press, 1990.

Jacob,.

R.J.K. Jacob, ‘‘The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get,’’ *ACM Transactions on Information Systems*, vol. 9, no. 3, pp. 152-169, April 1991.

Jacob,.

R.J.K. Jacob, ‘‘Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces,’’ in *Advances in Human-Computer Interaction, Vol. 4*, ed. by H.R. Hartson and D. Hix, pp. 151-190, Ablex Publishing Co., Norwood, N.J., 1993. http://www.eecs.tufts.edu/˜jacob/papers/hartson.txt [ASCII]; http://www.eecs.tufts.edu/˜jacob/papers/hartson.pdf [PDF].

Kamel,.

R. Kamel, ‘‘Guest Editor’s Introduction: Voice in Computing,’’ *IEEE Computer*, vol. 23, no. 8, pp. 8-9, August 1990.

Ledgard,.

H. Ledgard, J.A. Whiteside, A. Singer, and W. Seymour, ‘‘The Natural Language of Interactive Systems,’’ *Comm. ACM*, vol. 23, pp. 556-563, 1980.

Lewis,.

G.W. Lewis, L.J. Trejo, P. Nunez, H. Weinberg, and P. Naitoh, ‘‘Evoked Neuromagnetic Fields: Implications for Indexing Performance,’’ *Biomagnetism 1987, Proc. of the Sixth International Conference on Biomagnetism*, Tokyo, 1987.

Minsky,.

M. Minsky, M. Ouh-young, O. Steele, F.P. Brooks, and M. Behensky, ‘‘Feeling and Seeing: Issues in Force Display,’’ *Computer Graphics*, vol. 24, no. 2, pp. 235-244, 1990.

Nielsen,.

J. Nielsen, ‘‘Trip Report: CHI’90,’’ *SIGCHI Bulletin*, vol. 22, no. 2, pp. 20-25, 1990.

Schmandt,.

C. Schmandt and E.A. Hulteen, ‘‘The Intelligent Voice-Interactive Interface,’’ *Proc. ACM Human Factors in Computer Systems Conference*, pp. 363-366, 1982.

Schmandt,.

C. Schmandt, ‘‘Spatial Input/display Correspondence in a Stereoscopic Computer Graphic Workstation,’’ *Computer Graphics*, vol. 17, no. 3, pp. 253-259, 1983.

Shneiderman,.

B. Shneiderman, ‘‘Direct Manipulation: A Step Beyond Programming Languages,’’ *IEEE Computer*, vol. 16, no. 8, pp. 57-69, 1983.

Small,.

D.W. Small and L.J. Weldon, ‘‘An Experimental Comparison of Natural and Structured Query Languages,’’ *Human Factors*, vol. 25, pp. 253-263, 1983.

Starker,.

I. Starker and R.A. Bolt, ‘‘A Gaze-Responsive Self-Disclosing Display,’’ *Proc. ACM CHI’90 Human Factors in Computing Systems Conference,* pp. 3-9, Addison-Wesley/ACM Press, 1990.

Sturman,.

D.J. Sturman, D. Zeltzer, and S. Pieper, ‘‘Hands-on Interaction with Virtual Environments,’’ *Proc. ACM UIST’89 Symposium on User Interface Software and Technology*, pp. 19-24, Addison-Wesley/ACM Press, Williamsburg, Va., 1989.

Tufte,.

E.R. Tufte, ‘‘Visual Design of the User Interface,’’ IBM Corporation, Armonk, N.Y., 1989.

Ware,.

C. Ware and H.T. Mikaelian, ‘‘An Evaluation of an Eye Tracker as a Device for Computer Input,’’ *Proc. ACM CHI+GI’87 Human Factors in Computing Systems Conference*, pp. 183-188, 1987.

Ware,.

C. Ware and D.R. Jessome, ‘‘Using the Bat: A Six-Dimensional Mouse for Object Placement,’’ *IEEE Computer Graphics and Applications*, vol. 8, no. 6, pp. 65-70, 1988.

Ware,.

C. Ware and C. Baxter, ‘‘Bat Brushes: On the Uses of Six Position and Orientation Parameters in a Paint Program,’’ *Proc. ACM CHI’89 Human Factors in Computing Systems Conference,* pp. 55-60, Addison-Wesley/ACM Press, 1989.

Wenzel,.

E.M. Wenzel, F.S. Wightman, and S.H. Foster, ‘‘A Virtual Display System for Conveying Three-dimensional Acoustic Information,’’ *Proc. 32nd Annual Meeting of the Human Factors Society*, pp. 86-90, Anaheim, Calif., 1988.

Wickens,.

C. Wickens, A. Kramer, L. Vanasse, and E. Donchin, ‘‘Performance of Concurrent Tasks: A Psychophysiological Analysis of the Reciprocity of Information-Processing Resources,’’ *Science,* vol. 221, pp. 1080-1082, 1983.

Young,.

L.R. Young and D. Sheena, ‘‘Survey of Eye Movement Recording Methods,’’ *Behavior Research Methods and Instrumentation*, vol. 7, no. 5, pp. 397-429, 1975.