

# PERCEPTUAL USER INTERFACES

✎ MATTHEW TURK  
AND GEORGE ROBERTSON,  
*Guest Editors*

HERE IS NO MOORE'S LAW FOR USER interfaces. Human-computer interaction has not changed fundamentally for nearly two decades. Most users interact with computers by typing, pointing, and clicking. The majority of work in human-computer interfaces (HCI) in recent decades has been aimed at creating graphical user interfaces

(GUIs) that give users direct control and predictability. These properties provide the user a clear model of what commands and action are possible and what their affects will be; they allow users to have a sense of accomplishment and responsibility about their interactions with computer applications.

Although these endeavors have been very successful, and the WIMP (windows, icons, menus, pointer) paradigm has served to provide a stable and global face to computing, it is clear this paradigm will not scale to match the myriad form factors and uses of computers in the future. Computing devices are becoming smaller and ubiquitous, and interaction with them is becoming more and more pervasive in our daily lives. At the same time, large-scale displays are becoming more common, and we are beginning to see a convergence between computers and television. In all cases, the need arises for more general and intuitive ways of interacting with the technology. Pointing, clicking, and typing—though still appropriate for many uses of computers in the foreseeable future—will not be how most people interact with the majority of computing devices for long.

What we need are interaction techniques well matched with how people will use computers.

From small, mobile devices carried or worn to powerful devices embedded in homes, businesses, and automobiles—one size does not fit all. Is there a paradigm that captures the essence of such diverse future HCI requirements? We believe there is, and it is grounded in how people interact with each other and with the real world. This is the essence of *perceptual user interfaces (PUIs)*.

PUIs are characterized by interaction techniques that combine an understanding of natural human capabilities (particularly communication, motor, cognitive, and perceptual skills) with computer I/O devices and machine perception and reasoning. They seek to make the user interface more natural and compelling by taking advantage of the ways in which people naturally interact with each other and with the world—both verbally and nonverbally. Devices and sensors should be transparent and passive if possible, and machines should perceive relevant human communication channels as well as generate output that is naturally understood. This is expected to require integration at multiple levels of technologies such as speech and sound recognition and generation, computer vision, graphical animation and visualization, language understanding, touch-based

sensing and feedback (haptics), learning, user modeling, and dialogue management.

The accompanying figure illustrates how PUI encompasses research in several areas. Although the figure shows information flow in the context of a traditional computer form factor, PUI is intended for new form factors as well.

A *perceptive UI* (as opposed to PUI) is one that adds human-like perceptual capabilities to the computer, for example, making the computer aware of what the user is saying or what the user's face, body, and hands are doing. These interfaces provide input to the computer while leveraging human communication and motor skills.

*Multimodal UI* is closely related, emphasizing human communication skills. We use multiple modalities when we engage in face-to-face communication, leading to more effective communication. Most work on multimodal UI has focused on computer input (for example, using speech together with pen-based gestures). Multimodal output uses different modalities, like visual display, audio, and tactile feedback, to engage human perceptual, cognitive, and communication skills in understanding what is being presented. In multimodal UI, various modalities are sometimes used independently and sometimes simultaneously or tightly coupled.

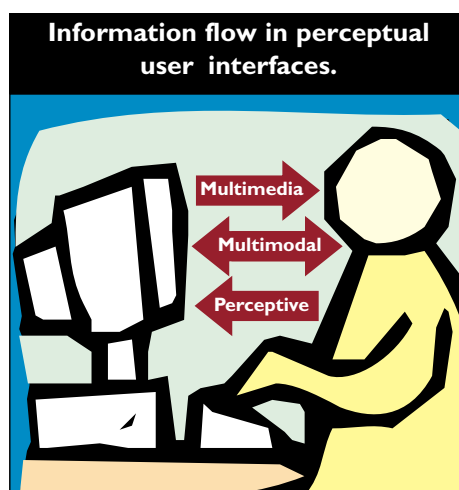
*Multimedia UI*, which has experienced an enormous amount of research during the last two decades, uses perceptual and cognitive skills to interpret information presented to the user. Text, graphics, audio, and video are the typical media used. Multimedia research focuses on the media, while multimodal research focuses on the human perceptual channels. From that point of view, multimedia research is a subset of multimodal output research.

PUI INTEGRATES PERCEPTIVE, MULTIMODAL, AND multimedia interfaces to bring our human capabilities to bear on creating more natural and intuitive interfaces. PUIs will enhance the use of computers as tools or appliances, directly enhancing GUI-based applications—for example, by taking into account gestures, speech, and eye gaze (“No, *that* one”). Perhaps more importantly, these new technologies will enable broad uses of computers as assistants, or agents, that will interact in more human-like ways. Perceptual interfaces will enable multiple styles of interaction—such as speech only, speech and gesture, text and touch, vision, and

synthetic sound—each of which may be appropriate in different circumstances, whether that be desktop apps, hands-free mobile use, or embedded household systems.

THERE ARE A NUMBER OF CHALLENGES FACING THE development and use of PUIs. It is an ambitious endeavor with diverse elements. The articles in this special section present both challenges and early results toward the goal of perceptual interfaces. They are not exhaustive, but rather serve as examples of efforts in this area. (See [1] for others.) Oviatt and Cohen summarize multimodal interfaces, emphasizing their extensive work on speech and pen-based systems. This work shows how multiple modalities can lead to more stable and robust systems (for example, reducing error and disfluency rates).

Pentland proposes *perceptual intelligence* as being key to interfacing with the coming generations of machines; he describes smart rooms and smart clothes—two classes of adaptive sensor-based environments—and technologies required to support them. Crowley et al. delve into the specific area of computer vision-based sensing and perception of human activity. They provide a broad view of the field and describe two projects that use visual perception to enhance graphical interfaces. Reeves and Nass address the need to better understand human perception and psychology as it relates to interaction with technology, and describe results from their human-centered experiments. The sidebars by Tan and Picard provide additional information about specific PUI research area, namely haptics and affective computing, while Bobick et al. describe a large-scale PUI application called the “KidsRoom.” **■**



## REFERENCES

1. Turk, M., Ed. *Proceedings of the 1998 Workshop on Perceptual User Interfaces*; [research.microsoft.com/PUIWorkshop](http://research.microsoft.com/PUIWorkshop); [www.research.microsoft.com/PUI-Workshop](http://www.research.microsoft.com/PUI-Workshop).

**MATTHEW TURK** ([mturk@microsoft.com](mailto:mturk@microsoft.com)) is a researcher in the Vision Technology Group at Microsoft Research in Redmond, Wash.  
**GEORGE ROBERTSON** ([ggr@microsoft.com](mailto:ggr@microsoft.com)) is a senior researcher at Microsoft Research in Redmond, Wash., working on 3D user interfaces and information visualization.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

© 2000 ACM 0002-0782/00/0300 \$5.00