Editor: Thad E. Starner ■ Georgia Institute of Technology ■ thad@cc.gatech.edu

# Wearable Agents

*Thad E. Starner, Georgia Institute of Technology*

Forty years ago, when batch processing was still the only common way of communicating with computers, Joseph Carl Robnett Licklider described a concept he called interactive computing. In time, Licklider would support research toward this goal by heading ARPA's new Information Processing Technical Office and, later, directing MIT's Project MAC. His reach was broad; M. Mitchell Waldrop's book The Dream Machine credits his ARPA funding program as the beginning of computer science graduate study as we know it today.[1] Although Licklider's ARPA researchers would go on to invent multitasking, hypertext, the mouse, the Internet, and many other mechanisms, Licklider's original goal was what he called "man–computer symbiosis," a concept he described in a 1960 paper by that name.[2]

## PERSONAL AGENT PIONEERS

As an MIT psychologist, Licklider seems to have thought of computers more as personal agents than as information processors. In an informal experiment, he kept a diary of his everyday actions: "About 85 percent of my 'thinking' time was actually spent getting into a position to think, to make a decision, to learn something I needed to know. Much more time went into finding or obtaining information than into digesting it."[2] In advocating a man–machine symbiosis, Licklider wanted to augment the thought process such that "human brains and computing machines will be coupled together very tightly and the resulting partnership will think as no human brain has ever thought."[2]

During the same time period, the early 1960s, computing pioneer (and Turing Award winner) Douglas Engelbart called for "augmenting intellect"[3]; his subsequent research foreshadowed much of what we take for granted in computer interfaces today (see Figure 1). The idea of mechanically augmenting human cognition goes back even further. In 1945, Vannevar Bush called for a device to aid human memory and thought (see Figure 2):

> *Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and, to coin one at random, "memex" will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.*[4]

Bush's vision of the future even included as one of its main ideas the ability for a user to capture information for his memex while mobile.

## DESKTOP VERSUS MOBILE

Unfortunately, in 1945, and indeed even in the 1960s, creating a cognitive coupling between a human and a computer was mostly limited to what was obtainable by sitting in front of a terminal and keying in requests. (A particularly interesting exception is the wearable computer Ed Thorp and Claude Shannon built in the 1960s to aid gamblers in predicting roulette.[5] I often refer to this machine as the first mention of an electronic wearable computer in the literature.)

Thus, by necessity, earlier researchers concentrated most of their efforts on creating systems appropriate for activities carried out on a physical desktop. However, human cognition is not confined to the desk; it is applied in the world at large. Often, a problem or situation's context in the physical world is the key to solving it. Much as a police detective must reconstruct clues in the context of a crime scene, an intelligent agent must reconstruct the context of its human partner to provide relevant assistance. Moreover, a mobile intelligent assistant has an advantage over a desk-bound one. It can observe and assist its user in situ and provide its informational services on a second-by-second basis. Thus, mobility could prove to be a key asset in creating intelligent agents.

Wearable computers provide a unique platform for creating such mobile intelligent assistants. Although current PDAs are powerful enough to run sophisticated programs, they usually reside in a user's pocket or briefcase and only see the light of day when the user wants to access them for specific functions, such


1536-1268/02/$17.00 © 2002 IEEE

as a calendar or phone list. These devices remain almost as deaf, dumb, and blind as their desktop counterparts. However, wearing a computer enables several opportunities for mounting sensing devices. For example, a camera could be bundled with a head-up display, or a microphone might be mounted next to the speaker on a pair of headphones. Such body-mounted sensors can see approximately the same view and be exposed to the same sounds as the user. This first-person view of the user's world is unavailable to desktop machines or pocketed PDAs.

Furthermore, if a user wears an interface continuously during waking hours, a wearable agent could collect a surprising amount of sensor data on the patterns of its user's everyday life. But how could this information be used? In the tradition of Licklider, one use could be training wearable agents to identify different user contexts and provide information support automatically during the day. In many senses, the wearable agent becomes the equivalent of an omnipresent and private virtual secretary. For example, the wearable agent—through the use of face recognition software and a camera mounted in the user's eyeglasses—could provide forgotten names during a dinner party.

**Body-mounted sensors can see approximately the same view and be exposed to the same sounds as the user.**

## WEARABLE INTERACTION AGENTS: AN EXAMPLE

My group at Georgia Tech is dedicated to creating wearable agents. By necessity, our work combines on-body perception, user modeling, and human–computer interfaces. The Web site, www.innovations.gatech.edu, demonstrates our first attempt at such an agent. The prototypes on the site predict the user's next location, perceive the user's availability, and help schedule an appointment based on a conversation with another researcher. The system is an effort to demonstrate how to make aspects of cell phone use more socially graceful through a wearable agent. As of this writing, an agent system that integrates all the demonstrated functionality is too imprecise and requires too many resources to be practical, but each perception subsystem described here is operational in the laboratory. For demonstration purposes, I have assumed much of the functionality that Chris Schmandt and his Speech Interface Group at MIT have demonstrated in the past.

Most of modern society is acquainted with the cell phone, and its ringing at inappropriate times is often the source of jokes and irritation. Although pundits often relate this phenomenon to connectivity addiction or information overload, inappropriate interruptions are often a result of not enough information instead of too much. If a phone agent knew more about its user's current context and the call's relative importance, it could better choose when to ring. Depending on available interfaces, the agent could choose alternative, less noticeable ways to announce a call, such as vibrating or identifying the caller on the user's head-up display. An even more substantive improvement would engage the caller in a dialogue with the agent to determine the call's importance relative to the user's current context.

Two of my students, Dan Ashbrook and Brad Singletary, are attempting to provide such context to a wearable agent. Ashbrook's system uses GPS coordinates gathered over a month of observing a user's movements.[6] His program clusters these coordinates to determine important locations in the user's life and then models the probability of moving from one location to another based on previous history. In a video on the Web site, Ashbrook's system determines that the user is walking from one building to another to prepare for class. Along the way, the user meets a student and begins a conversation. Software that Singletary is developing recognizes this event. Singletary's system uses a small head-mounted camera that is integrated on the earpiece of the user's eyeglasses. Through this camera, Singletary's system attempts to identify visual patterns associated with greeting a colleague and beginning a conversation.[7]

In the video, a hypothetical agent uses these perception systems to manage the user's cell phone calls. While the user walks to class, the agent intercepts a call from a colleague. Given that the colleague is a close collaborator, the agent asks if the call is important enough to interrupt the user's class preparation. If the caller indicates that the message is indeed urgent, the agent

Figure 1. In addition to inventing the mouse, Douglas Engelbart invented the first chording keyboard in his pursuit of augmenting intellect.

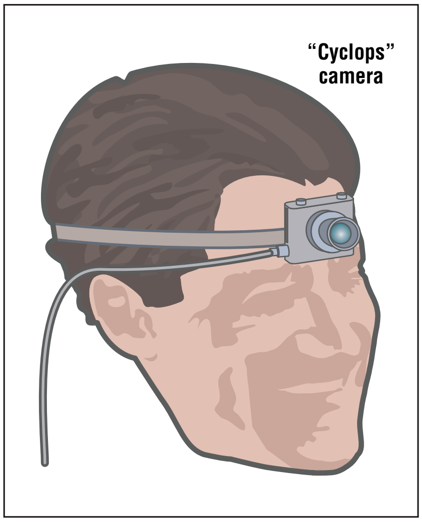


Figure 2. Besides foreseeing the Internet and machine speech recognition, Vannevar Bush described personal memory aids such as the Cyclops camera.

## next issue

**Is speech recognition appropriate for wearable interfaces? What makes speech interfaces so difficult to create? In the next issue, we'll explore popular misconceptions and the current research in the area.**

uses Singletary's system to determine if the user is immediately available and not currently engaged in conversation. The wearable agent determines that the user is currently "engaged" and asks the caller to leave a message. A note indicating the call and its originator appears on the user's display so that he or she can address the call later.

The video next shows the user returning the colleague's urgent phone call, and they begin negotiating a time to meet. Another student, Ben Wong, has written the Calendar Navigation Agent (CNA), which listens to the user's side of this conversation and helps determine his or her availability. Fragments of the conversation such as, "Let me see if we can meet next week" trigger the execution of the user's calendar program, which, correspondingly, displays next week's information. Similarly, "How does Tuesday sound?" zooms the calendar in on the entries associated with that week's Tuesday. In this manner, the agent can use the wearable's high-resolution head-up display to help negotiate an appointment quickly, simultaneously recording it in the calendar.

Creating conversational agents such as the CNA is extremely difficult. Recognizing conversational speech is significantly more complex than recognizing speech intended as dictation. The CNA only works because of the relatively severe constraints we place on the problem. Mobile speech recognition is an active research issue. Many wearable computer manufacturers, including IBM in its famous St. Mark's Square commercial, have emphasized speech as a primary interface, but is speech really appropriate? Three-thousand-word speaker-independent recognition is possible on 486-class machines, which is well within the realm of current portable computers. Why, then, hasn't speech overtaken pen input for mobile devices? I'll address these questions and more in the next issue.

### REFERENCES


1. M. Waldrop, *The Dream Machine: J.C.R. Licklider and the Revolution that Made Computing Personal*, Viking Press, New York, 2001.
2. J.C.R. Licklider, "Man–Computer Symbiosis," *IRE Trans. Human Factors in Electronics*, vol. HFE-1, no. 1, Mar. 1960, pp. 4–11.
3. D. Engelbart, *Augmenting Human Intellect: A Conceptual Framework*, tech. report, AFOSR-3233, Stanford Research Inst., Stanford Univ., Menlo Park, Calif., Oct. 1962.
4. V. Bush, "As We May Think," *Atlantic Monthly*, vol. 76, no. 1, July 1945, pp. 101–108.
5. E. Thorp, "The Invention of the First Wearable Computer," *IEEE Int'l Symp. Wearable Computers*, IEEE CS Press, Los Alamitos, Calif., 1998, pp. 4–8.
6. D. Ashbrook and T. Starner, "Enabling Ad-Hoc Collaboration through Schedule Learning and Prediction," to be published in *Workshop on Mobile Ad Hoc Collaboration*, 2002.
7. B. Singletary and T. Starner, "Learning Visual Models of Social Engagement," *Proc. 2nd Int'l Workshop Recognition, Analysis, and Tracking of Faces and Gestures in Real-time Systems*, IEEE Press, Piscataway, N.J., 2001, pp. 141–148.


### UPCOMING EVENTS AND USEFUL URLS

**ACM International Conference on Mobile Computing and Networking (Mobicom 2002),** 23–28 Sept. 2002, www.acm.org/sigmobile/mobicom/2002

**International Conference on Pervasive Computing (Pervasive 2002),** 26–28 Aug. 2002, www.pervasive2002.org

**International Conference on Ubiquitous Computing (Ubicomp 2002),** 29 Sept.–1 Oct. 2002, www.viktoria.se/ubicomp

**International Symposium on Wearable Computers 2002,** 7–10 Oct. 2002, http://iswc.gatech.edu

**Innovations @ Georgia Tech,** www.innovations.gatech.edu

**The Speech Interface Group,** www.media.mit.edu/speech

**Thad E. Starner** is an assistant professor of computing at the Georgia Institute of Technology, where he directs the Contextual Computing Group in the Institute's College of Computing. His research interests include wearable computing, augmented reality, machine and human vision, intelligent agents, and gesture recognition. He received four degrees, including his PhD, from the Massachusetts Institute of Technology, while working with the MIT Media Laboratory. He also cofounded the IEEE International Symposium on Wearable Computers and is a cofounder of Charmed Technology. Contact him at Georgia Tech., 801 Atlantic Dr., Atlanta, GA 30332-0280; thad@cc.gatech.edu.