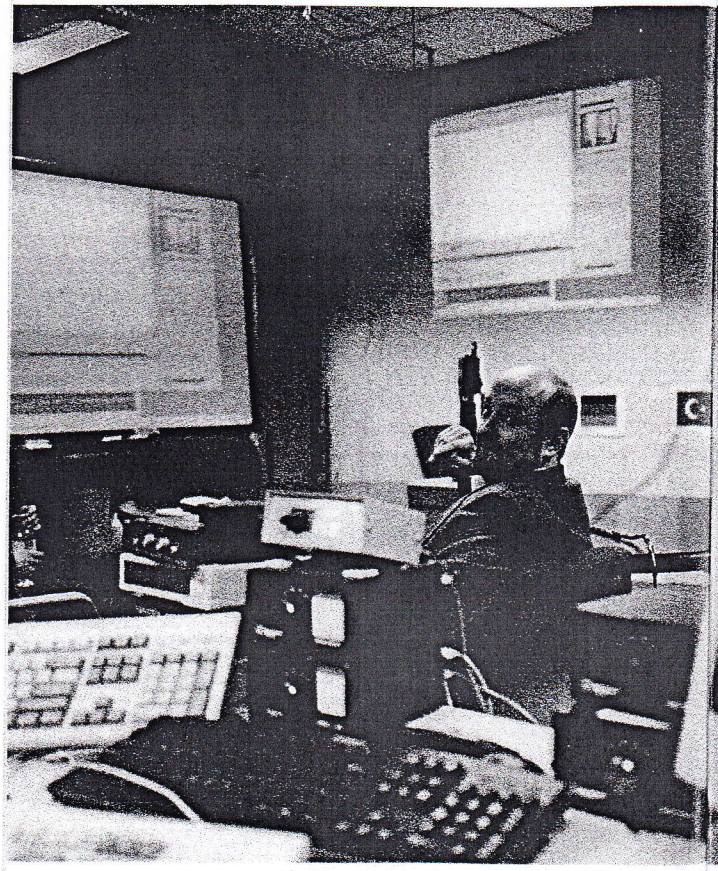
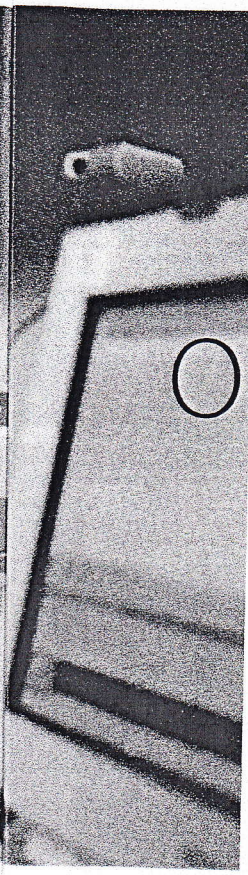


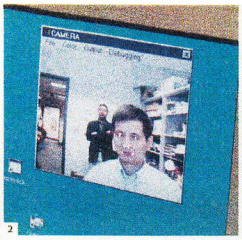


DEMO

Imagine a computer that listens to you, watches you, and gives you what you want exactly when you want it. **Alex Waibel** is building that computer. **Photographs by Beth Perkins**

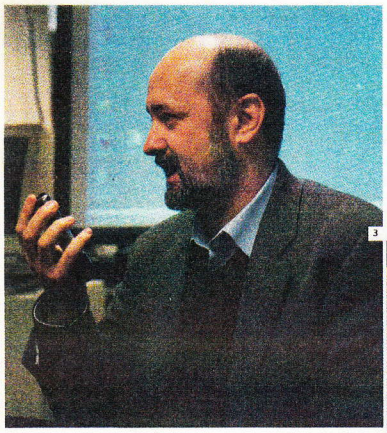
The Observant Computer

Alex Waibel is disappointed in his computer. "It doesn't care what I do and who I am and where I sit," says the director of Carnegie Mellon University's Interactive Systems Laboratories. "It just doesn't do anything until I do something—until I hit some button." Splitting his time between his lab at Carnegie Mellon and a sister lab at Universität Karlsruhe in Germany, Waibel is working to free humans from that forced interaction with machines. His model for the ideal computer? "A good butler or a good secretary—someone who invisibly hovers in the background, guesses your very needs, and serves them up before you even ask." That way, he says, humans would be free to interact with other humans and, as he puts it, "do the human thing." Computers would observe what the humans were doing—and understand it sufficiently to guess how to help the humans out. Sitting at a table in his office with *Technology Review* senior editor Rebecca Zacks, Waibel explains how that might play out. "I want to talk with you and then," he says, craning his head toward his desktop machine, "say, 'Oh, by the way, write a letter to so and so and tell him I can't do the review.' But how should the machine know that I'm now talking to it and not to you? If I say something about deleting the files, I don't want to have my computer go off and delete all the files. It needs to know who is being addressed." That's just one of a number of obstacles standing between Waibel and the computer of his dreams. He told Zacks about a few more hurdles and showed her what his team is doing to clear them.



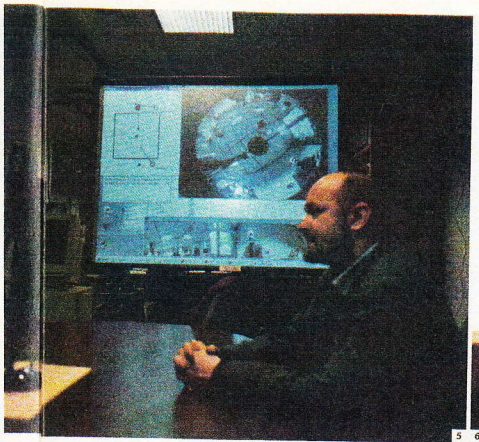
1 2

1-2. Some of the first tasks for an observant computer, Waibel says, are visual: "finding out that there are people in the room, that these people are moving, and also what are they doing." Down the hall from Waibel's office at Carnegie Mellon, research computer scientist Jie Yang's desktop computer is outfitted with a small black camera and image-processing software capable of doing just that. The computer finds *Technology Review* photo editor Lisa Petrie and Waibel in the room *(1)* and colors them red and green to mark each as unique. Yang (foreground) then poses for the camera himself *(2)*, and the software demonstrates its ability to distinguish a human face by marking Yang's eyes, nostrils, and the corners of his mouth with small white marks. "Once we have the face," Waibel says, the computer can begin to ask, "Where is the face pointing? What is the person looking at or attending to?"



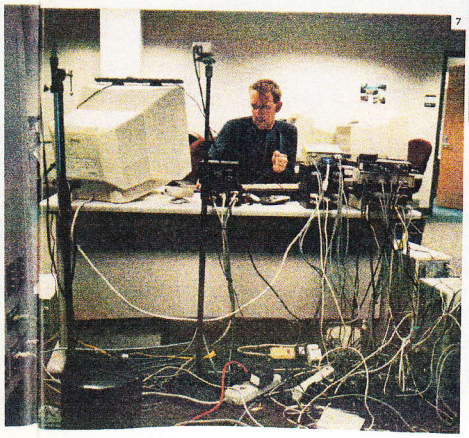
3 4

3-4. To understand people, computers will also have to understand their speech—in a variety of languages—so Waibel's ideal computer might one day draw on the expertise of a handful of translation systems his lab has produced. He demonstrates the prowess of one such system—a handheld computer loaded with software that helps doctors and patients converse when one party speaks Arabic and the other speaks English.



5 6

5-6. Waibel's most ambitious attempt to date at developing computers' powers of observation and comprehension takes the form of a makeshift meeting room. The room is rigged with microphones and a camera that sits in the center of a large conference table; the camera is aimed downward at a domed mirror, so it is afforded a simultaneous view of everybody at the table. The idea is, at first blush, a simple one: a computer will observe the meeting and act as a stenographer and assistant. "To put you in the right sense of awe of how incredibly difficult all of this is," Waibel says, "imagine at the end of the meeting, you ask a simple question of the computer: 'Why did Joe get angry at Bob when they talked about the budget?' And now imagine what knowledge you would have to have in order to answer it. First of all, you need to know that Joe was present and Bob was present. So somehow you need to do speech identification, face identification. Second, you need to know that Joe has addressed Bob and that the emotion during that time was angry. So you need to know what the focus of attention was and that the emotion was angry and that the topic of discussion was budget."



7

7. Even determining that Joe is addressing Bob is more complicated than simply noting that Joe has turned his face toward Bob, Waibel says; people tend to be lazy, moving their bodies only as much as necessary to see what they want with a sidelong glance. To understand the focus of somebody's attention, he says, "we actually have a statistical model that combines head direction over time together with some notion of what potential targets of interest are—human faces and people speaking, for example." The computer, off to one side, takes in data from the microphones and camera and feeds the information into a host of such models, crunching the numbers in a valiant attempt to see, hear, and understand.