TRN's
Making the Future Report
The State of an Emerging Technology and a Look at What Lies Ahead

Report Number 3 February, 2003

Computer Interfaces: Hands, Eyes, Voice and Mind

Executive Summary

Research about graphical user interfaces aims to make them dynamic without distracting, able to convey rough sketches, and able to present information at the periphery of users' awareness.

But the greatest potential for change in the human-computer relationship is conveying information from people to computers more quickly and efficiently.

One avenue is to augment or altogether replace keyboards and mice. Three major technologies that promise to allow people to control computers more naturally are speech recognition, gesture recognition and gaze tracking. These technologies are particularly useful in situations where manual control of a computer is difficult, which is increasingly the case as computers shrink and become embedded in clothing, jewelry, buildings and cars.

With the cost of digital cameras falling and researchers improving software that distinguishes arms, hands and fingers, simple gesture interfaces could soon be ready for applications like videoconferencing, kiosks and digital command posts.

Much of human-computer interaction research is dedicated to multimodal interfaces, which allow for mixed inputs that enable computers to recognize, for example, communications that combine speech and gesture.

Computers that are aware of where people are and what they are looking at will become increasingly important as the devices blend into the fabric of daily life. Several technologies allow computers to determine where people are in space: visual object recognition, infrared heat sensing, pressure sensing and sound localization.

Giving computers the abilities to see people, understand speech, decipher gestures and process and interpret natural language will go a long way toward the ultimate goal of making interacting with a computer as natural as interacting with another person.

Another, longer-term route to improving human-computer communications is to directly connect computers and neurons. Several research projects have shown that monkeys can control cursors on computer screens using signals from electrodes surgically implanted in their brains.

What to Look For

Existing input devices:

Low-cost and more accurate gaze tracking system
Low-cost and more accurate gesture recognition system
Low-cost and more accurate object recognition system

Combined Input:

Fully integrated speech, gesture and object recognition

Interfaces that tap human subtleties:

Speech recognition software that uses prosody
Systems that recognizes conversational gestures
Systems that recognizes basic emotions

Direct connections:

A monkey using a brain implant to consciously control a robot arm
A brain implant connected to thousands of neurons
A brain implant that restores paralyzed limb control

When reading is writing

Researchers at the University of Cambridge in England have made a computer interface that is like something out of Alice in Wonderland—you write by reading. The computer tracks your eyes, strings together the letters you look at, and puts the most likely next letters in the spot where you're eyes are most likely to go.

The unusual interface is one of many ways scientists and engineers have devised for people to interact with computers. Creativity is not a problem in computer interface research. The bottom line, however, is finding methods people will use.

People have a low tolerance for interfaces that are hard to learn or taxing to use, simple inertia makes them reluctant to venture from the familiar even when superior technology comes along, and they are reluctant to pay a premium for new ways to use their computers.

The principal challenge is producing high-quality, intuitive interfaces that won't hog a computer's resources or require expensive hardware.

The ultimate goal is making interacting with a computer as natural as interacting with another person.

You just don't understand

There is a long way to go. Computers may be blazingly fast at many tasks, but they are slow to take in instructions and data directly from people. Current human-computer interaction is a very unequal relationship, in fact. Computers offer a rich mix of text, graphics, video and audio information, but humans provide a paltry trickle of keyboard and mouse clicks—even though we are capable of much more.

So while computers could do a better job of presenting information—good synthesized speech comes to mind—the greatest potential for change in the human-computer relationship lies in blasting open the narrow pipeline into computers. Improving computer interfaces will speed computer input, which should significantly increase computers' productivity.

This is a tall order, however. It means giving computers the senses of sight, hearing and touch so they can recognize these types of information, and giving them a facsimile of the human mind's powerful pattern recognition capabilities so they can interpret the information.

This report maps out the four major thrusts in human-computer interaction research.

- Improving display screens controlled by pointing devices, which is currently the principal communication channel between people and computers.
- Conveying information to computers in alternative ways, including speech recognition, gesture recognition, and gaze tracking.
- Giving computers a broader awareness of humans through mixed modes of communication and tracking people's positions and movements.
- Providing a direct link between the human mind and computers.

On screen

Although today's graphical user interfaces are far more advanced than the command-line interfaces of the past, they are still restrictive. Window-and-mouse interfaces are document-centric and evolved from the rectilinear layouts of the print world.

How It Works

People detect meaning in even subtle facial expressions. Because this seemingly basic ability to read each other has evolved over millions of years, however, we tend to take our communications skills for granted.

Computers, on the other hand, have more in common with toasters than the human brain. There's much work to be done to make computers communicate with us on anything like the terms we're accustomed to.

The underlying technology that lets computers hear words, follow a gaze, pick up gestures, and keep track of a person moving around a room is pattern recognition software.

The first task of pattern recognition software is finding patterns in streams of raw data like digital audio or video. The software then matches the patterns to structures it knows. Patterns can include words, gestures and human bodies.

Four types of pattern recognition software are key to computer interfaces: artificial neural networks, the hidden Markov model, nearest neighbor and support vector machines.

Artificial neural network software mimics the basic structure of biological brains. The software learns what a hand looks like by building a pattern of connections among brain-cell-like components through repeated exposure to stimuli like digital video images of hands.

The hidden Markov model divides sensory input into a series of extremely brief events — typically thousandths of a second — and makes predictions about the nature of an event based on the event before and the one after. A sequence of events could be the sounds that make up a spoken word.

Nearest-neighbor algorithms map sensory input as points in an imaginary space, then classify the points assuming that points near to each other are similar. Two points representing video images of hands, for example, would be closer than a point representing a hand and a point representing a head.

Support vector machines map sensory input features statistically. The software recognizes patterns by comparing the ways a feature map resembles the maps of examples it was trained on.

Tracking and Coordinating

Recognizing patterns, however, gets a computer only so far. Once it learns to recognize an eye, a hand or a human, a computer has to be able to track the object as it moves.

Object tracking software measures an object, predicts where the object will be next, narrows the area to be measured based on the prediction, and uses new measurements to improve the predictions.

One way to loosen things up is to use animation techniques that make on-screen objects more tangible and dynamic. Techniques refined by cartoonists over more than half a century can make screen elements that change size to convey the impression that they occupy three-dimensional space, and stretch and warp to give the illusion that they react to stimulus. It's a major challenge, however, to produce helpful effects and avoid distracting ones.

Researchers are also loosening up the information that graphics convey. A team from the University of California at Berkeley and Carnegie Mellon University has developed a set of design tools that allows users to make rough sketches of, among other things, user interface designs. The usually polished look of information presented on computer screens turns out to hinder the design process because people are hesitant to suggest broad changes to something that appears to be nearly finished. Giving computers the ability to portray rough information expands their usefulness by allowing them to serve a role still largely filled by paper and pencil. (See Rough Tools Smooth Design, page 8)

Another avenue of research is a bit making computer interfaces mesh more closely with the ways humans shift attention and use peripheral awareness. Several research teams have developed ambient displays that convey information at the edge of a computer screen, on a separate screen or on appliances or furniture. The displays update information through changes to color, shape, patterns and movements. The information changes subtly in order to avoid interrupting the user.

Graphical user interface research benefits from well-developed tools and methods that allowed the techniques to be implemented quickly. But graphical user interfaces are also a firmly entrenched aspect of the computer industry. How research initiatives influence interface development is largely determined by business considerations.

Getting the point across

Pointing devices go hand in glove with graphical user interfaces. Given the plethora of choices available, including mouse, trackball, pen tablet, joystick and data glove, there is little research into entirely new devices.

One new approach, however, is is to reproduce an input device's functions without requiring the user to hold a physical object. Researchers at NASA have developed a joystick alternative based on electrodes that monitor the electrical activity generated by muscles in the arm. (See Muscles Tapped for Virtual Input, page 16)

Researchers are also working to remove the burden of using input devices. Three major technologies allow people to control computers sans pointing devices:

- Speech recognition
- Gesture recognition
- Gaze tracking

These efforts, some of which began decades ago, have the potential to be particularly useful in situations where manual control

Even being able to track and interpret the types of input humans use to communicate — gestures, words and facial expressions — is not enough. Meaning is often conveyed by a combination of different types of sensory input. Words and gestures, for example, can go together to produce meaning that cannot be determined from simply examining the inputs separately.

To tackle this problem, the computer needs to recognize and interpret each type of input, track the timing of the inputs, group segments of sensory input from each type chronologically, then combined segments, for example, words denoting space with pointing gestures. Then all that's left is interpreting the combination to extract its meaning.

Who to Watch

Screen:

Paul Aoki, Palo Alto Research Center
Palo Alto, California
www2.parc.com/csl/members/aoki/

Paul Dourish, University of California at Irvine
Irvine, California
www.ics.uci.edu/~jpd/

Scott Hudson, Carnegie Mellon University
Pittsburgh, Pennsylvania
www.cs.cmu.edu/~hudson

James Landay, University of California at Berkeley
Berkeley, California
www.cs.berkeley.edu/~landay

Brad A. Myers, Carnegie Mellon University
Pittsburgh, Pennsylvania
www.cs.cmu.edu/~bam

Terry Winograd, Stanford University
Stanford, California
hci.stanford.edu/~winograd/

Speech:

Julia Hirschberg, Columbia University
New York, New York
www1.cs.columbia.edu/~julia/

Elizabeth Shriberg, SRI International
Menlo Park, California
www-speech.sri.com/people/ees/

Victor Zue, Massachusetts Institute of Technology
Cambridge, Massachusetts
www.sls.lcs.mit.edu/zue/zue.html

of a computer is awkward or impossible. These techniques are becoming increasingly important as computers shrink and become embedded in clothing, jewelry, buildings and cars.

The most widespread alternative computer input technology is speech recognition, which has been commercially available for more than a decade. In addition to taking dictation, speech recognition software can control desktop computers and navigate telephone-based voice response systems.

Speech is the primary mode of communications between people, but talking to today's speech-enabled computers is vastly different from talking to a person. Much of the research into speech interfaces aims to narrow the gap by improving a computer's ability to pick words out of the continuous sound stream of speech. One approach is to allow computers to listen for more than just words.

Human speech contains several streams of communication; words are only the most prominent. The embellishments of speech—duration, pitch and emphasis—convey a multitude of meaning. Experiments in having computers listen for these prosody attributes of speech in recordings show that paying attention to between-the-lines information improves automated segmentation of recorded speech. Existing speech recognition software may have to be rewritten to better incorporate prosody.

Gaze tracking uses digital cameras and computer vision software to sense where on a screen a user is looking. A person controls a computer or composes words by scanning on-screen buttons or letters and blinking at or lingering over his selections. Gaze tracking is relatively slow as a sole means of controlling computers, but it is an important tool in the larger effort to increase the bandwidth of communications with computers. It is also a good recourse for the severely disabled.

Pointed gestures

For more than a decade, researchers have programmed data gloves to recognize hand positions and movements as commands. Data gloves are widely used to navigate virtual reality environments because they are easy to use and relatively unobtrusive. They are still input devices, but they point out the potential for more natural, nonverbal communication with computers.

Alex Waibel, Carnegie Mellon University
Pittsburgh, Pennsylvania
www-2.cs.cmu.edu/afs/cs.cmu.edu/user/ahw/www/

Gesture/Multimodal:

Gregory Abowd, Georgia Institute of Technology
Atlanta, Georgia
www.cc.gatech.edu/fac/Gregory.Abowd/

Justine Cassell, Massachusetts Institute of Technology
Cambridge, Massachusetts
web.media.mit.edu/~justine/

Joe Paradiso, Massachusetts Institute of Technology
Cambridge, Massachusetts
web.media.mit.edu/~joep/

Rajeev Sharma, Pennsylvania State University
University Park, Pennsylvania
www.cse.psu.edu/~rsharma/

Mandayam A. Srinivasan, Massachusetts Institute of Technology
Cambridge, Massachusetts
rleweb.mit.edu/rlestaff/p-srin.htm

Matthew Turk, University of California at Santa Barbara
Santa Barbara, California
www.cs.ucsb.edu/~mturk/

Neural:

Peter Fromherz, Max Planck Institute for Biochemistry
Martinsried, Germany
www.biochem.mpg.de/mnphys/

Miguel Nicolelis, Duke University
Durham, North Carolina
www.neuro.duke.edu/Faculty/Nicolelis.htm

Observe two people engrossed in conversation: chances are they gesture frequently. There are three types of conversational gestures: pointing, mimicking and marking rhythm. Two basic elements that allow computers to recognize gestures are vision systems that capture gestures and artificial intelligence software interprets them.

Decades of gesture recognition research has resulted in computers that are reasonably good at recognizing pointing, the simplest human gesture and the most useful one in human-computer interaction.

In recent years the cost of digital cameras has decreased and researchers have improved software that distinguishes arms, hands and fingers. This means simple pointing interfaces could be deployed on personal computers for applications like augmented reality and videoconferencing within a few years. The technology is already a basic component of smart spaces and emerging command-post applications.

Pointing is the metaphor used by graphical user interfaces, and researchers have adapted the interfaces to allow people to literally point at windows and icons. These pointing interfaces usually require an oversized screen or a display projected on a wall or surface. Because people are more likely to project maps and images than desktop computer displays, most of these interfaces are designed for kiosks and command posts.

Pointing is also used in augmented reality systems that recognize objects like books and sketch pads and project images onto objects or nearby surfaces. This allows a person to point at a page in a book, for example, to command a computer to project related information next to the book. Researchers at the University of Electro-Communications in Japan and the University of Tokyo have developed this type of interface for textbooks. The interface also allows users to control projected information with gestures. (See PCs Augment Reality, page 22)

Mimicking gestures, which describe objects, shapes or movements, are much more difficult for computers to recognize and understand. The challenges include determining when one gesture ends and another begins, distinguishing among different types of gestures and interpreting them. Researchers have developed interfaces that let people issue commands by holding up a certain number of fingers and by opening and closing their hands.

Finding a way to enable computers to follow conversational gestures is a major challenge in artificial intelligence research, and full gestural literacy for computers is at least a decade away.

Multiple doors of perception

Alternative inputs are only a step toward fulfilling the promise of human-computer interaction. Though we might never develop truly intelligent computers, existing technology holds the potential for building interfaces that allow us to communicate with a computer as though we were communicating with another person.

The next step is giving computers the ability to interpret several signals at once. This means more than simply recognizing more than one type of input. The challenge is giving computers the ability to match different inputs so they can interpret the mixed commands common in human-human communications. For example, when you tell someone to pick "that" up as you point to an object, they automatically combine the two forms of communications into a single instruction. The spoken command doesn't specify the object to perform the action on, and the gesture doesn't convey the action to perform on the object.

Much of the research on human-computer interaction is dedicated to developing multimodal interfaces, and much of that work is focused on making it possible for computers to recognize a mix of speech and gesture. The initial work in this field was centered on integrating speech recognition software and pen tablet input.

One recent thrust is using prosody to improve gesture recognition. Multimodal interfaces could be ready for use in smart kiosks and command posts in two to five years. (See Interface Gets the Point, page 23)

A full body experience

Another line of research aims to give computers the ability to track people's positions, postures and gazes. Making computers aware of where people are and what they are looking at will become increasingly important as computers blend into the fabric of daily life. There are several technologies that enable computers to determine where people are in space:

- Visual object recognition
- Infrared heat sensing
- Pressure sensing
- Sound localization

Object recognition software allows computers to recognize that the shapes characteristic of two eyes, a nose and a mouth compose a face. Recognizing a face allows the computer to determine which way a person is facing, and hence where her attention is likely to be. Gaze tracking, which senses the orientation of a person's eyes to determine where she is looking, is also a prime candidate for letting a computer know where human attention lies.

These technologies, along with gesture recognition, can give a computer a reasonable sense of where a person's attention is focused, assuming the person is making unambiguous movements, not dividing her attention, and is in an uncluttered environment. These capabilities have been demonstrated in carefully prepared laboratory facsimiles of controlled environments like command posts and conference rooms.

The technology could become practical for use in those types of environments within five years. Crowded public places are likely take more than a decade to master.

Getting emotional

Beyond the hard realities of physical presence lies the fuzzy realm of human emotion. Computers would be a lot more useful if they could gauge human emotion and adjust their responses accordingly. And making human-computer interaction as close to human-human interaction as possible also means giving computers the ability to express emotion.

Researchers are mapping facial expressions in an effort to prepare a rudimentary emotional vocabulary for computer vision systems, and similar audio efforts are aimed at giving computers the ability to hear emotion in human voices.

There are also efforts to add biometrics to the mix by tapping capabilities computers can have that humans don't, like infrared vision. Researchers at the Mayo Clinic and Honeywell Laboratories have developed a method for spotting a telltale pattern in the heat distribution of a lying person's face. (See Hot Spots Give Away Lying Eyes, page 29)

Though work is well underway on emotion recognition, little has been done to make computers respond differently based on emotional cues.

Eventually emotion will become part of both sides of human-computer communications. The emerging field of affective computing aims to give computers the ability to provide facsimiles of human emotion. This work is just beginning, and much of it involves artificial intelligence learning systems that require expensive training and feedback from human handlers.

Fully realizing the potential of human-computer interaction might require embodied computer interfaces, namely robots. Research has shown that people respond viscerally to the physical presence of robots. (See Manners Matter for the Circuit-Minded, page 30)

People are also prone to projecting human qualities onto animals and machines, and researchers are tapping this tendency in order to create the illusion that computers and robots are intelligent and emotional. A Carnegie Mellon University project has produced an animatronic robot that tells jokes and engages in rudimentary conversations. (See Interactive Robot Has Character, page 31)

The ideal computer interface could well turn out to be a humanoid robot that would follow your eyes, understand your gestures, pick up on your emotions, and be able to respond in kind. This would require pulling together many pieces of technology, including natural language processing software that gleans meaning from words and phrases, dialogue management tools that handle the back and forth of conversations, context awareness software that recognizes the subject of a conversation, and natural language generation programs that produce natural phrases and sentences, not to mention the extraordinary engineering challenge of producing mobile, expressive humanoid robots. This level of sophistication in human-computer interaction is decades away.

Brain jacks

Another approach—long a favorite of science fiction writers—is aimed at more transparent human-computer interaction. There have been several attempts to use brain waves to control computers. These interfaces, however, have provided very coarse-grained control, and people often find deliberately altering their brain waves to be an exhausting experience.

One way to make human-computer communications more direct is to connect to the rich variety of signals inside the brain. Several research projects have surgically implanted electrodes in the brains of monkeys.

Duke University Medical Center, Massachusetts Institute of Technology (MIT) and the State University of New York (SUNY) Health Science Center researchers tapped signals generated by a monkey's motor neurons as it moved its arm, and used the signals to duplicate the arm motions with a robot hundreds of miles away over the Internet.

Brown University researchers used a similar setup to allow a monkey to control a cursor on a computer screen by moving its arm. A neural probe experiment at Arizona State University gave monkeys control over a 3D cursor, and one of the monkeys learned how to control the cursor without visibly moving its arm. (See Monkey Think, Cursor Do, page 33 and Brain Cells Control 3D Cursor, page 34)

Researchers have also used neural probes to control animals, showing that the brain interfaces can be two-way. Researchers at SUNY Downstate Medical Center and Drexel University implanted a radio-controlled electrode in a rat's brain and used the signals to remotely control the rat's movements. (See Virtual Touch Controls Rats, page 35) The project's lead researcher shared the Defense Advanced Research Projects Agency (DARPA) 2002 Award for Sustained Excellence by a Performer for the work.

These technologies could eventually give people who have severe disabilities greater control over computers. The brain surgery the technology requires is likely to severely limit the use of these interfaces, however.

Cyborg science

On the other hand, why stop at neural probes? Why not more extensive electronics a la the cyborgs of science fiction?

Researchers are laying the groundwork by exploring how nerve cells and semiconductor electronics interact. Researchers at Max Planck Institute for Biochemistry have induced a feedback loop between a nerve cell and a field effect transistor. (See Neuron-Chip Link Advances, page 36)

The main challenge is that cell communication is electrochemical, and in order to communicate with cells electronic circuits need to be able to induce and interpret the chemical changes that cells undergo. Initial forays suggest that nerve-electronics integration is possible. Any practical developments, however, are decades away.

It's not clear what would happen if we were able to integrate electronics into living human brains. It's not known how people would perceive signals from imbedded electronics, whether electronic devices could be configured to transmit meaningful and useful signals to the brain, and whether the mind would be able to receive input from an imbedded electronic device more quickly than it can process information through the senses.

Until we have a much better understanding of how the brain works, it will be difficult, and probably foolish, to try to augment it.

Integrating nerve cells and electronics could, however, be a big help in learning how the brain works, and could lead to instruments that would collect significantly more data about the brain than the electroencephalogram (EMG) and functional magnetic resonance imaging (fMRI) systems in use today.

Changing relationship

The human-computer relationship is poised to improve dramatically, fueled in part by the evolution of the computer from a box sitting on your desk to a component of everyday objects around you.

The good news is it seems likely that computers will have plenty of horsepower to process the huge amounts of raw data needed to absorb the dense and varied information conveyed by humans. The bad news is that the real challenge lies in re-creating the intelligence behind the interpreting abilities we take for granted, especially since we don't fully understand how we ourselves work.

Recent Key Developments

Advances in screen interfaces:

- Software tools for roughly sketching designs (Rough Tools Smooth Design, page 8)
- An experiment showing that animation techniques can improve graphical user interfaces (Cartoons Loosen Up Computer Interfaces, page 10)
- A PDA graphical interface that discreetly fades out of the picture (PDA Interface Keeps a Low Profile, page 11)
- A Web interface that adjusts to user preferences on the fly (Software Guides Museum-Goers, page 13)

Advances in hands-free input:

- A gaze-tracking interface for composing text (Software Turns Reading Into Writing, page 14)
- An arm muscle sensor that replaces the joystick (Muscles Tapped for Virtual Input, page 16)
- A winking input device based on a wireless sensor in a skin patch from the Academy of Finland

Advances in speech and sound interfaces:

- A project to incorporate prosody — volume and inflection — into speech recognition software (Hearing between the Lines, page 17)
- An experiment that shows that beginners, intermediates and experts need different ways of correcting speech recognition errors (Correction Choices Key for Speech Software, page 18)

- Software that makes it easier to query large databases using speech recognition (Two-Step Queries Bridge Search and Speech, page 20)
- Software that improved programmers' efficiency by having errors generate distinct sounds (Programming Tool Makes Bugs Sing, page 21)

Advances in gesture and multimodal interfaces:

- A gesture touchpad developed at the University of Delaware
- A system for projecting a computer display onto a desk and controlling it with gesture recognition (PCs Augment Reality, page 22)
- An interface that uses prosody to improve gesture recognition (Interface Gets the Point, page 23)
- An interface that matches pointing gestures and speech (Interface Lets You Point and Speak, page 25)
- Software that coordinates speech, eye-tracking and data glove input (Integrated Inputs Improve Interactivity, page 26)

Advances in computer senses and emotion:

- A computer vision system directed by sound localization technology (Sounds Attract Camera, page 27)
- A smart chair that tracks a person's posture (Biometrics Takes a Seat, page 28)
- An infrared vision system that spots liars by the blood flow around their eyes (Hot Spots Give Away Lying Eyes, page 29)
- A study that shows that humanoid robots will need to be taught socially acceptable behavior (Manners Matter for the Circuit Minded, page 30)
- An interactive animatronic robot that answers questions and tells jokes (Interactive Robot Has Character, page 31)

Advances in brain interfaces:

- A brain implant that lets a monkey control a computer cursor by moving its arm (Monkey Think, Cursor Do, page 33)
- A brain implant that lets a monkey control a virtual ball in a 3D space by thought alone (Brain Cells Control 3D Cursor, page 34)
- A wireless brain implant that let researchers control a rat's movements (Virtual Touch Controls Rats, page 35)
- A feedback-loop between a nerve cell and a transistor (Neuron-Chip Link Advances, page 36)
- A method for binding bits of semiconductor to nerve cells (Nerve-Chip Link Closer, page 37)
- An implant that melds with nerve cells (Implant Links Nerve Cells to Electronics, page 39)

Rough Tools Smooth Design

By Kimberly Patch, Technology Research News
May 16, 2001

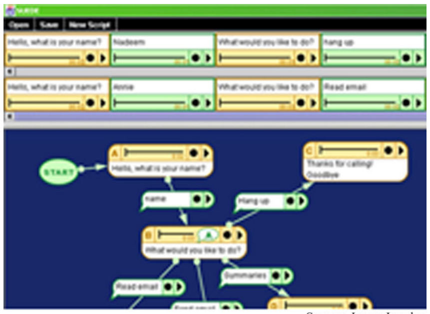
It's clear that the computer is not the best tool for every step of a design process despite its organizational advantages.

Many designers, from architects to Web interface builders, use pencil and paper during the initial stages of creating a design. One reason is that a design that is laid out neatly on a computer looks done, which is an impression that squelches the creative process. The workings of a computer can get in the way as well, especially with interfaces like speech recognition that regularly produce mistakes.

One group of researchers has created three tools that sidestep these drawbacks by allowing computer interface designers to create rough designs on a computer. Silk is a tool for graphic designers, Denim a tool for Web site designers and Suede a tool for designing speech interfaces. What they have in common are interfaces that make it clear that the project is an early draft, and that eliminate technical issues that tend to get in the way of creativity.

"There's been a lot of work that shows that if the designer uses a pen or pencil on paper, they're more creative than someone who uses... a drawing program," said James Landay, an assistant professor of electrical engineering and computer sciences at the University of California at Berkeley. “The reason is that the ambiguity in the sketch actually encourages them to look at more ideas, and in the early stages of a design it’s really important to be creative and to look at many different approaches,” he said.

The same goes with designers showing clients rough ideas, said Landay. “A sketch... communicates that the idea is rough and unfinished and open to suggestion and also gets them to focus on those high-level issues,” he said.



Source: James Landay

The Suede tool allows designers to roughly plot out speech recognition interfaces.

Conversely, when a designer shows a client something on a computer, “it looks done so they tend to focus on small visual details that are important at some stage but not in the early stages where you’re trying to get high-level feedback,” he said.

But there are also reasons to design on a computer, Landay said. “It’s easy to edit, easy to reuse... easy to track different versions [and] easy to test, because when it’s on a computer we can actually turn it into something that runs and test it with people... whereas on paper we just have to pretend,” he said.

Silk allows a designer to literally sketch out an interface that works, meaning when a scribbled button is clicked, the button will execute an action. “You can actually try them out and see if they make sense, as well as show them to other people to illustrate how [the interface] is intended to work,” said Landay.

The Silk tool, which was Landay’s first project, goes the extra step of converting the sketch interface to a finished interface when the designer is done with the project. The other two tools don’t include this step, however, because the conversion wasn’t very useful for the designers, he said. “It’s interesting from a computer science perspective but it’s not as important to the designers,” said Landay. Once the designers finish and test a design in run mode to see that it works properly, they are likely to switch to a Web building tool anyway, he said.

Denim is a similar tool for Web interfaces that keeps track of how different pages, which can be roughly sketched out, link to each other. “We found that [Web designers] were using three different representations—storyboards, site maps and pages, and they had to use different tools for all of them. The idea was to unify these different representations all on one interface in one screen and use zooming as a way of moving between them,” said Landay.

A slider bar allows the designer to zoom in to see a single page and out to see a site map. “Arrows represent how [different pages] are related. [When] you zoom into an individual page, you can just draw on the page,” he said. The interface recognizes the arrows, and has a run mode where it works like a Web browser. “You can save this in HTML and run it as a Web browser, but it will look... sketched,” said Landay.

Suede brings the same basic ideas to designing voice interfaces. What gets in the way of being creative in designing a voice interface is the technology of speech recognition, which doesn’t always work perfectly. “Often these technologies lead you to start having a conversation with [the speech recognition software] rather than focusing on your task,” said Landay.

The interface sidesteps speech technology snafus simply by not using speech recognition at all. Instead, it uses something speech interface designers have historically used before they put a design on a computer. “The Wizard of Oz method... means someone pretends to be the computer... to test the interface,” said Landay.

Suede allows the designer to organize what the system will say by mapping out the different prompts and responses on the screen. It also facilitates test sessions where a person is acting as the voice recognizer, then analyzes the data. “There’s no speech recognition, there’s no speech synthesis. We can still design something really rapidly, and that’s what’s in common [between Suede] and the sketching systems that we built,” said Landay.

The next step is making the interfaces multimodal, said Landay. This would allow for two different types of input, like speech and pointing, at the same time. For example, a person could point to something with a pen and at the same time say ‘move this here.’

Source: James Landay

The Denim tool allows Web designers to build working Web sites by roughly drawing pages and the links among them.

The idea behind all three projects is to adapt computers to the way people work naturally instead of the other way around, he said. “My argument is that most computer programs and applications make us operate on the terms of the computer [where] everything is precise.”

They’re neat applications, said Terry Winograd, a professor of computer science at Stanford University. “What [Landay has advanced] is using informal things that actually work. The main issue here is the faster the feedback cycle, the more iterations you can do, [and so] the more you can debug your ideas as you develop them," he said.

"There's a whole body of work on prototyping techniques that emphasize things like informal sketching and ways of getting feedback from users based on various early designs" that usually involve media like paper, video and flipbooks, Winograd said.

The researchers' work brings a similar type of prototyping to computers "in a particular way—combining informal sketching with computer response," said Winograd. It is easy with some combination of techniques to go through the same process that the interface design tools enable, but the tools may allow you to learn more with less work, he added.

Source: James Landay

The top image shows a graphical interface drawn using the Silk tool. The drawn buttons work like the actual buttons of a finished user interface. The lower image shows the design converted into a standard interface.

Landay's research colleague for the Silk project was Brad Myers of Carnegie Mellon University. They published the research in the March, 2001 issue of IEEE Computer. Landay worked with Berkeley graduate students Mark Newman, James Lin, and Jason Hong on Denim, and Scott Klemmer, Anoop Sinha, and Jack Chen on Suede. The Silk project was funded by the Department of Defense (DOD) and Fuji Xerox Palo Alto Laboratories (FX Palo). The Denim project was funded by NEC, Qualcomm, and FX Palo. The Suede project is funded by SRI International, FX Palo, and the National Science Foundation (NSF).

Timeline: Now
Funding: Corporate, Government
TRN Categories: Software Design and Engineering; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "Sketching Interfaces: Toward More Human Interface Design," March, 2001, IEEE Computer. More information on the projects, including the videos of the interfaces in action and a downloadable version of Denim, is available at guir.berkeley.edu

TRN

Cartoons Loosen Up Computer Interfaces

By Ted Smalley Bowen, Technology Research News
July 18, 2001

Wile E. Coyote misdirects the elastic force of a giant slingshot; he plummets off a cliff, the inevitable boulder looming palpably overhead. The appeal of Saturday morning cartoons, which may have hit their peak with gems like Road Runner, stemmed in large part from the animators' knack for evoking an exaggerated sense of the laws of physics at work.

Researchers at two universities in South Australia are looking to adapt the mechanics of serious amusement to the minutiae of serious computer work by adding cartoon animation effects to graphical user interfaces (GUIs).

The work is an attempt to lend substance and dynamism to the generally flat and less-than-engaging graphical user interface, according to Bruce H. Thomas, director of the wearable computer lab at the School of Computer and Information Science at the University of South Australia.

With very few exceptions, today's GUIs are minimally responsive. When users select and drag objects, or pull down menus, the screen elements react with jumpy movements, hasty transitions, and an overall lack of stimulating feedback.

Although GUI animation has been an active field of research, its practical uses have been limited by a lack of suitable programming tools and the relative lack of computing power available to run applications, according to Thomas.

Although today's machines are about 100 times faster than the original Apple Mac, applications like word processors are not proportionally faster because system software has steadily claimed more of the computer's raw power. "The computers are fast, but when you add the system software the entire system is [relatively] slow," said Thomas. In addition, "the animation software tools have not been built into the user interface toolkits. Until it is easy to add animation, programmers will be reluctant to do so."

But as computing power increases, and as animation tools are added to the GUI programmer's palette, users could benefit from interfaces whose elements seem more substantial and responsive, according to Thomas.

Several animation techniques can bring GUI elements to life: keeping the cursor in contact with the object being manipulated, adding a sense of resistance to the object, showing change in a continuous manner and presenting a clear response for each action.

By warping, magnifying and shrinking objects, animators can give them the appearance of existing in three-dimensional, physical space, according to Thomas.

To test the effectiveness of animated GUIs, the researchers created a simple drawing application that used cartooning techniques to animate screen elements as people moved and changed them.

The researchers measured peoples' reactions to animation feedback as they moved objects on a screen. The researchers also measured how the feedback affected performance.

They tested four types of visual cues: no visible feedback during the move; handles added to a selected object; animation that showed the object stretching in the direction of the cursor but resisting the move as if rooted by gravity; and handles added to the animation effect.

Somewhat to the researchers' surprise, the feedback types yielded almost identical performance, leading them to speculate that the task was too simple to reveal different levels of effectiveness, said Thomas.

"The first task was very simple and repetitive. The subjects quickly learned to perform the task by rote learning," Thomas said. "I feel the greatest benefit is making the user's actions more understandable or legible. In more complex tasks users could make more mistakes and the time savings [would be] in the rectifying [of] those mistakes."

Subjects rated the effects on a scale of 1 to 7, with one representing strong affinity and 7 strong dislike. The animation-plus-handles feedback was most popular, rating a mean score of 2.4 compared to a 3.1 for handles only and 3.4 warping in the direction of the cursor, according to the research.

A second test gauged the users' preference for the degree of animation by allowing them to adjust the strength of the effects, from 0 to 20, with a slider control.

Most subjects played with the full range before composing with more than one setting. The average setting was 3. The subjects generally preferred animation, although there was no consensus that animation improved their work, according to the study.

Making the best use of animation in GUIs, according to the research, means showing subtle changes relating to the task at hand, and avoiding superfluous, distracting effects.

Indiscriminate animation of screen elements with exaggerated and sustained effects, for example, can turn off users. If every sweep of the cursor causes a dialog box, icon or block of text to move for no apparent reason, the GUI becomes a hindrance.

For example, early animated desktop icons were distracting because they were always running, said Thomas. "So you would have ten or twenty canned animations going on simultaneously on the desktop. It was too much motion on the screen, and distracted the user," he said.

The next step in the research is to add animation to computer-aided design (CAD) and mapping applications to provide visual cues of the constraints affecting objects, said Thomas. "Warping and animation effects can greatly enhance the visualization of constraints, which are very prevalent in both our mapping tool and CAD systems. We wish to provide animated visual cues to highlight the large and varied set of constraints associated with graph manipulation," he said.

Better support for graphics and animation in programming languages like Java is likely to crop up over the next two years, and more animated user interfaces will follow, said Thomas.

Thomas' research colleague was Paul Calder of Flinders University. A technical paper on the study is slated for publication in the September 2001 issue of ACM Transactions on Computer Human Interactions. The work was funded by the University of South Australia.

Timeline: >2 years
Funding: nbsp; University
TRN Categories: Software Design and Engineering
Story Type: News
Related Elements: Technical paper, "Applying Cartoon Animation Techniques to Graphical User Interfaces," slated for publication in the September 2001 issue of ACM Transactions on Computer Human Interactions.

TRN

PDA Interface Keeps a Low Profile

By Kimberly Patch, Technology Research News
February 21, 2001

Although humans are capable of doing several things at once, it's nice to be able to concentrate on, say, an 18th-century artifact without having to interrupt the experience to curse at a piece of electronic gadgetry that promised to give you more information about the artifact.

With that in mind, a team of researchers at Xerox tested a personal digital assistant (PDA) interface on tourists at a historic house, including those who said they did not get along well with computers.

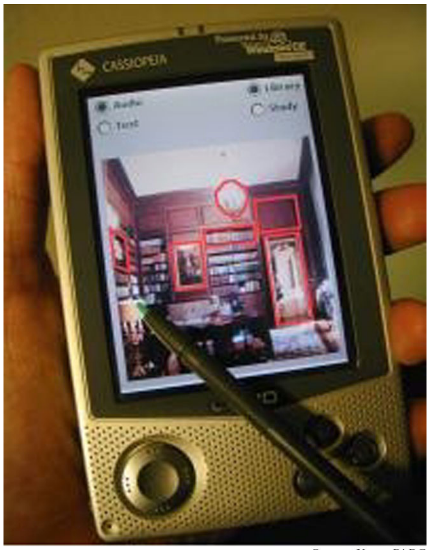
The Tap Tips guidebook interface is simply a picture of the part of the historic house that a person is looking at. For information on an artifact, the person uses the PDA stylus to tap on the picture.

The historic house was a challenge because, unlike a "buy" button on a Web site, it was not obvious that some of the objects pictured on the guidebook screen could be relevant. "In a house, something as simple as a ladder in a library to get up to books might be a historically interesting object," said Paul Aoki, a computer scientist at the Xerox Palo Alto Research Center.

However, giving the museum goers that explicit cue on the screen would require either having them remember how to turn the cue on and off, or make the cue permanent, which would ruin the picture and the flow of the experience, he said.

The researchers solved the problem simply. Although there's nothing on the interface to tell a person where to tap, touching a place that does not harbor information causes outlines to appear briefly on the objects that do yield information. Because the outlines appear only when they are needed, then fade on their own, no knowledge is required to turn them on and off.

"The point is when... you're trying to make [an interface] very simple for people to understand, the more of those special-purpose little buttons and gadgets and widgets on the screen that you have to understand, the more difficult it is... to remember them all," Aoki said.



Source: Xerox PARC

When a non-active region of the Tap Tips interface is tapped with a stylus, outlines appear briefly around objects that will serve up information when tapped.

The interface turned out to be intuitive, allowing museum goers to spend more time looking at the artifacts. "We did a usability study which showed that people—some of them self-described technophobes—were able to stay in that visual mode where they were looking at things and... asking the computer about things without being distracted by the interface," said Aoki.

Although the information was available in both text and audio forms on the Tap Tips Guidebook, museum goers preferred the audio version, Aoki said.

The study yielded some further interesting results. With the museum goers free from computer gadgetry angst, they began to use the interface in ways the researchers hadn't expected.

Instead of simply listening to their own guidebooks, the museum goers began to share, incorporating the audio from the device into a conversation. "We found that people actually use the devices as a tool to interact with each other [and] share information. They treated the combination of a conversation with a friend, [the] things that they were hearing in their own guidebooks and the things that they were able to overhear from the friend's guidebooks [as a whole environment] that contributed to not only their understanding... but it changed the way in which they interacted with their friends," said Aoki.

In a larger context, the researchers gained a strong lesson from the study, said Aoki. "It makes you sit back and look at the way you've programmed these things—you don't necessarily just want to stick with the mechanisms that are provided by the tool kits that the vendors supply. It may take some creativity in order to come up with interaction mechanisms that are specific to the task and intuitive for the task... it may be okay to not use something exactly like what we use on a computer with a mouse and a screen," he said.

The Tap Tips type of interface is a logical step for Palm-type interfaces, said Jodi Forlizzi, assistant professor of human-computer interaction and design, at Carnegie Mellon University. It will likely prove more useful as handheld computers continue to percolate throughout society, she said. "The biggest challenge for PDAs and cell phones is that the devices need to be usable in mobile contexts—often while the user's attention is divided by another task [like] walking [or] driving."

The interface addresses two important trends in human-computer interaction research, said Forlizzi. "One, to try to understand the contexts in which mobile applications might work, so as to make their interfaces more usable; and two, understanding how dynamic graphical information could make information more salient to the user," she said.

The Xerox researchers are now working on a system that will facilitate information sharing in a public place while not impinging on the experience of others, according to Aoki. "The question is how can you share with just the people you're interested in sharing with. We think technology can help with this. We're still at an exploratory phase. The approaches we envision have very lightweight visual mechanisms... combined with audio presentation of the information," he said.

Aoki's research colleagues were Amy Hurst and Allison Woodruff of Xerox Palo Alto Research Center. A pair of technical papers on the research have been accepted for presentation at the Association for Computing Machinery (ACM) Conference on Human Factors in Computing Systems (CHI 2001) in Seattle, Washington, March 31-April 5, 2001. The research was funded by Xerox.

Timeline: Now
Funding: Corporate
TRN Categories: Human-Computer Interaction
Story Type: News
Related Elements: Technical papers, "Tap Tips: Lightweight Discovery of Touch Screen Targets" and "The Guidebook, the Friend and the Room: Visitor Experience in a Historic House," ACM Conference on Human Factors in Computing Systems (CHI 2001) in Seattle, Washington, March 31-April 5, 2001. Technical paper, "Improving Electronic Guidebook Interfaces Using a Task-Oriented Design Approach," Association for Computing Machinery (ACM) Conference on Designing Interactive Systems (DIS 2000), New York, August 17- 19, 2000.

TRN

Software Guides Museum-Goers

By Kimberly Patch, Technology Research News
June 12/19, 2002

Reading the written material that goes along with museum exhibits is always a little tricky. If you're the type who has to read every word, you're likely to see the same background information over and over again, and if you're the type who likes to dip in and out of the text, you'llrobably end up missing at least some of background material.

Researchers from Europe have built a system designed to tap the powers of hypertext, information databases, and natural language generation to allow people to go as deeply or as quickly as they wish through the written material in museum-type settings without repeating or missing much. "It occurred to me that... these problems can be addressed by using natural language generation technology," said Jon Oberlander, a reader in cognitive science at the University of Edinburgh.

The information can be displayed in several forms in physical places like museums and virtual spaces like the World Wide Web. "The same information server and generator can dynamically supply information to wireless handhelds in a real museum gallery, or drive synthetic speech over a mobile phone, or build Web pages on-the-fly to describe a virtual gallery," said Oberlander. The system is also designed to work with any language.

There are several inherent problems with museum labels, according to Oberlander. First, they are generally designed to be accessed in any order. This means they must each represent all the relevant information about their object, which can mean overly wordy and redundant descriptions. "Small differences between two objects may be submerged in a sea of similar details," he said. Using traditional labels, the only way to avoid massive redundancies is to force visitors to read the descriptions in a certain order "and that's not great for their sense of freedom," he said.

"Secondly, there's no guarantee that [visitors] will actually find what they need," he added. In contrast, a live curator can find out what museum-goers want, present options, and, if necessary, steer them to objects they were not aware of, said Oberlander.

The researchers' system addresses those problems by generating answers to visitors' questions on-the-fly. It keeps track of what a visitor has seen in order to tailor the descriptions appropriately.

Someone visiting via the Web would start from a page of icons showing a gallery of objects, and when the visitor clicked on a particular icon, a new page would be generated, with a larger image, a title, a description and a list of links to related objects. "At this point they can return to the main page and choose another object, where they can follow one of the suggested links, or they can ask for more information about the current object. Either way a new page is generated for the chosen object [and] the description of the page will take into account what other descriptions have been generated so far, tailoring both content and form," he said.

Under the hood is software that includes four key components: a content potential module, a text planner, a surface realizer and a module that chooses the best presentation for the generated description.

The content potential module keeps track of, and links together, facts extracted from museum databases and curator interviews. It also places different values on each fact, depending on how important the curator judges it to be and how interesting and familiar it is expected to be to the visitor. This familiarity value changes throughout the course of a visit.

When a visitor requests information, the text planner module selects a subset of facts from the content potential module. "It starts from the... selected object, and includes all the facts which are nearby and sufficiently interesting, important and unfamiliar," said Oberlander.

The module takes into consideration the number of facts available for the current type of user, and organizes the information into a coherent order that signals explicitly how the facts fit together, Oberlander said. "The text structure built up this way is still essentially independent of the language which is used to express the information," he said.

The surface realizer takes this abstract information and chooses the best way of expressing it using grammatical constructions, words and connectives. "This is also where the system takes into account the different ways we refer to objects when we mention them the first time, [than] on subsequent occasions," said Oberlander. For example, the first time you mention a designer, you might say 'a British designer named Jessie M. King', then later refer to her as 'Jessie M. King', 'King', or 'she'.

The final module decides whether to wrap the textual description in HTML with live links, send it as pure text, or put it through a speech synthesizer.

In theory, the software can work with any language. The researchers are currently working with English, Italian and Greek. "One of the key challenges in the current project has been to cleanly separate the parts of the system that are independent of English, Italian or Greek from the parts that have to rely on knowledge of the particular language," he said.

In some ways, English is the easy language, Oberlander added. "The sophistication of the system [had] to be considerably increased for languages with complex word-information rules like Greek. But once you've done Greek, Italian is relatively easy," he said. In the end, it shouldn't cost much to add a new language, he said.

As part of the project, the researchers and a partner, the Foundation of the Hellenic World in Athens, have constructed an immersive view of the ancient city of Miletus using the software.

The researchers are also looking to use the software to mine many types of existing textual information, including online catalogs. "It will work with almost any kind of online catalog and in customer relationship management," said Oberlander. The researchers are also planning on using the system for tutoring, he said.

The software combines work in several different areas in a very interesting way, said Paul Aoki, a research scientist at the Xerox Palo Alto Research Center. "They're able to [make] previous technologies really deployable," he said. "You can imagine that typical audio guide content like overviews, jokes and dramatic stories would be tough to generate on-the-fly, but something like [this] could be used to weave pre-recorded pieces together with dynamic factual content."

The overall approach of generating text from a database of descriptive elements could have many uses, Aoki said. "There are many different... scenarios where this kind of technology can be applied—walks through historic districts, botanic gardens, historic houses. Another example might be an audio restaurant guide that knows you care about parking and price... and gives you natural-sounding descriptions that are tailored to those preferences," he said.

Oberlander's research colleagues were Ion Androutsopoulos and Aggeliki Dimitromanolaki of the Greek National Center for Scientific Research in Greece, Vassiliki Kokkinhai of the Foundation of the Hellenic World in Greece, Jo Calder of the University of Edinburgh, and Elena Not of the Trentino Cultural Institute in Italy. They presented the research at the 29th Conference on Computer Applications and Quantitative Methods in Archeology held in Gotland, Sweden, April 25 to 29, 2001. The research was funded by the European Union.

Timeline: Now
Funding: Government
TRN Categories: Human-Computer Interaction; Databases and Information Retrieval
Story Type: News
Related Elements: Technical paper, "Generating Multilingual Personalize Descriptions of Museum Exhibits — The M-PIRO Project," presented 29th Conference on Computer pplications antitative Methods in Archeology in Gotland, Sweden, April 25-29, 2001.

Software Turns Reading Into Writing

By Kimberly Patch, Technology Research News
September 4/11, 2002

Humans communicate in many ways—by speaking, with our eyes, with gestures, and through touch. In comparison, the keyboard is fairly primitive, and its near-monopoly on computer text entry is a high hurdle for those who find it slow or impossible to push the requisite buttons—PDA users and the disabled alike.

Researchers from the University of Cambridge in England are getting ready to release an open source software program that promises to speed computer use for people who are unable or unwilling to use a keyboard.

The software, dubbed Dasher, lets users spell words by steering through a landscape of letters, said David J. C. McKay, a reader in natural philosophy in the physics department at the University of Cambridge.

The software can be used with a stylus, mouse, track pad, rollerball or eye tracker to enter text into handheld or desktop computers. People who have become expert at using Dasher with a mouse have entered text as fast as 34 words per minute, said McKay.

The eyetracker version, which works in conjunction with a camera that follows a person's gaze, allows a person to produce text using only eye movements, according to McKay. With an hour of practice on the eye-tracking system, a novice user can achieve 25 words per minute, he said.

The interface presents the user with letter choices that change as the user points to or looks at a letter. "Imagine sky diving onto a world painted with alternative letters, each its own field, and within each of those fields are smaller fields... painted with one letter from the alphabet," said McKay. "By steering through the big fields into smaller fields you choose a sequence of letters."

The sizes of the fields vary, with more probable possibilities in bigger fields. This makes the more likely letter choices easier to select, said McKay.

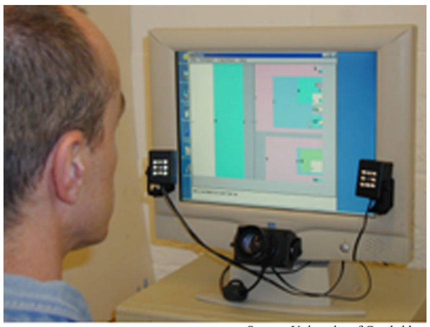
The field size and continuous selection methods make the system less laborious than other ways of selecting a letter at a time, he said. Existing systems that use eye tracking to select letters tend to require the user to briefly stare at a letter in order to choose it. This can be fairly tiring and is relatively slow.

Spelling with Dasher is more like steering through a landscape, said McKay. "The user has the feeling that whole syllables, whole words, even whole phrases are simply leaping towards him," he said.

Because the user is looking at the fields he wants, it is possible to dispense with a pointing device altogether, and use eye tracking only, he said. The system can "simply track the user's gaze to get the steering signal," he said.

Under the hood, the software uses a text compression method called arithmetic coding. Text can be compressed simply because it has lots of repeated letters. This redundancy can be described by a logical model that defines the probability of the first character in a document, the probability of the second character given the first, and so on. Instead of compressing, or telescoping existing words, Dasher uses these possibilities to generate new words.

The method can be thought of as a ruler that measures the probabilities of certain letters, rather than centimeters and inches, according to McKay. The top five percent of the ruler is reserved for "A", the next two percent for "B", and so on. The "A" area could be further subdivided into "AA", "AB", "AC" and so on, in proportion to the respective probabilities. Every possible string of characters is associated with a little fragment of ruler, said McKay.



Source: University of Cambridge

An eye-tracking system watches where the user is looking, allowing him to choose letters to build words. The more likely next letters appear in larger areas that are easy to spot, speeding the process.

To write with the eye-tracking version of Dasher, a person would look on the screen for the first syllable of the word, and it would zoom past, replaced by the next sets of possible syllables, said McKay. A little to the left of where the person is looking, the characters she has already chosen are queuing up, and to the right the user can see possible continuations arranged alphabetically. Most of the time the next syllable is easy to spot, he said. "No conscious control of the eyes is needed."

Spelling errors are also fairly rare with this type of software, he said. This is because if the user makes one, he typically notices right away because the model doesn't make a good prediction of the character he wants to write next. "You can correct errors by backing up—looking to the left of the screen instead of the right—then going forward again," McKay said.

The software also addresses a problem common to spelling programs that offer a list of complete words to choose among as a shortcut to continuing to spell a word. Having a word completion list means that after choosing each character a person must decide whether to check the candidate words or continue typing. It's a subtle point, but switching to a slightly different mental activity takes cognitive effort.

In contrast, the Dasher interface makes no distinction between word-completion and ordinary writing, but allows users to simultaneously see the last few characters they've chosen and the most probable options for the next few, said McKay.

The technical challenge the researchers faced in writing the software was balancing the trade-off between tapping the computer's processor power to refresh the moving image on the computer screen, and tapping it to compute additional predictions of the language model, said McKay.

The software is "a clever invention," said Robert Jacob, an associate professor of electrical engineering and computer science at Tufts University. It may take a fair amount of concentration to use, however, because of the constant on-screen changes, he said. "The ground shifts under you—every time you move, the picture changes," he said.

In the right situation, however, having to pay close attention to the screen is a small price to pay for not having to move very much to write words on the screen, he said. "It does solve the problem of optimizing a person's motions," he said.

The researchers are releasing the program as open source software that anyone can use and improve.

There are three distinct groups of potential users, said McKay. The program offers a way to enter text into handheld computers using an eyetracker or a miniature track pad. Disabled users can control Dasher on the desktop using a touchpad, joystick, mouse, head-mouse, roller-ball or eyetracker, he said.

Finally, the software may be especially useful to Japanese computer users who wish to write using the Hiragana alphabet, he said. Currently most computers in Japan have standard QWERTY keyboards, but using them to generate the 46 characters of the Hiragana alphabet is quite slow, said McKay. "Dasher offers a way to write in Japanese that bypasses the QWERTY keyboard," he said.

The software could also eventually be used in conjunction with both translation and speech recognition software to give users an alternative way to correct errors, McKay said.

The researchers are currently working on improving the eyetracker's automatic calibration. "At present, poor calibration of the eye tracking limits performance," McKay said. "We believe we can automatically tune the eyetracker on-the-fly using the information supplied by the user's steering corrections," he said.

The researchers are also working on improving the language model. "We can imagine... an extra 20 percent improvement in speed by improving the language model's compression by 20 percent," said McKay.

McKay's research colleague was David J. Ward. They published the research in the August 22, 2002 issue of the journal *Nature*. The research was funded by the Gatsby charitable foundation and by IBM Zürich Research Laboratory.

Timeline: 1 year
Funding: Corporate; Private

TRN Categories: Computer Science; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "Fast Hands-free Writing by Gaze Direction," *Nature*, August 22, 2002. The software is available for download at the researcher's site: www.inference.phy.cam.ac.uk/dasher

Muscles Tapped for Virtual Input

By Kimberly Patch, Technology Research News
February 7, 2001

It has long been a fantasy of computer junkies and more recently a dream of people who find typing painful to be able to communicate with computers more directly.

A group of NASA researchers is bringing the idea closer to reality by intercepting the electrical signals the brain and muscles use to communicate in order to provide phantom joystick and keypad control.

"We're looking at ... connecting human beings to machines at a neuroelectronic level [by] trying to understand the underlying nervous system signals," said Charles Jorgensen, who heads the neural engineering laboratory at the NASA-Ames Research Center.

The researchers are using electrodes to literally listen in on the body's communications system. "Dry electrodes... are sitting on the surface of the skin and they're picking up the electrical signals [that tell the muscles to contract] as they're being sent to muscles," said Jorgensen.

The signals are then sent to a computer, which interprets them using the hidden Markoff model algorithm commonly used for speech recognition, and neural net software, which is also good at recognizing patterns.

The researchers used a sleeve containing eight electrodes to allow a pilot to control a simulation of a 757 passenger jet, according to Jorgensen. "You can reach into the air and grab an imaginary joystick and move the wrist without any joystick actually in your hand and the neural signals are then interpreted and sent to a class four aircraft simulation," he said.

They've also used the system to enter data into an imaginary keypad. "We're understanding... slight movements of the fingers in order to enter data on a keypad without [an actual] keyboard—you can type in the air or on a picture of a keyboard," Jorgensen said.

One motivation for developing the system is to make it possible for astronauts to use computers while wearing bulky spacesuits and to better control things like robot arms, said Jorgensen. "There are a number of applications — everything from a wearable cockpit to exoskeletal manipulation to emergency communication modes such as in hazardous suits where the suit inflates."

The difficult part of tapping nerve signals is isolating and interpreting them correctly, said Jorgensen. "These are extremely weak signals in a high noise environment because we are measuring them on the surface of the skin," he said.

Neurons fire by building up chemical energy in the form of sodium ions crossing a membrane, then release it all at once, causing a spike of electrical activity. "Each individual neuron has a little pulse of energy. The electrode is picking up thousands of individual nerves that are firing" at once, said Jorgensen. When the middle finger moves from the five to the eight on a keypad, for instance, the other fingers move as well, creating that much more noise on the surface of the skin.

Source: NASA

A set of electrodes records nerve signals in the arm, allowing a pilot to control a jetliner simulation without actually touching the joystick.

"In the aggregate this winds up looking like a waveform," he said. This electromagnetic effect "propagates from cell to cell and [is] picked up by the sensors on the skin. We're just getting the average energy at different locations," Jorgensen said.

Electrical noise outside the body, like the electric fields produced by the flow of electricity in electrical hardware, also contribute to the electrical cacophony on the surface of the skin.

In addition to the noise, there are other problems in interpreting nerve signals through skin and over time. The electrodes are sensing the signals through a fat layer under the skin, and so the electrode position changes relative to the nerves as the hand moves, said Jorgensen. A cup of coffee can change things as well. "If you get caffeine in your system you are chemically a little different critter and your nervous system fires differently," he said.

The signals the electrodes find are amplified, filtered, and then interpreted using the hidden Markoff and neural net algorithms. "There are some distinctly different wrinkles in... the signal processing algorithms. You've got to get moving averages, generate probability density clusters, identify transition states that the information is going through, and you have to map [these pieces of information] to a pattern recognition scheme that lets you label them in one way or another," said Jorgensen.

Once the signals are sorted out, they're used to make changes in the position of the joystick or keys pressed on the keypad. The computing can be done using a 230 megahertz

Pentium III chip, making it possible to use the system with a wearable computer, said Jorgensen.

It's an interesting approach that is part of the general move in computer communications towards a wider range of sensory inputs and outputs, said Terry Winograd, a professor of computer science at Stanford University. "It might... be relevant to mobile computing, where, for example, you could have a keyless keyboard, just detecting the muscle motions of the fingers," he said. It may also prove useful for helping people with disabilities who can't use ordinary devices, Winograd said.

The NASA researchers are working on expanding the keypad control to the entire keyboard. Also in the works are plans to make the system trainable. "We want to look at adaptive algorithms to permit customization for different users like you have for speech recognition," said Jorgensen.

The researchers are also exploring a project that will attempt to combine electromyographic signals from the muscles of the neck with electroencephalographic signals from the brain to interpret speech. "One of the things that we're exploring is whether or not we might be able to do silent speech recognition—electrodes would pick up subvocalization behavior [in the neck] — and combine that with [brain wave] information and that may be enough for us to tease out what's being spoken," Jorgensen said.

The type of device control used in the joystick and keypad should be available for practical applications within five years, said Jorgensen.

Jorgensen's research colleagues were Kevin Wheeler of NASA and Slawomir Stepniewski of Recom Technology Corp. They presented research pertaining to the aircraft simulation model at the World Automation Congress Third International Symposium on Intelligent Automation and Control in Hawaii, June 11-16, 2000. The research was funded by NASA.

Timeline: < 5 years
Funding: Government
TRN Categories: Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "Bioelectric Control of a 757 Class High Fidelity Aircraft Simulation," World Automation Conference, June 11-16, 2000, Hawaii.

Hearing between the Lines

By Kimberly Patch, Technology Research News
July 19, 2000

When humans talk, we exchange a lot of audio information along with the words. Computers, however, don't hear between the lines, which is one reason speech recognition applications can seem so frustratingly stupid. Essentially, today's computers are socially inept, blind to the meanings of subtle pauses or even drastic changes in tone.

The technical reason for this is the Hidden Markov Model (HMM) most speech recognition programs rely on only looks at tiny, 10 millisecond slices of speech. The model works well for picking out words, but misses contextual cues that span words, phrases or sentences.

"When you pause at the end of the sentence or you lengthen or you drop your pitch, that [spans] a region that's at least 10 times larger than the HMM can capture and sometimes 100 times larger," said Elizabeth Shriberg, a senior research psycholinguist at SRI International.

Hoping to remedy the situation, Shriberg and other researchers have shown in a pair of experiments that computers can use speech attributes like prosody—information gleaned from the timing and melody of speech — to better understand human speech.

In one experiment, prosody significantly improved a computer's accuracy in adding punctuation and paragraphs to databases of speech from news broadcasts and phone conversations. Prosody proved even more helpful in sorting the broadcast feed into topics. (See chart)

Prosody includes the duration, pitch and energy of speech. Duration, or the way people stretch or speed certain parts of speech, is most important, said Andreas Stolcke, a senior research engineer at SRI International. "People use the duration of speech sounds in certain ways to emphasize things," he said.

The researchers found that pauses and pitch were most useful in segmenting news speech, while pauses, duration of syllables and word based cues proved significant in the more difficult task of segmenting natural conversation.

Prosodic information is slowly being recognized as an important source of information in speech understanding, said Julia Hirschberg, Technology Leader in the Human-Computer Interface Research Department at AT&T Labs. "The SRI work applies prosodic information to a very important task, topic segmentation, with considerable success. [It's] the first that I know of which improves topic segmentation performance," she said.

In another experiment, researchers used word choice and order as well as prosodic cues to improve the task of automatically categorizing telephone conversations into 42 types of phrases like statements, opinions, agreement, hedging, repeated phrases, apologies, and phrases that signal non-understanding.

Prosody's ability to mark emotional levels of speech may eventually help in certain types of searches, like news footage of politicians having an argument. A similar, real-time, application could be call center operators wanting to know "who the angry customers are right away because you don't want them to have to [continue listening] to a computer," said Shriberg. Prosody also allows computers to gauge attention levels, which may allow educational applications to automatically adjust the difficulty of a task. Prosodic information, because it differs among languages, may also prove useful in discerning what language is being spoken.

The researchers are also looking at using prosody to make speech recognition more accurate — "the holy grail right now," said Stolcke. "The general idea is simply to have a more comprehensive model of everything that can vary within speech. [You] can get significantly better speech recognition if you know the type of utterance," he said.

Better recognition based on prosody is also likely to create a feedback loop that will make talking to computers more natural, said Shriberg. "If the machine is using [pitch and emphasis, people] will put that in their speech because it's getting a response from the machine. They'll adapt to what the machine is able to do—that's a well-known principle."

Real world applications of prosody are at least two years away, said Shriberg.

Shriberg and Stolcke were joined in the prosody topic segmentation research by Dilek Hakkani-Tür and Gökhan Tür of Bilkent University in Ankara, Turkey. They were joined in the automatic tagging of conversational speech research by Noah Coccaro and Dan Jurafsky of the University of Colorado Boulder, Rebecca Bates of the University of Washington, Paul Taylor of the University of Edinburgh, Carol Van Ess-Dykema of the U.S. Department of Defense, Klaus Ries of Carnegie-Mellon University and the University of Karlsruhe in Germany, Rachel Martin of Johns Hopkins University and Marie Meteer of BBN Technologies.

The researchers' work on prosody for topic segmentation was funded by the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA). The work on automatic tagging of conversational speech was funded by the Department of Defense (DOD).

Timeline: > 2 years; > 5 years
Funding: Government
TRN Categories: Databases and Information Retrieval; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper "Prosody-Based Automatic Segmentation of Speech into Sentences and Topics" posted in the Computing Research Repository; Technical paper "Dialog Act Modeling for Automatic Tagging and Recognition of Conversational Speech," posted in the Computing Research Repository.

Correction Choices Key for Speech Software

By Kimberly Patch, Technology Research News
September 5, 2001

One broad trend in getting humans to interact more easily with computers is multimodal input—essentially giving us the breadth of choices we are used to rather than restricting communications to tapping on keys and positioning a cursor.

Researchers from Carnegie Mellon University have found that giving users who are not experts several ways to correct speech recognition errors makes them considerably more successful in using speech recognition software, substantially lowering the method's steep learning curve.

This is true even though some ways of correcting are clearly more efficient than others.

"The general goal of the research was to see how much better one could do using multimodal correction," said Brad Myers, a senior research scientist at Carnegie Mellon University.

The results also point out the importance of being able to quickly and easily correct the words the computer mishears, a capability that is generally underrated, said Myers. "Error correction really needs to be one of the fundamental things you take into account when you're designing a speech system, because they're never going to work perfectly and the ways that are available for people to correct the errors have an enormous impact on their usability," he said.

The researchers tested the abilities of three different types of users—novice, average and skilled—to correct speech-generated text in three different ways: using only the keyboard, using the keyboard and speech as in conventional dictation systems, and multimodal correction, which allowed users to choose among keyboard, speech and handwriting correction methods.

In general, using speech to correct speech recognition errors allowed all three types of users to create text faster than using only typing to correct speech errors, according to the research. There was also a considerable learning curve, with experts doing much better than average users, and average users doing much better than novices.

Also, all three types of users dictated somewhat slower than is possible with commercial speech dictation systems because the researchers' system does not adapt to a user's voice. This decreased dictation accuracy from 90 percent or more to about 75 percent, which also increased the need to correct.

Correcting by typing only, experts produced about 40 words per minute, average users 30, and novices about 11. Using typing or speech to correct errors, experts produced about 55 words per minute, average users 40 and novices 22.

There are three basic ways to correct speech using speech: saying a word again, choosing from a list of possibilities, and spelling a word. All of these methods are used by commercial speech recognition systems.

The researchers found that the most instinctive way for humans to correct mistakes using speech was the computer's worst. "The most obvious way to correct when the system mishears is to say it again. But it turns out that most speech systems... actually do much worse when you try that," said Myers.

There are several reasons for this, he said. First, a misheard word is likely to be a difficult one to understand in the first place. Also, speech recognition systems are designed to understand a word in context, but users are more likely to say a single word at a time when correcting a mistake. "When you say a word in isolation it actually sounds totally different," said Myers. In addition, when forced to say something over again, "people tend to hyper-articulate—it works when you try to make someone else understand [but it] sounds different," to a computer, said Myers.

The problem with allowing users to correct a word by choosing from a list of the computer's top 10 or so possibilities is that the correct word or phrase is often not listed, and when this happens, it slows correction down considerably, said Myers.

There's also a catch to using speech to spell a word in order to correct it. "The problem with that is [the system] doesn't do a very good job of recognizing the switch to spelling," said Myers. So, as in today's commercial speech recognition systems, the researchers had to provide a command so the users could tell the system that they were spelling, which slowed the correction down. However, once users learned to make the switch, spelling was the best of the speech modes of correcting, Myers said.

The researchers found that the experts tended to recognize that spelling words was the most efficient way to correct, and so did so consistently. Beginners, however, tended to use the least effective and most frustrating speech method of saying words over again.

One hypothesis going into the study was that people would eventually try to pick the technology that worked best, said Myers. "That was only partially supported" by the research, said Myers, because novices kept trying to correct an error by repeating the word even after many unsuccessful tries. This is probably because it works well for communicating with other people and so is a hard habit to break when talking to a machine, said Myers.

This loop became even less successful with time because "as you get more emotional your enunciation changes," said Myers.

Giving users the ability to correct using handwriting as well as speech increased the correction speed of novices and average users considerably. "Handwriting with a pen-based interface in general worked pretty well," Myers said.

Using the multimodal system, novices' dictation speed nearly doubled to about 40 words a minute; average users' speed increased slightly to 44 words a minute.

The multimodal system didn't help experts who were already proficient in the spelling correction method. In fact, it slowed them down a little, from 55 to 48 words a minute.

The main message of the study is that the error rate and techniques for correcting errors have to be taken into account in order to improve usability of speech systems, Myers said. "Allowing multiple choices for how to correct errors really makes a big difference in the success of the system," he said.

Speech systems should take into account the ingrained tendency to simply say a word again when it is not heard correctly, said Myers. "Our recommendation would be that speech systems... have different language models... in correction mode, [that account for] hyper articulating and the same words in isolation as compared to context—that might make the correction more successful," said Myers.

This type of research is not only very relevant and applicable to existing interfaces, but will prove more important in future interfaces, said Matthew Turk, an associate professor of computer science at the University of California at Santa Barbara. It's one piece of a larger movement towards interfaces that accommodate people in all kinds of situations, he said.

As electronics grow even more widespread, these types of interfaces will ensure that "we're not going to have to spend hours and hours learning detailed computer systems and different systems for different purposes," said Turk. "We can just use whatever were comfortable with. If someone does have a particular skill like... good voice input, they can take advantage of it, but if they don't, they have these other alternatives."

There are no technical barriers to implementing multimodal speech correction in today's products, said Myers. "It would require some engineering work to tune the parameters, but there is reasonably good speech, handwriting and [even] gesture recognition already on the market. [They're] just not integrated into the same system," he said.

The Carnegie Mellon researchers are currently working to more fully meld different ways of interacting with a computer—for instance using a combination of speech and gesture to evoke a command—to make communicating with computers more natural for humans, said Myers. The trick is figuring out how to get the different input mechanisms, or recognizers, to cooperate at a more basic level than they usually do, he said.

Myers' research colleagues were Bernhard Suhm, a former Carnegie Mellon University graduate student who is now at BBN Technologies, and Alex Waibel of Carnegie-Mellon University and Karlsruhe University in Germany. The researchers published their work in the March, 2001 issue of the journal *ACM Transactions on Computer-Human*

Interaction. The research was funded by The Defense Advanced Research Projects Agency (DARPA).

Timeline: Now
Funding: Government
TRN Categories: Human-Computer Interaction
Story Type: News
Related Elements: Technical paper "Multimodal Error Correction for Speech User Interfaces", *ACM Transactions on Computer-Human Interaction*, March 2001.

TRN

Two-Step Queries Bridge Search and Speech

By Kimberly Patch, Technology Research News
July 24/31, 2002

After 30-odd years of computer speech recognition development, researchers are still looking for ways to make it easier for a computer to sift individual words from the constant stream of syllables that is spoken language.

What is most difficult for speech software is recognizing the borders of a word that's not in its dictionary. Researchers from the Japanese University of Library and Information Science and the Japanese National Institute of Advanced Industrial Science and Technology have made this a little easier.

The researchers have found a way to help speech recognition programs used to retrieve information from data collections like the Web identify out-of-vocabulary sequences of syllables. In a sense, the researchers have given computers a faster way to sound out words they don't already know.

State-of-the-art information retrieval systems allow users to input any number of keywords into a vocabulary. "It is often the case that a couple million terms are indexed for a single information retrieval system," said Atsushi Fujii a research assistant at the University of Library and Information Science in Japan.

State-of-the-art speech recognition systems have to limit vocabulary size to a few tens of thousands of words in order to match each syllable sequence to a word in real-time.

Because of the limited speech recognition vocabulary sizes, however, when speech is used to query information retrieval systems, some of the words may not be in the speech recognition vocabulary.

The trick to finding these words is knowing where to look. When someone uses speech recognition as an interface to search a collection of data, he naturally utters words related to the unrecognized query term, said Fujii.

To take advantage of this, the system carries out the query using the words the computer does recognize, then looks in those documents for words that are phonetically identical or similar to the unrecognized syllable sequences. The system then queries the documents again using the new-found words. This two-step process makes it possible for the computer to match an unrecognized syllable sequence to a real word relatively quickly, according to Fujii.

The researchers tested their method by dictating queries to archives of newspaper articles. The method improved the information retrieval system's accuracy and did not increase the search time, according to Fujii.

The researchers also used their data retrieval method to beef up a speech recognition system's vocabulary with appropriate new words. "We used a target collection to recover speech recognition errors so as to improve the quality of [both] speech recognition and information retrieval," Fujii said.

The method is a way to improve speech-driven information retrieval systems, which could lead to interactive dialogue and question-answering systems that allow users to control computers by speech, according to Fujii. These include car navigation systems, and Web search using telephones and mobile computers, he said.

The researchers have come up with a "clever trick" for turning sequences of syllables that are not in a speech recognizer's vocabulary into words, said Brian Roark, a senior technical staff member at AT&T Research. "This takes a step toward solving the problem of turning... syllable sequences into [correctly spelled] words," he said.

The method is potentially useful for speech recognition in general, Roark said. "If you can somehow leverage a particular task to give an indication of likely [out-of-vocabulary] words in a particular context, it might be possible to exploit this," he said.

But because large vocabulary recognition programs don't come across a lot of out-of-vocabulary sequences the total possible gain in recognition from this method would probably be fairly small, Roark added.

The researchers' next step is to do larger-scale experiments using different types of document collections, such as technical papers and Web pages, said Fujii.

The researchers' current experiments use Japanese speech that is dictated directly to the computer, said Fujii. Ultimately, the researchers are aiming to be able to process spontaneous speech in different languages, he said.

Practical applications using dictated speech are technically possible within two years, said Fujii. Applications that can handle spontaneous speech will take more than three years, he added.

Fujii's research colleagues were Katunobu Itou of the National Institute of Advanced Industrial Science and Technology in Japan, and Tetsuya Ishikawa of the University of Library and Information Science. The research was funded by the University of Library and Information Science and the Japan Science and Technology Corporation (JST).

Timeline: < 2 years, > 3 years
Funding: University, Corporate
TRN Categories: Databases and Information Retrieval; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "A Method for Open-Vocabulary Speech-Driven Text Retrieval," posted in the arXiv physics archive at xxx.lanl.gov/abs/cs.CL/0206014.

TRN

Programming Tool Makes Bugs Sing

By Kimberly Patch, Technology Research News
August 7/14, 2002

Over the past few centuries we've found ways to cram more data into graphics like maps, and dynamic elements like slider bars have allowed computers to display even more data per inch. Sound, however, is largely absent from these tools, despite music's long history as a universal language.

Researchers from the University of Northumbria in England are tapping the auditory sense by allowing programmers to listen, rather than simply look, for software bugs.

The work may eventually give sound a bigger role in computing. It is also a step toward making programming and other computing activities more accessible to people who are visually impaired.

The researchers first tested the ability of the average non-musician to distinguish differences like pitch using sounds similar to those musical instruments make. The results were good. "Most people could discriminate between separate pitches, and describe melodic contours quite well," said Paul Vickers, a principal lecturer at Northumbria University.

They then set up software that mapped pitch and melodic contour information to structural elements in the programming language Pascal. "[We] aimed to see if information about the structure of Pascal programs could be communicated using such musical phrases," said Vickers.

The researchers used similar musical phrases, or motifs, to represent similar programming constructs.

The constructs of the Pascal programming language can be categorized hierarchically into two classes—selections and iterations—which, in turn, have sub-classes, Vickers said. For instance, in the iteration class, the language has a pair of similar bounded loops,: "FOR... TO" and "FOR... DOWNTO". It also has a pair of unbounded loops: "REPEAT" and "WHILE".

The researchers used a common theme for each class, and wrote musical motifs that were variations on that theme to make the similar REPEAT and WHILE loops sound distinct, but more similar to each other than to the bounded loops. At the same time, all four of the phrases representing loops sounded similar enough to each other that they could be distinguished as being in the iteration class rather than the selection class.

Once again the results were good. "Programmers [can] derive useful program information from musical representations," said Vickers.

Once they had established the basics, the researchers worked out a melodic system for debugging that follows several basic principles.

In writing the motifs, the researchers used the common diatonic, seven-note scale, which is easy to memorize and recognize, and they made use of several sound variables that are easy to follow.

They used meter and rhythm, which previous music cognition research showed are easily retained, to delineate different parts of the program. A song remains recognizable, for example, even when the underlying cords are changed significantly, if its basic rhythm stays intact, said Vickers. The meter, or time signature of a musical phrase indicates a certain number of beats per measure.

They added percussion to changes in melody. "Percussive devices can enhance the music and provide extra clues... to help users recognize significant events such as a change in condition of a Boolean evaluation," said Vickers.

And they used continuous tones, or drones similar to those used by bagpipes, to indicate continuous states like loops where many nested operations may take place. "The use of a continuous tone can indicate that the program is inside the loop," said Vickers.

Once the researchers worked out the musical rules, the scheme was easy to implement, said Vickers. "We used simple Musical Instrument Digital Interface (MIDI) devices," he said.

When the researchers tested the auralization tool on 22 undergraduate computer science students, they found that the subjects could correctly locate bugs in a program more often when sound was added. What surprised the researchers was the response times for subjects who used the sound and those who did not were similar, said Vickers. "Because the musical auralizations took time to listen to—up to two minutes in some cases—we might have expected... an increase in the time spent locating the bugs," he said.

The researchers work is a good exploration of the ways people can absorb information presented musically, said Mike O'Donnell, a professor of computer science at the University of Chicago. "We have so little information on that topic that every experiment is very helpful," he sai
http://xxx.lanl.gov/abs/cs.CL/0206014
d. "I'm not aware of anyone else using musical phrases to render the execution of a computer program."

There are at least two challenges to using sound to represent information, said O'Donnell.

First, time is perceived in at least three different ways and these three ways have radically different structures: the speed of vibration, or frequency, which we hear as a certain pitch; sudden changes in the frequency, which we perceive as part of the timbre that makes up a particular type of sound; and sequences of sounds over time, which we perceive as, for instance, a sequence of notes in a musical melody.

The problem is that these three different ways of perceiving overlap. “We can’t separate them cleanly,” said O’Donnell. In contrast, a moving picture can be separated into a rapid sequence of static frames.

Second, because aural information is based on a sequence of sounds over time, its perception must be explicitly programmed. In contrast, we can browse a static picture by shifting the focus of our eyes, said O’Donnell. “Eye motion generates a lot of valuable interaction between a human user and a graphical display for free,” he said.

It would be interesting to condense the musical material so that a more experienced programmer can listen for patterns in much larger program runs, said O’Donnell. Looking at the concept backwards may also be enlightening, he said. Writing programs that make interesting musical segments “may give [programmers] an intuition for the impact of different programming constructs,” he said.

The researchers are currently looking at how musical representations of programming languages might be used in teaching computer programming, said Vickers.

Another line of research is to see how auditory signals can be used to build accessible user interfaces, he said. They are “investigating using musical messages to assist elderly users with monitoring and controlling household devices,” he said.

Sound has several potential benefits for human-computer interaction that could eventually be exploited, said Vickers. “Today’s graphical user interfaces are increasingly complex, leading to cluttered screens and a general lack of screen real estate. If some of the communication burden can be transferred onto the audio channel, then perhaps interfaces can be made more efficient, more effective, less cluttered [and] easier to use,” he said.

Sound could also help visually impaired users in applications like audio-enhanced Web browsers, he said.

Sound also offers a new paradigm for exploring data, he said. “Its ability to transmit multiple streams in parallel—consider the different instrumental parts of the symphony — and its ability to transmit in time-based rather than spatial domains offer us new ways of interacting with data,” he said.

Vickers’ research colleague was James L. Alty of Loughborough University in England. They are publishing the research in an upcoming issue of the journal *Interacting with Computers*. The research was initially funded by Liverpool John Moores University.

Timeline: 4-5 years
Funding: University

TRN Categories: Computer Science; Human-Computer Interaction;
Multimedia
Story Type: News
Related Elements: Technical papers, “Using Music to Communicate Computing Information,” slated for publication in *Interacting with Computers*; “Musical Program Auralisation: a Structured Approach to Motif Design,” Interacting with Computers, slated for publication in *Interacting with Computers*; “When Bugs Sing,” slated for publication in *Interacting with Computers*.

TRN

PCs Augment Reality

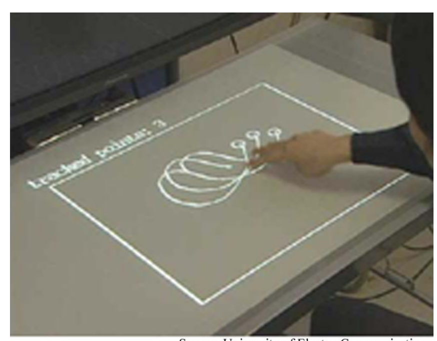
By Eric Smalley, Technology Research News
June 26/July 3, 2002

If you’re reading this story online, chances are you are looking at a desktop monitor and your hand is resting on a mouse. For years, researchers around the world have been working to eliminate the monitor and mouse.

Using digital cameras, projectors, and software that tracks hands and fingers, several research teams have turned desktops into displays and fingers into pointing devices. These augmented reality systems blend real objects and digital information, turning a Web address on a printed page, for instance, into a clickable link.

But the systems generally use expensive, high-powered graphics workstations to handle the difficult task of recognizing and tracking fingers in real-time, which tends to limit augmented reality to venues like research laboratories and art exhibits.

A team of researchers in Japan has brought augmented reality to a standard PC by finding a way to track users’ hands and fingertips that uses less computer power.



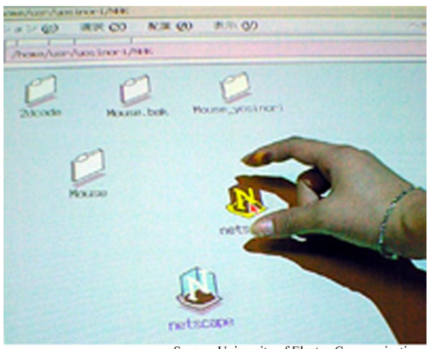
Source: University of Electro-Communications

This computer interface tracks fingertips in real-time...

The researchers added an infrared camera to make it easier for their system, dubbed EnhancedDesk, to distinguish fingers amid the clutter of a desktop. The infrared camera sees heat, a hallmark of human skin. Display images projected onto a desk can sometimes overlap the users’ hands, making even state-of-the-art computer vision techniques fail to detect hands and fingers in real-time, said Hideki Koike, an associate professor of information systems at the University of Electro-Communications in Japan.

They also gave the computer a little common sense. Ordinarily, if you instruct a computer to watch for fingers it will scan the entire arm looking for the telltale shape of a finger. The researchers' software knows to look for fingers only at the end of an arm, and to recognize that the semicircles at the ends of fingers are fingertips, which lightens the workload for the computer.

Armed with the knowledge of where a person's hands and fingers are, the system uses a standard camera that pans and tilts to follow the tip of the person's index finger as it moves around the desktop.



Source: University of Electro-Communications

... which allows a person to drag an icon with her fingers.

Users can put objects in the camera's field of view and tell the system what they are and what to do when someone points at them. When the system detects that a finger is pointing to one of these registered objects, it can, for instance, display related digital information that the person can then control with hand gestures, said Koike.

The idea is to allow people to read printed material and access accompanying digital information without having to use a keyboard or mouse to launch a program and find a Web page or multimedia file, said Koike. "In this environment, users [will] not need to be aware of the existence of computers," he said. "The system [will] automatically recognize what the users are doing by recognizing their gestures and objects on the desk."

The researchers have developed an educational program that works with the system. Interactive Textbook tracks the position and orientation of a textbook, recognizes certain pages and projects text, graphics, movies or Web pages at the appropriate angle and place on the desktop.

When students reading a physics textbook reach a page describing an experiment that shows the effect of a weight on a spring, for instance, the Interactive Textbook projects a simulation of the experiment to the right of the textbook. The students use their hands to control the simulation by moving or exchanging the weight and observing the effects on the simulated spring. "The students do not need to move their focus to using computers," said Koike.

EnhancedDesk is state-of-the-art, said Simon Baker, a research scientist at Carnegie Mellon University. "The complete system is very impressive. The [textbook] application is great," he said.

Augmented reality systems have a lot of potential, Baker said. "It is easy to imagine such tools being used in the very near future. [They] add a fun component to education, making the process far more enjoyable for kids," he said.

The one drawback to the system is the relatively high cost of infrared and pan-tilt cameras, said Baker. "It may... be possible to realize the basic idea using less expensive hardware," he said.

The researchers are adding a method for tracking a person's gaze, so the system can detect what a person is looking at, said Koike. The researchers' are also working on extending the system to whole rooms where it can track people's positions and project images on tables, walls and ceilings, he said.

EnhancedDesk could be ready for use in practical applications in a few years, said Koike.

Koike's research colleagues were Yoichi Sato of the University of Tokyo and Yoshinori Kobayashi of the University of Electro-Communications in Japan. They published the research in the current issue of the journal *ACM Transactions on Computer-Human Interaction*. The research was funded by the Ministry of Education, Culture Sports, Science and Technology in Japan.

Timeline: 2-3 years
Funding: Government
TRN Categories: Human-Computer Interaction; Computer Vision and
Image Processing
Story Type: News
Related Elements: Technical paper, "Integrating Paper and Digital Information on EnhancedDesk: A Method for Real-time Finger Tracking on an Augmented Desk System," *ACM Transactions on Computer-Human Interaction*, December, 2001.

TRN

Interface Gets the Point

Kimberly Patch, Technology Research News
January 1/8, 2003

Tone of voice can mean a lot. Your colleague can be giving you a complement or an insult depending on how she inflects the phrase "great work." Gestures can be just as expressive.

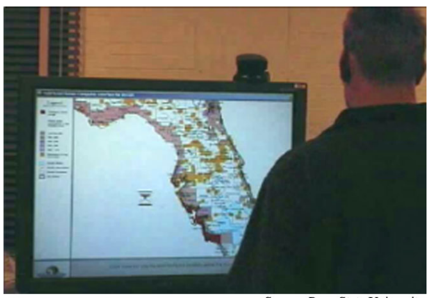
Communicating with computers is much more basic. Try to insult an uncooperative speech recognition system by telling it where to go, and, assuming your diction is clear, it will simply show the words on-screen without gleaning anything about your dark mood. Adding an appropriate gesture would make things very clear to even a tone-deaf human, but computers are generally gesture-blind as well.

Researchers from Pennsylvania State University and Advanced Interface Technologies are trying to change that.

They are working to untangle the relationships between prosody—the loudness, pitch, and timing of speech—and gestures in an attempt to improve the way computers recognize human gestures.

The research could eventually be applied to many different situations where humans try to get information across to computers, including computer games, surgical applications, crisis management software, and security systems, according to Rajeev Sharma, an associate professor of computer science and engineering at Penn State University and president of Advanced Interface Technologies, Inc.

Although it's child's play for humans, getting a computer to recognize gestures is difficult, said Sharma. Gestures "do not exhibit one-to-one mapping of form to meaning," he said. "The same gesture... can exhibit different meanings when associated with a different spoken context; at the same time, a number of gesture forms can be used to express the same meaning."



Source: Penn State University

This speech and gesture computer interface is designed for emergency management planning.

In previous work, the researchers analyzed hours of tape of meteorologists giving weather forecasts in order to link prosody to gestures.

The researchers increased their understanding of the phenomenon by plugging speech pitch and hand velocity into the Hidden Markov Model, which breaks information into very small pieces and makes predictions about a given piece of information based on what comes before and after it. The model is commonly used to predict words in commercial speech recognition systems.

The researchers used the system to help detect speech segments that commonly occur along with a particular class of gesture. "We [combined] visual and speech signals for continuous gesture recognition," said Sharma. "The basic idea... is to detect emphasized parts of speech and align them with the velocity of the moving hand."

For instance, a pointing gesture commonly precedes these emphasized segments of speech, a contour-type gesture is more likely to occur at the same time as an emphasized speech segment, and auxiliary gestures, which include preparation and retraction movements, tend not to include emphasized speech segments at all, according to Sharma.

The researchers are using the method in a geographical information system prototype that uses a large screen display, microphones attached to the ceiling and cameras that track users gestures.

The state-of-the-art in continuous gesture recognition is still far from meeting the naturalness criteria of a true multimodal human-computer interface, said Sharma. Computers have achieved accuracies of up to 95 percent in interpreting isolated gestures, but recognizing significant gestures from a full range of movements is much harder, he said.

Taking into consideration prosody when trying to interpret gestures, however, increased the accuracy of gesture recognition from about 72 percent to about 84 percent, Sharma said.

One of the challenges of putting together the system was to define when the visual and audio signals corresponded, said Sharma. "Although speech and gesture... complement each other, the production of gesture and speech involve different psychological and neural systems," he said.

Further complicating things, speech contains both phonological information, which are the basic sounds that make up words, and intonational characteristics, which include some words louder than others and raising the pitch at the end of a question. The system had to accurately pick up changes in intonation amidst the phonological variation in the speech signal, Sharma said.

Modeling and understanding prosody in systems that combine speech and gesture is important in the long run to help transition from a low-level, or syntax-based, to a high-level, or semantics-based understanding of communication, said Matthew Turk, an associate professor of computer science at the University of California at Santa Barbara.

The field has applications in "just about every human-computer interaction scenario, and in many computer-mediated human-to-human communication scenarios [like] remote meetings," Turk said.

The researchers are currently working on incorporating the prosody-based framework into a system to manipulate large displays. The researchers' next step is to run a series of laboratory environment studies to investigate how it works with real people, according to Sharma.

The researchers are ultimately aiming for an environment where a user can interact with the gestures he is accustomed to in everyday life rather than artificially-designed gestural signs, said Sharma.

The system could eventually enable more natural human-computer interfaces in applications like crisis management, surgery and video games, Sharma said.

Another possibility is using the method in reverse for biometric authentication, said Sharma. "This research [could] enable a novel way to identify a person from [a] video sequence... since a multimodal dynamic signal would be very hard to fake," he said.

Understanding how humans and computers can interact using several different types of communication will become increasingly important "as we deal with the need to interact with computing devices... embedded in our environment," said Sharma.

The first products that incorporate the prosody-based system could be ready within two years, said Sharma.

Sharma's research colleagues were Sanshzar Kettebekov and Muhammad Yeasin. The research was funded by the National Science Foundation (NSF) and Advanced Interface Technologies, Inc.

Timeline: > 2 years
Funding: Corporate, Government
TRN Categories: Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "Prosody Based Co-Analysis for Continuous Recognition of Co-Verbal Gestures," posted at the computing research repository at arXiv.org/archive/cs/intro.html.

TRN

Interface Lets You Point and Speak

By Kimberly Patch, Technology Research News
July 25, 2001

One of the reasons speech recognition software remains inferior to human speech recognition is computers can't read hands.

Humans convey a surprising amount of information through the gestural cues that accompany speech. We point things out, convey concepts like 'big' or 'small', get across metaphorical ideas, and provide a sort of beat that directs conversational flow.

No matter how often or how vigorously you shake your fist at at your computer screen, however, it won't help the computer tune in to your mood.

Researchers from Pennsylvania State University are working on a human-computer interface that goes a step toward allowing a computer to glean contextual information from our hands. The software allows a computer to see where a human is pointing and uses that information to interpret the mixed speech and gestural directions that are a familiar part of human-to-human communications.

These pointing, or deictic gestures are commonly mixed with speech when talking about things like directions, for example, saying "from here to here," while pointing at a map.

The researchers used Weather Channel video to glean a database of deictic gestures, which include directly pointing to something, circling an area, or tracing a contour. "Looking at the weather map we were able to classify pieces of gestures, then say which pieces we can interpret, and what kind of gestures would be useful. We came up with algorithms [that] extract those gestures from just the video," said researcher Sanshzar Kettebekov, a Pennsylvania State University computer science and engineering graduate student.

The researchers used this database to create a pair of applications designed for large screens that allow the computer to interpret what people mean when they use a mix of speech and pointing gestures.

One application, dubbed IMAP, is a campus map that responds to pointing and spoken queries. "It brings the computer into the loop with the human," said Kettebekov. For example, if a person asks the map for a good restaurant in an area she is circling with her hand, the computer will reply based on the spoken request for a restaurant and the gestural request for a location, according to Kettebekov.

The second application is a battlefield planning or city crisis management simulation that allows a person standing in front of a large screen to direct vehicles around a battlefield or city. "A person has limited resources [and there are] alarms going off all over the city. The person is using... a 50-inch display... to direct the resources to where the alarm is going [off]," said Kettebekov.

Even though it seems easy to us, giving a computer the ability to sense and make sense of gestures in a verbal context is a complicated problem that involves several steps, according to Kettebekov. The computer must be able to track the user's hands, recognize meaningful gestures, and interpret those gestures.

The first problem is tracking. "We have a vision algorithm that tracks a person and tries to follow a person's hand," Kettebekov said. The second stage is picking out the pointing gestures. "You're trying to delimit gestures from a continuous stream of frames where the hands are just moving—saying 'from here to here was this gesture'," he said. "The third stage is interpretation when you really associate [the gesture you have isolated] with parts of speech and try to extract meaning," he said.

Multimodal human computer interaction is an active research topic with a long history, said Jie Yang, a research scientist at Carnegie Mellon University. "Coordination of speech and gestures is an old but still open problem," he said, noting that there was a paper published 20 years ago on a computer system that integrated speech and gesture, and there have been many studies on the advantages of using speech and gesture. "Yet, we cannot naturally interact with a computer using speech and gesture without constraints today."

When all the difficult computer problems have been worked out, however, systems that recognize speech and gesture will allow a person to "efficiently manipulate multimedia information regardless of whether the person is communicating with a computer or with another human," he said.

The Penn State researchers are working on improving their gesture recognition algorithms by adding an understanding of the prosodic information that lends speech its subtle shades of meaning, said Kettebekov.

"We're working on using prosodic information in speech: tone of voice, stresses, pauses... to improve gesture recognition and interpretation," he said.

The toughest of the three gesture problems is improving gesture recognition, said Kettebekov. Currently the system identifies keywords and tries to correlate them with gestures. Adding prosodic information would help the system to both recognize gestures and interpret them, he said.

For example, when a TV meteorologist wants to emphasize a keyword, he raises the tone of his voice, said Kettebekov. "If I want you to pay attention I not only point, but my voice would change so that I would attract more attention to that concrete point," he said. "You can extract those most prominent parts of speech, and those parts of speech nicely relate with the gestures—in this case it was pointing," he said.

The researchers may eventually turn their sights to iconic, metaphoric and beat gestural information, but there is a lot of work to be done in the deictic area first, said Kettebekov. In addition, understanding what these subtler gestures mean from a linguistics point of view "is not there yet—so there's not enough theoretical basis," to use to give that understanding to computers, he said.

Kettebekov's research colleague was Rajeev Sharma of Pennsylvania State University. They presented the research at the Engineering for Human-Computer Interaction conference in Toronto in May, 2001. The research was funded by the Army Research Laboratory and the National Science Foundation (NSF).

Timeline: Now
Funding: Government
TRN Categories: Human-Computer Interaction; Computer Vision and
Image Processing
Story Type: News
Related Elements: Technical paper, "Toward Natural Gesture/ Speech Control of a Large Display," presented at the Engineering for Human-Computer Interaction conference in ronto, May 11-14, 2001.

TRN

Integrated Inputs Improve Interactivity

By Kimberly Patch, Technology Research News
September 13, 2000

Communicating with a computer through the usual channels—pressing plastic keys and pointing with a mouse—is pretty limited compared to the sound, eye contact, gesture and touch of human conversation.

For years, researchers have been working on widening human-computer bandwidth via speech recognition, eye and gesture trackers, and force feedback devices designed to allow a computer to communicate tactile sensations.

Each one of these technologies promises to help human-computer communications, but the real trick is being able to use them all at once in a way that feels natural. To that end, Researchers at Rutgers University have put together a desktop system, dubbed Stimulate, that coordinates input from all of these means of communication.

"Our focus was ... to achieve more natural communication between the human user and the networked computer system," said project leader James Flanagan, VP for Research at Rutgers University. Flanigan defines natural human interaction as including things like facial expression and manual gestures "in a hands-free mode where you don't have to wear or hold sound pickup equipment in order to transmit your message."

To do this, Stimulate uses a camera to track the user's face and eye movements; an array microphone mounted on the monitor to pick up the user's voice and distinguish it from background noise; and a three-ounce glove to track finger gestures and provide tactile feedback via pneumatic pistons.

The system uses speech recognition software to interpret users' words, and text-to-speech synthesis so it can answer back.

The camera is gimbaled, meaning it is able to swivel in all directions to track fine movements. It uses an ultrasonic range finder and a face recognition algorithm to find the user' s face and watch for visual gestures. Software maps the cursor movement to the user's eye movement so "you can just move the cursor by looking," said Flanigan. The camera tracks pupil movement by shining an infrared beam at the eye and computing the angle between the center of the pupil at the beam's reflection off the cornea, he said.

The glove includes a position detector and little pneumatic thrusters that can apply pressure to the fingertips. (See picture.) "You can reach into a complex scene and move an object—you can detect the position of it, the shape of it [and] the squishiness of it—how much it pushes back when you grab it," said Flanigan.

The difficult part was coordinating the different inputs, said Flanigan. The researchers' Fusion Agent software interprets sensory inputs and estimates the user's intent by putting everything in context, which is essentially a type of semantic analysis," he said.

For instance, a user might point to an object and say 'move this to there.' Interpreting the command involves knowing what 'this' is and knowing where 'there' is. The software must look at all the inputs simultaneously because the user might point with an eye movement or a hand gesture. It gets more complicated when inputs are redundant or contradictory, Flanigan said. "You might speak and point. ... so the software agent has to maintain some context awareness of what the transaction is and what objects are being addressed and what actions are requested. In order to interpret all this, the software must perform syntax analysis and semantic analysis, Flanigan said.

"I think it is a significant work," said Jie Yang, a research scientist at Carnegie Mellon University. " In terms of the microphone array they are on the leading edge. [And the] major problem is you have to coordinate all the components together. This is a tough problem—it's not trivial."

It is this combination and coordination of sight, sound and tactile technologies that, "even though they are quite primitive technologies, transcend the capabilities of the traditional mouse and keyboard," said Flanagan. For instance, it is somewhat difficult to rotate a virtual object 22-and-a-half degrees to the right with a mouse and keyboard. "But if you wanted to reach into the scene and twist the object 22-and-a-half degrees you can do that, or if you want to say 'rotate that 22-and-a-half degrees clockwise' by speech, that's fairly convenient as well," he said.

The researchers are currently working on a wireless version of the system. "It's at a very early stage," said Flanigan, but the goal is "to be able to walk around with your personal digital assistant and use conversational interaction, eye tracking... manual gesture [and] stylus gesture." Toward that end, the researchers are working on a miniature gimbaled camera, said Flanigan.

Although there's a lot of work to do on both the input technologies and the software, the utility of more natural human computer communications is clear, said Flanagan. "I could imagine you will see applications in selected places in less than five years," he said.

Flanigan's colleagues in the project were Rutgers professors Greg Burdea, Joe Wilder, Ivan Marsic and Cas Kulikowski. The project was funded by the National Science Foundation.

Timeline: < 5 years
Funding: Government
TRN Categories: Human-Computer Interaction
Story Type: News
Related Elements:

TRN

Sounds Attract Camera

Chhavi Sachdev, Technology Research News
July 25, 2001

When Steven Stills penned the lyrics, "Stop! Hey, what's that sound? Everybody look what's going down," he was describing a natural phenomenon that we seldom think about consciously—sounds make us look. When people clap, shout, or whistle to get our attention, our heads instinctively swivel towards them. Imagine the potential of a robot that reacts the same way.

University of Illinois researchers are taking steps towards that goal with a self-aiming camera that, like the biological brain, fuses visual and auditory information.

In time, machines that use vision systems like this one could be used to tell the difference between a flock of birds and a fleet of aircraft, or to zoom in on a student waving her arm to ask a question in a crowded lecture hall.

Source: University of Illinois

The self-aiming camera uses the microphones and camera in the foreground to find sounds and motion, which it then records with the camera in the top right corner.

The self-aiming camera consists of a video camera, two microphones, a desktop computer that simulates a neural network, and a second camera that records chosen information. The microphones are mounted about a foot apart to mimic the dynamics of an animal's ears.

The heart of the researchers' system is a software program inspired by the superior colliculus, a small region in vertebrates' brains that is key in deciding which direction to turn the head in response to visual and auditory cues. The colliculus also controls eye saccades—the rapid jumps the eyes make to scan a field of vision.

In determining where to turn the camera, the system uses the same formula that the brain of a lower level vertebrate like a barn owl uses to select a head-turning response, said Sylvian Ray, a professor of computer science at the University of the Illinois at Urbana-Champagne.

Source: University of Illinois

The camera aims in the direction of the sound of a researcher clapping his hands.

The computer picks out potentially interesting input and calculates the coordinates where sound and visual motion coincide, Ray said. "The output [is] delivered to a turntable [under] a second camera. The turntable rotates to point the second camera at the direction calculated by the neural network."

In this way, the second, self-aiming camera captures the most interesting moving object or source of noise on film for further analysis, saving a human operator the chore of sifting through all the data.

The camera re-aims every second toward the location most likely to contain whatever is making the most interesting noise. Because it always chooses an estimate of the best location for all available input, the system works even if several motions or noises happen at once, according to the researchers.

The self-aiming camera could be used to pare down meaningless data captured by surveillance systems that use several cameras to take pictures of their surroundings, according to Ray. It could be used as an intelligent surveillance device in hostile environments, and for ordinary security, he said.

To have the system differentiate among different types of inputs, the researchers plan to add specializations that will give certain inputs more weight. In nature, different vertebrates respond to particular targets; a cat likes different sounds and motions than an owl, for instance, said Ray. "One specialization of the self-aiming camera would be to train it to like to look for human activity," he said.

"This is a nice example of exploiting ideas from biology to better engineer systems since this pairing of stimuli increases the reliability and robustness of the self-aiming camera," said John G. Harris, an associate professor of electrical and computer engineering at the University of Florida. A better understanding of the underlying neural mechanism is still needed, he said.

Eventually the system will have to deal with multiple objects as well as noise and room reflections, Harris said. "Such a system needs an attention mechanism in order to attend to objects of interest while ignoring others. This is a higher level behavior that requires different sets of neurons and is beyond the scope of the current demonstrated system," he said.

The system could be in practical use in two to three years, according to the researchers. Ray's research colleagues were Thomas Anastasio, Paul Patton, Samarth Swarup, and Alejandro Sarmiento at the University of Illinois. The research was funded by the Office of Naval Research.

Timeline: 2 to 3 years
Funding: government
TRN Categories: Neural Networks; Computer Vision and Image Processing
Story Type: News
Related Elements: Technical paper, "Using Bayes' Rule to Model Multisensory Enhancement in the Superior Colliculus," Neural Computation 12: 1141-1164.

TRN

Biometrics Takes a Seat

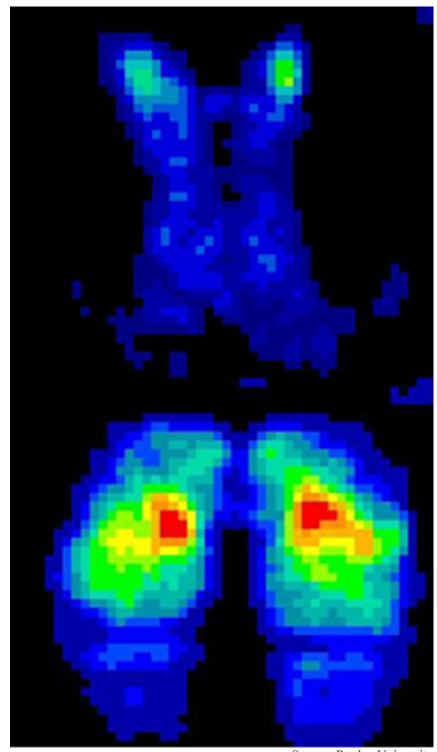
By Kimberly Patch, Technology Research News
November 15, 2000

Your chair may be privy to a considerable amount of information, being so close to you so much of the time.

Researchers at Purdue University are taking advantage of the close relationship by adding pressure sensors to the seat and back of a chair so it can track how a person is sitting.

Currently, the chair can sense only static postures, but the researchers' eventual aim is a real-time tracking system that can sense how a person shifts throughout the day. This may eventually allow the chair to react in useful ways to how a person is sitting, or even to serve as a computer interface.

The chair's sensors are thin sheets of plastic embedded with circuits that give pressure readings from 4,000 separate spots, or pixels, on the chair back and seat. When someone sits in the chair, individual points of pressure at each pixel are recorded as numbers between zero and 255, resulting in a 3-D pressure map.



Source: Purdue University

This shows a full, 4,000-point pressure distribution map from a person sitting upright in a pressure-sensitive chair.

To enable the chair to determine how a person was sitting, the researchers had 30 people of different sizes sit in the chair in 10 distinct positions, including upright, leaning in one of several directions, slouching, or crossing either leg.

"The computer builds up a model of what [a given posture's] pressure should look like, and the population variation associated with it, because nobody sits down exactly the same way twice, and different people sit slightly differently," said Hong Z. Tan, assistant professor of electrical and computer engineering at Purdue University.

An image-processing algorithm used in face recognition did a principal component analysis to reduce the 4000 points of data to 15 potentially important features. Features are things like the position of the highest pressure point, the distance between sitting bones, or the total contact area, Tan said.

This data reduction process pinpointed both common features within a sitting position and the distribution of variation within those features. For example, "what's the mean pressure map for everybody sitting upright, and then what's the range of variation," said Tan.

Once the computer had the measurements for those samples, it was able to determine the posture of a person sitting in the chair by comparing pressure readings, she said.

"It's a very interesting first piece of work in this area," said Irfan Essa, assistant professor of computing at Georgia Tech. "I have not seen much work that other people have done on instrumenting chairs like this. In essence it's a first step in that direction," he said. In addition, the current work "could be used in a very limited sense [to] try to help improve... a person's posture," he said.

The chair's static posture abilities are a first step in a more ambitious dynamic tracking project, said Tan.

Once the chair can determine static postures, it can go on to show how a person moves in a chair. "[To] do the continuous tracking, we need to catch static postures very well and then model the transition. So this is the beginning portion of our work [toward] a continuous tracking system," said Tan.

To that end, the researchers are investigating two different methods. "One technique is the Hidden Markov Model, which people use to model speech. We think that could be used to automatically learn the pattern between different sitting postures when people switch between one and another," Tan said. The researchers are also looking into using neural networks in a similar way.

The researchers are aiming to make a user-independent system where "anybody could just sit down and the computer can start to track their posture," said Tan.

Potential practical uses of this ability range from ergonomics to driving safety to computer control.

For instance, it could be used as a tool to evaluate both the sitting habits of people and the ergonomics of chairs. A chair that keeps track of how a person moves all day could keep statistics and predict consequences, said Tan. "It can say something like, 'for 20 percent of today you're slouching. This is the typical pressure distribution when you slouch, and [this] part of your body is not properly supported when you slouch,'" she said.

As this type of technology matures there could be a lot of interesting possibilities like sensing if a person is falling asleep or getting tense while driving, said Essa. "This has a lot of potential," he added.

It could also serve as a sensitive human-machine interface, said Tan. "Here's an object that everyone sits in when they interact with computers... what if I use the chair [as a] control?"

For example, in an intelligent teleconferencing room application, "if I turn, maybe a remote camera pans left and right. If I lean forward maybe that means I'm interested — it should zoom into whatever is the center of the scene. If I lean back maybe I want to get a shot of the whole room. This would a very intuitive kind of interface [that] you wouldn't even have to think too much about," said Tan.

And further into the future, a sensing chair may be able to at least assist in biometric identification. "It's pretty hard to fool a computer, the way the pressure distributes," Tan said.

For instance, from a chair's point of view, the parts of a person that produce the most pressure are the sit bones. And one of the very distinct features the chair can determine is the distance of the sitting bones, said Tan.

Even the current version of Tan's chair can tell with a fair amount of accuracy the difference between a male and female based on the size of the space between these bones. "If you are close to 16 cm [it's] very likely a female, if you're close to 14 cm it's very likely it's a male," she said.

A real-time posture tracking chair robust enough for practical applications is three to five years away, said Tan.

Tan's research colleague is Lynne A. Slivovsky of Purdue University. Slivovsky presented the research at the 2000 International Mechanical Engineering Congress and Exposition in Orlando, Florida on November 9, 2000. The research was funded by Purdue University and by the National Science Foundation (NSF).

Timeline: 3-5 years
Funding: University, Government
TRN Categories: Applied Technology; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "A Real-time Static Posture Classification System," presented at the 2000 International Mechanical Engineering Congress and Exposition's Symposium on Haptic Interfaces for Virtual Environments and Teleoperation in Orlando, Florida, November 9.

TRN

Hot Spots Give Away Lying Eyes

By Kimberly Patch, Technology Research News
January 23, 2002

Researchers from the Mayo Clinic and Honeywell Laboratories have come up with a way to measure these heat changes in a person's face in order to tell whether the person is lying or telling the truth.

The system consists of a high-definition thermal imaging camera and a computer. The camera takes pictures of heat emanating from a subject's face, and the computer provides a quick analysis of any changes.

Monitoring blood changes in the face is similar to the traditional polygraph exam, according to James Levine, a consultant at the Mayo Clinic. The polygraph lie detector test measures changes in a subject's breathing, pulse rate and blood pressure. It also measures sweating by sensing changes in skin conductance via electronics attached to the skin.

The thermal method measures infrared lightwaves, or heat, around the person's face. The infrared light shows up on the computer screen as red areas. The theory behind using thermal changes in a person's face to detect lying is similar to the principal behind the polygraph. When someone is not telling the truth there is likely to be instantaneous warming around the eyes, which is probably a natural response produced by the sympathetic nervous system, said Levine.

The thermal method is as accurate as traditional polygraph tests, according to Levine. It is also faster than polygraph tests and doesn't require the subject to be connected to a device, he said.

The researchers tested their theory and the system at the U.S. Department of Defense Polygraph Institute. Twenty volunteers were randomly assembled into two groups. One group was instructed to stab a mannequin, take $20 from it, then lie about what took place.

The thermal imaging system correctly identified six of eight of the subjects who were lying, and 11 of 12 who were innocent. The subjects were also put through traditional polygraph tests, which correctly identified the same number of guilty subjects, but correctly identified only eight of the 12 innocent subjects.

Thermal imaging operators would not need the type of training to carry out the tests that traditional polygraph tests require, according to Levine.

"The technique sounds interesting and promising," said Christoph Koch, a professor of cognitive and behavioral biology at the California Institute of Technology. "For mass security and screening applications, you need a technology that can rapidly, at low-cost and with a low false alarm rate, screen people."

The accuracy of polygraph methods is controversial, said Koch. However, if the thermal imaging technique is faster than polygraphs and if it is less prone to label truthful statements lies it is worth investigating further, he said.

The researchers are continuing to test the method and are aiming to turn it into a practical security application, said Levine.

Levine's research colleagues were Ioannis Pavlidis From Honeywell Laboratories and Norman L. Everhardt from the Mayo Clinic. They published the research in the January 3, 2002 issue of *Nature*.

Timeline: 2-4 years
Funding:
TRN Categories: Computer Vision and Image Processing; Applied Computing
Story Type: News
Related Elements: Technical paper, "Seeing through the Face of Deception," *Nature*, January 3, 2002.

Manners Matter for the Circuit-Minded

By Ted Smalley Bowen, Technology Research News
April 25, 2001

In his dystopian futuristic comedy, "Sleeper", Woody Allen's twentieth-century time traveler, on the lam as a domestic robot, is revealed when, among other breaches in automaton etiquette, he betrays a fondness for his owner's euphoriant orb.

While sophisticated androids are still the stuff of science fiction, robotics technology is creeping closer to the point when mobile robots will be commonly employed for personal use.

Anticipating frequent human-robot interaction, researchers are trying to get a sense of how people will be affected by the activities of their mechanized assistants. Such observations could lead to the design of well-behaved, and thus more effective, robots.

To this end, University of Kansas researchers put robots through their paces in the presence of human subjects and gauged the humans' reactions.

Among the lessons they learned: personal robot etiquette frowns on rushing headlong at people. This may come as no surprise, especially in the case of large robots, but relatively little quantitative research has been done on the psychological responses mobile robots elicit in humans, according to Arvin Agah, an assistant professor of electrical engineering and computer science at the University of Kansas.

Working with a commercially available mobile robot configured in two basic body types, the researchers recorded the reactions of forty people as robots approached and went around them, and when the robots simply moved about in their presence. The robots, which were based on the Nomadic Scout II made by Nomadic Technologies Inc., moved on two wheels and a caster.

The small robot body type was 35 centimeters high and 41 centimeters in diameter, or about the size and shape of a wide mop bucket. To make the larger body type, the researchers topped the small version with a rudimentary humanoid form to give it a height of 170 centimeters, or about five-and-a-half feet.

To determine the most acceptable ways robots might approach humans, the researchers guided robots of each size toward the human subjects in several ways.

In a direct approach, a robot went straight toward a human at the speed of 10 inches per second or at a faster clip of 40 inches per second.

In an avoidance mode, a robot moved around the subjects either by stopping to change direction or by making a continuous turn. The avoidance mode speed was 10 inches per second, but the evasive moves were made at a slightly faster 15 inches per second.

The robots were also set to work moving around the space while not interacting with the human subject. This involved both random movement and a more methodical sweep of the floor space.

The researchers carried out the experiments in the relatively close quarters of a lab room measuring about nine-by-fifteen feet. The subjects recorded their responses in a survey, rating them on a one-to-five numerical scale, with one representing very uncomfortable and five very comfortable.

In general, the humans liked the small robot better than the larger, humanoid version, said John Travis Butler, a software engineer at Lockheed Martin who participated in the investigation when he was a University of Kansas graduate student. "The smaller robot body was preferred in cases where the robot was moving fast or close to the subject due to the intimidation factor of the more massive-bodied robot," he said.

In the direct approach experiments, the humans were generally comfortable with the slower approach, and were not at ease with the fast approach.

The avoidance mode was met with general approval, with the most positive reception given to the nonstop pass-by performed by the robot in its smaller incarnation.

While generally at ease with both types of non-interactive behaviors, the subjects were slightly less comfortable with the structured movements, which involved frequent and slightly faster turning.

Some of the behavioral concepts gleaned from such experiments are already being used in experimental designs, said Agah. "In research laboratories, the behavioral research is starting to be incorporated into the design of personal robots. In the industry, mostly entertainment/companion/pet robots, this will be happening in the next five years," he said.

While the behavior studies could inform the design of robots for both workplace and home settings, the requirements for those venues will likely differ, said Butler. "I would expect a work environment to be more structured and easier for a robot to operate in. [The] home would be a more dynamic environment," he said.

Workplaces will also be much more concerned with the amount of work done per dollar spent on the robot and less concerned about the attractiveness or noise of the robot, he said. "A robot working in someone's home will have to be something you can tolerate looking at every day. This will be something that the user will have to live with much like a pet. The expectations will be much higher," he said.

The University of Kansas research largely confirms similar studies of human reactions to robot actions, said Dieter Fox, assistant professor of computer science and engineering at the University of Washington. "This is an interesting article on design issues involved in the development of human-friendly service robots. Our experience [also] suggests that high acceleration is the major factor that makes people uncomfortable when being approached by mobile robots," said Fox.

However, Fox's own research shows one difference in human acceptance of robots. "In slight contrast to the results presented in this article, we had good experience with taller robots carrying human features," he said.

The next step in this type of research, said Butler, is evaluating more complex human-robot interactions by having robots perform more varied tasks with human subjects. "More interaction would give a better understanding of how people and robots will fit in the same environment," he said.

University of Kansas researchers are working on extending the work using robots that interact with people by responding to verbal and visual commands such as 'put the green one over there,' said Agah. "This requires dealing with ambiguity resolution, a concept that necessitated our multidisciplinary team of researchers including faculty from departments of electrical engineering and computer science, psychology, and linguistics," he said.

Additional work might include more detailed evaluations of human subject's behavior when they share space with mobile robots, said Butler. "Monitoring subjects as they perform normal daily activities while in the presence of an active robot would provide very interesting results," he said.

The researchers described their experiments in the March, 2001 issue of the journal *Personal Robotics*. The work was funded by the University of Kansas department of electrical engineering and computer science.

Timeline: 5 years
Funding: University
TRN Categories: Robotics; Computers and Society
Story Type: News
Related Elements: Technical paper "Psychological effects of behavior patterns of a mobile personal robot," *Personal Robotics*, March, 2001.

Interactive Robot Has Character

By Eric Smalley and Susanna Space, Technology Research News
March 6, 2002

Combine some of the most advanced human-computer interaction technology with one of the oldest forms of entertainment—puppetry—and you get Horatio Beardsley.

Doc Beardsley is an animatronic robot, a descendant of the mechanical humans and beasts that rang bells and performed other actions as parts of the clocks of medieval European cathedrals. Modern science, however, has carried Doc far beyond these ancient automata, endowing him with the ability to see, understand spoken words and carry on a conversation.

Researchers at Carnegie Mellon University made the amusing, forgetful inventor as a literal embodiment of a computer interface. Doc performs for audiences, answering questions about himself. He claims to have been born on a mountaintop in Austria to a family of goatherds, and to have invented endless bread, the milkbed, the antisnooze and the foon.

Source: Carnegie Mellon University

Doc Beardsley's animatronic body gives him mobility and facial expressions. Unlike ordinary animatronic figures, however, Doc can see, hear and hold up his end of a conversation.

In addition to paving the way for a future generation of theme park animatronic characters, the technology could lead to embodied personal digital assistants with personalities, interactive electronic pets, animated historical museum figures, and robotic waiters and salespeople, said Ron Weaver, a graduate student at Carnegie Mellon.

Several layers of software drive Doc's apparent wit. Synthetic interview software, which includes speech recognition abilities, allows Doc to react to spoken questions. The technology, developed at Carnegie Mellon for use with video characters, gives a character sets of lines to deliver on given topics. This allows Doc to give appropriate answers to questions that match an anticipated query closely enough, said Weaver.

If the question hasn't been anticipated, another layer of software takes over. A discussion engine tracks the questions and answers during a conversation and allows Doc to make relevant comments by keying off individual words even if he doesn't understand a specific question, said Weaver. And if that doesn't work, the discussion engine tosses the conversation back to the questioner, he said.

"The discussion engine will first try to deliver a comment that is still relevant based on whatever individual keywords can be found in the text," said Weaver. "Failing that, the character gives a random comment that either pretends to reflect what is being discussed to try to keep [the conversation] going, or transfers the onus of the conversation back to the guest."

The researchers made Doc forgetful to help keep his lapses from breaking the illusion of a realistic personality.

Here's an actual exchange:

Questioner: "How are you feeling today?"

Doc: (pause) "My—my dog just died. No, I'm just kidding. He just got run over. He's still alive. Well, it's a she really. I just refer to her as a boy to bust her chops, keep her honest, you know?"

Questioner: "What's your best invention?"

Doc: "Well, you know, there was this one invention I made accidentally while working on genetic recombination. I crossed a spoon with a fork and thus created the foon. It was used either as a spoon or a fork."

Microphones and cameras that track the sounds and movements of an audience provide Doc's audiovisual perception, said Todd Camill, a research engineer at Carnegie Mellon University's Robotics Institute. "Microphones listen for sounds made by people in the room, and a camera subsystem tracks the movement of people by finding areas of skin tone in the room. The audio and vision systems generate position data that tell Doc where to turn his head."

One aspect of making Doc Beardsley a believable character is keeping the technology in a supporting role, according to Tim Eck, another Carnegie Mellon graduate student. "Character and story are the most important aspects to creating believable, entertaining characters," he said. "We are striving to provide the illusion of life, to create an entertaining experience, which is an important distinction. We are not trying to create artificially intelligent agents. We are creating the illusion of intelligence with time-tested show business techniques: drama, comedy, timing and the climactic story arc."

As with many creative endeavors, serendipity plays an important role. "From time to time, we find ourselves caught off guard by conversations that seem to make sense in ways we did not intend," said Camill. "For example we've recently heard this exchange:

Guest: 'Doc, why are you wearing a Carnegie Mellon University sweatshirt?'

Doc: 'I've spent time at many universities. You'd be surprised at the things they throw away.'"

In addition to using traditional storytelling and theatrical techniques, the researchers are studying the human side of human-computer interaction. "Since our goal is the illusion of human intelligence or intent in the service of a story, a large part of our results concern the human audience rather than the robot," said Camill. "We are exploring the social dynamics between human and machine by exploiting the tendency of people to project human qualities on the objects around them."

From the entertainment perspective, the ultimate goal is creating synthetic characters that seem to possess dramatic human qualities, like a sense of humor, comic timing, personal motivations and improvisation, said Camill. "When an audience can get so engrossed in interacting with Doc's dialogue and story that the technology is completely forgotten, then we know we have accomplished our goal," he said.

The next steps in the project are improving the character by adding skin and a costume, building a set and props, creating a show, building puppeteering controls for the props, and writing software for producing other shows, said Camill.

The technology is not yet ready for the entertainment industry, said Eck. "The main reason [is] speech recognition technology. We believe once the overall accuracy of speaker-independent speech recognition is 80 percent or higher, applications such as ours will be seen in the entertainment industry. This will be approximately 5 to 8 years from now," he said.

The research is funded by the Carnegie Mellon University.

Timeline: 5-8 years
Funding: University
TRN Categories: Robotics; Artificial Intelligence
Story Type: News
Related Elements: Project website: micheaux.etc.cmu.edu/~iai/web/newIAI/doc.html

TRN

Monkey Think, Cursor Do

By Eric Smalley, Technology Research News
March 20/27, 2002

Computers fall far short of being able to read our minds. This could change. Brown University researchers have shown that, with the right filters, computers can interpret the electrical signals brain cells send to move limbs.

The researchers implanted electrodes in the brains of three rhesus monkeys and recorded neural activity as each monkey used a hand control to move the cursor on a computer screen. The implant, a centimeter-wide silicon chip covered with tiny spikes, recorded the signals from a small number of motor neurons in the monkeys' brains.

The researchers then built a mathematical algorithm that converted these neural signals into a control signal that moved the cursor. The algorithm translated the brain signals into computer signals in real-time, which allowed the monkeys to pursue a moving spot on the computer screen with the cursor just by moving their arms.

The algorithm averages the signals from 7 to 30 motor neurons to estimate where each monkey intends to move its hand. "It's as if each neuron gets a series of votes on where it thinks the hand is," said Mijail Serruya, a graduate student at Brown University. "Some of the votes relate to how the neuron feels right now, some relate to how it felt up to one second ago. The model... uses this to guess new hand positions from the neural activity alone," he said.

The researchers' system was able to produce a control signal after recording only a few minutes of the monkeys' manual control of the cursor, according to Mijail Serruya, a graduate student at Brown University.

"The scientific principle of decoding [motor neuron] activity rapidly, online, in a useful manner is now proven," said Serruya. the method could eventually help people who are paralyzed control electronic devices, he said. "This paves the way for possible development of a medical device that could help paralyzed patients."

One monkey eventually learned to control the cursor without visibly moving its arm. The researchers could not determine whether the monkey was using subtle muscle movements to produce the neural signals, however, and so do not yet know whether thought alone can be used to produce the control signal.

The neural control was as efficient as hand control at the task of pursuing the spot on the screen, said Serruya.

Paralyzed humans have already used brain implants to control computer screen cursors. But in those experiments, the subjects took months to learn how to use the system, said Serruya. "Any neuroprosthetic system requires both the machine and the person to learn," he said. "We believe that by having our machine—the mathematical algorithm — do a lot of learning, it makes it much easier and faster for the subject to learn their part."

In a similar experiment last year, researchers at Duke University, the Massachusetts Institute of Technology and the State University of New York Health Science Center used a monkey's motor neuron signals to control a robotic arm. In that case, the robot arm simply mimicked the actions of the monkey's arm, and the monkey did not consciously control the robot arm.

The Brown University monkeys controlled the cursors consciously in order to win rewards.

The researchers are considering applying their technique to other output devices, said Serruya. It's too soon to estimate when or if the technique could be applied to humans, he said.

Serruya's research colleagues were Matthew R. Fellows and John P. Donoghue of Brown University, and Nicholas G. Hatsopoulos and Liam Paninski who are now at the University of Chicago. They published the research in the March 14, 2002 issue of the journal *Nature*. The research was funded by the National Institute of Neurological Diseases and Stroke, the Defense Advanced Research Projects Agency (DARPA) and the Burroughs Welcome Foundation.

Timeline: Unknown
Funding: Government, Corporate
TRN Categories: Biotechnology; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "Instant Neural Control of a Movement Signal," *Nature*, March 14, 2002.

TRN

Brain Cells Control 3D Cursor

By Eric Smalley, Technology Research News
June 12/19, 2002

Researchers at Arizona State University have developed a feedback system that lets monkeys use brain signals to move a virtual ball within a computer-generated box, an advance that increases the chances that scientists will be able to give disabled people neural control of prosthetic limbs.

The research also suggests that surgeons will eventually be able to rewire bodies to give people control over paralyzed body parts.

The feat is the latest in a string of advances that allow brains to directly control electronics. Scientists have been planting electrodes in the brains of animals to record electrical activity for decades, but they have only recently been able to use these neural signals to control robots and computers.

In 2000 researchers from Duke University, MIT and the State University of New York Health Science Center tapped a monkey's brain signals to make a remote robotic arm mimic the movements of the monkey's own arm. Earlier this year, researchers at Brown University showed that the method could allow monkeys to consciously control a computer cursor, and found that one monkey learned to move the cursor without physically moving its arm.

The Arizona State experiment goes beyond two-dimensional cursor control to give a pair of rhesus macaque monkeys direct cognitive control of a virtual ball in a three-dimensional space. The Arizona monkeys also showed greater control over the ball than the Brown monkeys had over their cursors, said Andrew Schwartz, a research professor of bioengineering at Arizona State University. The ball control resembles "real biological movement," he said.

The key to the researchers' success is a feedback system between the monkeys' neurons and the software algorithm used to translate brain signals into computer signals.

Each of the many billion neurons in a primate brain is connected to as many as 10,000 other neurons. Learning occurs when the brain adapts to different conditions by changing the patterns of signals neurons transmit and receive.

The researchers added their software to the learning loop, allowing it to adapt along with the changing neural signals within a small portion of a monkey's brain. "We are using a more sophisticated approach that allows two-way learning to take place," said Schwartz. "The animal learns to move the cursor using biofeedback to change the discharge patterns of its neurons. Our decoding algorithm tracks these changes as they occur" in order to make better predictions about the new neural patterns, he said.

The monkeys were able to move the ball using brain signals alone almost as well as they were able to control it with arm movements, said Schwartz.

To teach the monkeys this cognitive control, the researchers implanted electrodes in the motor cortex region of their brains. Motor neurons coordinate muscle activity. The researchers rewarded the monkeys for using arm movements to move the ball to a particular spot, and recorded the neural activity. They used this recording to calibrate the software that translates the neural activity into the control signal for the computer.

The researchers then restrained the monkeys' arms so that the monkeys could not physically move their arms as they attempted to move the virtual ball. At first the monkeys pushed against the restraints in the direction they wanted the ball to move, but stopped straining as their performance improved. Measurements of the monkeys' brain activity showed that eventually the monkeys could control the ball without using the normal brain signal patterns associated with muscle movements, indicating that the neurons had adapted to a new circumstance, according to Schwartz.

Both the Brown and Arizona State research teams were able to use a surprisingly small number of neurons to generate a control signal. Monkeys have millions of neurons in the motor cortex, but the Arizona State researchers used the signals from less than two dozen motor neurons to generate the ball control signal, according to Schwartz.

Using so few neurons would not be practical for controlling a prosthetic device, said Miguel Nicolelis, a professor of neurobiology and biomedical engineering at Duke University, and the lead researcher on the robotic arm project. "If you [were to] lose a couple of neurons your entire implant [would] become useless," he said. "Much larger neuronal samples are needed" to make an implant practical over time.

Controlling a neuroprosthetic arm is also considerably more complicated than moving a cursor in a three-dimensional space, said Nicolelis. "To reproduce complex 3D hand and arm trajectories and to mimic the force required to move objects with a prosthetic arm, hundreds of neurons would be needed," he said.

The Arizona State researchers' next step is to replace the virtual cursor with a robot arm that will be used by a monkey to retrieve food while its arms are restrained, said Schwartz.

And over the next two or three years, "we would like to try these implants... in human patients," said Schwartz. Such a neural bypass could be used to give disabled people control over computer-driven prosthetic limbs. The technique could also be combined with electrical signals that stimulate muscle movement in order to let paralyzed people regain control of their own limbs, said Schwartz.

Schwartz's research colleagues were Dawn Taylor and Stephen Helms Tillery of Arizona State University. They published the research in the June 7, 2002 issue of the journal *Science*. The research was funded by the National Institutes of Health, the Whitaker Foundation, the Philanthropic Education Organization and the U.S. Public Health Service.

Timeline: 2-3 years
Funding: Government, Private
TRN Categories: Biotechnology; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "Direct Cortical Control of 3-D Neuroprosthetic Devices," *Science*, June 7, 2002.

Virtual Touch Controls Rats

By Kimberly Patch, Technology Research News
May 15/22, 2002

Since the late 1930's scientists have known that electrical stimuli to different regions of the brain could be used as teaching cues. In the '50's, researchers found that stimulating the medial forebrain bundle in the hypothalamus is rewarding enough that an animal will work to obtain the stimulation just as it will for food or water.

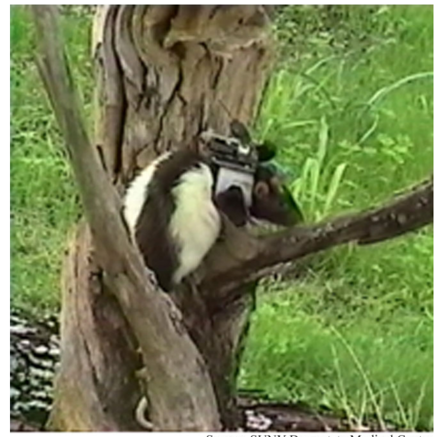
Researchers from the State University of New York Downstate Medical Center and Drexel University have combined reward stimulation with stimulation of a part of the brain that processes the sense of touch in order to remotely control the movements of a rat up to 500 feet away.

The achievement provides insights into how animals learn, and may eventually lead to better prosthetics that provide tactile feedback. "A larger aim behind the guided rat was to begin studying the kind of effects, both on behavioral control and perception, that stimulation of sensory regions of the brain may have," said Sanjiv Talwar, a research assistant professor at SUNY Downstate Medical Centre. "This would enable us to experiment with ways to [use] sensory feedback in order to better control a robot arm," he said.

It also shows that it is possible to control wired rats as if they were intelligent robots. "The guided rat could easily be used as a robot platform in situations where the terrain is too complex for [mechanical] robots to traverse or where stealth may be a prerequisite," said Talwar. The rats could be equipped with small video cameras, GPS systems and other sensors, he said.

The key insight that led to this type of control was thinking of conditioning animal behavior wholly in terms of virtual cues and rewards rather than external cues and rewards like sounds and food, Talwar said.

To train a rat this way, the researchers implanted electrodes that were about the diameter of a hair into three parts of the rat's brain. One electrode stimulated the reward center in the hypothalamus. The other two stimulated areas of the right and left somatosensory cortex that receive input from a rat's whiskers. The somatosensory stimulations probably give the rat the illusion of touch, said Talwar. It will seem to the rat like it is being touched on its left whisker, for instance, when electricity passes through the electrode on the right side of the somatosensory cortex, he said.



Source: SUNY Downstate Medical Center

Researchers can control this rat's actions remotely by sending signals to a computer chip the rat carries on its back. The signals trigger three electrodes wired into the rat's brain, two of which give the rat the illusion it is being touched on its left or right whiskers. The third electrode is implanted in the reward center of the brain.

The electrodes were connected to a small adapter cemented to the top of the rat's skull. The adapter was connected to a microprocessor carried in a small pack on the rat's back. The researchers controlled the electrode stimulation remotely using a PC that sent radio signals to the microprocessor. "The system we came up with was built using simple, off-the-shelf electronics," said Talwar.

The researchers showed that rats could be trained using the electrodes and a figure-eight maze. When the rat moved forward, the researchers gave it a hypothalamus reward stimulation, and within ten minutes the rat learned that when it continued to move forward it received a reward every second.

Whenever the maze presented the rats a choice of turning left or right, the researchers stimulated one side of its whiskers. When a rat made the corresponding turn, it was immediately given a hypothalamus reward stimulation. If it made an incorrect choice it was deprived of hypothalamus reward stimulation for five seconds.

After seven or eight 10-minute training sessions, "the rats learned to move forward continuously and respond with near 100 percent accuracy to turning commands," said Talwar.

The researchers then tested the rats in open environments and found that they still responded to the turning commands. “Even in open space they would move forward for periodic... rewards and respond correctly and instantaneously to turning commands,” said Talwar.

The researchers were also able to get the rats to climb or jump over objects using the stimuli, Talwar said. “By giving [reward] stimulations along with right-turn and left-turn cues we... were able to guide the rats over any terrain that was within their capability to overcome. These terrains included wide-open spaces, both indoors and outdoors, complex 3-D mazes, trees, pipes, narrow ledges at considerable heights, stairs, and large concrete-block rubble piles,” he said.

The research is exciting, and bears on the whole field of neural prosthetic development, said Miguel Nicolelis, a professor of neurobiology and biomedical engineering at Duke University. “It’s a very important step because it shows that you can actually, in a very predictable way, deliver signals to the brain of an animal and condition the animal to interpret the signals, particularly using the sense of touch,” he said.

The research shows that you can get touch-like feedback to the brain and train animals to interpret it as if they were being touched on the face, said Nicolelis. Guiding animals using a sense of touch “is very new, no one has done it before,” he said.

The research has several potential uses, said Nicolelis. “One is on the basic science level as a new paradigm to study fast learning,” he said, pointing out that with the stimuli, the rats were learning to do a maze on the first try. “This is something that allows you to look at the fully behaving animal and measure what is going on in [brain] circuits,” he said.

The research is also significant for the development of prosthetic devices that use a touch-like interface to restore motor, sensory or even cognitive function, said Nicolelis. “This was a missing step that several people were trying to achieve — [the researchers have] done it.”

A sense of touch would be an important type of feedback for a paralyzed patient using a robotic arm, for example, said Nicolelis. “You need to provide some sort of feedback information to this patient,” he said. One possibility is visual feedback, “but when you’re grabbing an object like a glass of water, vision alone is not going to help you—you need to have this touch feeling of what it is you’re touching, how heavy it is, otherwise you’re going to either drop it or smash it because the force that you’re going to apply is not going to be appropriate,” he said.

Another potential application is to use remote-control rats to search rubble for victims of earthquakes and bombings, said Nicolelis. “If you can steer a rat... through rubble, it would be [useful] because they are smaller than dogs, they’re very good sniffing animals, and they have these phenomenal facial whiskers... that can discriminate very fine objects,” he said. “Theoretically that could be a... development if it proved to be reliable,” he said.

The researchers’ next steps are to explore the possibility of a working sensory prosthesis, said Talwar.

Although the capabilities of the guided rat could easily be developed further, “at present, however, we’re not aiming at this,” he said. It would take two to three years to develop the concept into practical processes shaped around specific applications, he added.

“Ethical considerations may play a role in future development, and a wider debate will be required for this to happen,” said Talwar. The rat experiments were performed within National Institutes of Health guidelines, and “in addition our behavioral model is based only on a reward system with no associated food or water deprivation,” he said.

“Nevertheless, for some there may still appear to be something creepy about using a guided rat for real-world tasks,” said Talwar. “This must be acknowledged—after all it will be easy to extend the same method to any species,” he said.

Talwar’s research colleagues were Shaohua Xu, Emerson S. Hawley, Shennan A. Weiss and John K. Chapin SUNY’s Downstate Medical Centre, and Karen A. Moxon of Drexel University. They published the research in the May 2, 2002 issue of the journal *Nature*. The research was funded by the Defence Advanced Research Projects Agency (DARPA).

Timeline: 2-3 years
Funding: Government
TRN Categories: Biotechnology; Applied Technology
Story Type: News
Related Elements: Technical paper, “Rat Navigation Guided by Remote Control,” *Nature*, May 2, 2002.

TRN

Neuron-Chip Link Advances

By Kimberly Patch, Technology Research News
March 7, 2001

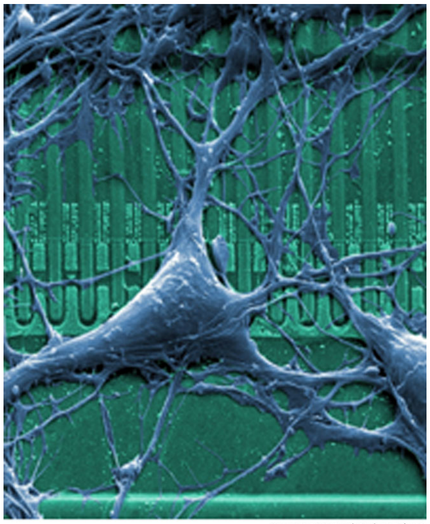
The differences in biological versus electronic communications are far beyond Mars/Venus comparisons.

Cells communicate when ions suspended in water flow through membranes, while signals pass among electronic components when electrons flow through solid metal or semiconductor. Electrons flow through silicon and metal about 100 billion times faster than ions flow through water. And there’s also a considerable size problem: cellular ion channels, at about 3 nanometers in diameter, are smaller than today’s tiniest electronic components.

A direct ion channel-electronic interface, however could eventually prove useful in biosensors, brain-computer interfaces, or even open up the possibility of neural prosthetics.

Although devices that combine biology and electronics—like pacemakers and cochlear implants—exist today, the electronics are too large to actually interface with ion channels.

Researchers from the Max Planck Institute for Biochemistry have taken a step toward more specific biological-electronic communications with a device that allows a field effect transistor on a silicon chip to sense the electric potential created by a nerve cell, and the nerve cell to sense a voltage pulse in the chip.



Source: Max Planck Institute

This electron micrograph shows a nerve cell connected to an oxidized silicon chip coated with collagen. The ion current the cell uses for communication flows along the narrow gap between the cell and chip and affects the silicon electrons' flow.

The researchers cultured the cells on an oxidized silicon chip coated with collagen. The cell's ion current flowed through the cell membrane and then continued along the narrow gap between the cell and the chip. This gap current gave rise to a voltage drop that changed electron flow in the silicon on the other side of a thin insulating oxide layer.

In this way, "the elementary electrical current of the cell [was] translated into a charge of electrical current in the chip without current flow across the interface," said Peter Fromherz, a professor of biophysics at the at the Max Planck Institute for Biochemistry in Germany.

The setup is in essence a direct connection of ion channels in a cell membrane with the electron channel, said Fromherz. An ion is an atom that carries electric charge because it has lost or gained one or more electrons. A neuron fires when it collects enough ions to build up a substantial electric charge, or potential on the inside of the cell.

The researchers also took the communication a step further by creating a feedback loop where the current modulation triggered the chip to pass a voltage pulse to a stimulation spot in the neuron, which made it fire again.

The researchers also used a pair of neurons connected both to each other and to the chip to perform this feedback loop sequentially and in parallel.

The neuron-chip device is smaller and the method more precise then existing ion-electron converters that connect electrodes to cells and allow current flow across the interface, said Fromherz. It is also different because it directly integrates a transistor with a biological component.

The researchers' device "talks directly with these ion channels, a new step down in size and step up in precision," said Richard Granger, a professor of information in computer science at the University of California at Irvine and CEO of Thuris Corporation, a neurotechnology company. The work "foreshadows incipient conversations between devices and living tissues," he said.

Granger cautioned, however, that there are many steps between this research and practical ion channel-computer chip communications. The research has "advanced understanding of how we might talk to neurons, but we still do not know what neurons say to each other, nor even what language they use."

The research, ironically, may eventually help with the language problem, said Granger. "Bioelectronic interfaces for the brain may enable us to listen to brain circuits before we can translate what they say," he said.

The researchers next steps are to improve the interface and to try to control the ion channels from the computer chip, said Fromherz.

There are two classes of biological ion channels. One type is controlled by voltage, the other type is opened when a special type of molecule, a ligand, binds to a receptor from outside the cell. The researchers also have plans to control the second type of channel, which could lead to devices that could sense biological, environmental or pharmaceutical agents.

"We shall replace the voltage sensitive... channel by ion channels that are sensitive to biological antagonists or to related pharmacological agents. In these ligand-gated channels the channel itself is coupled to a chemical receptor," said Fromherz.

For instance, one important ligand-gated channel is the glutamate receptor in the synapses of the brain, where glutamate molecules open a communications channel by binding to the membrane. This channel is the target of numerous pharmacological agents, said Fromherz.

The interface could be used to design pharmacological biosensors within five years, said Fromherz. Neurocomputers and neural implant devices are more than 20 years away, he said.

Fromherz's research colleagues were Bernhard Straub and Elisabeth Meyer of Max Plank Institute for Biochemistry. They published the research in the February, 2001 issue of *Nature Biotechnology*. The research was funded by the Max Plank Society and the German Federal Ministry of Education and Research.

Timeline: 5 years, > 20 years
Funding: Government, Private
TRN Categories: Human-Computer Interaction; Neural Networks;
Integrated Circuits
Story Type: News

Related Elements: Technical paper, "Recombinant Maxi-K Channels on Transistor, a Prototype of Iono-electronic Interfacing," Nature *Biotechnology*, February, 2001.

Nerve-Chip Link Closer

By Kimberly Patch, Technology Research News
December 5, 2001

Although human-machine hybrids are likely to remain in the realm of science fiction for decades, researchers are beginning to meld tissue and technology at the cellular level.

Bridging the wide communications gap between biology and electronics by connecting a cell to a semiconductor means both the cell and the electronic device can potentially take advantage of the best attributes of the other.

Researchers at the University of Texas at Austin have taken a step toward cell-semiconductor communications by soldering semiconductors to nerve cells. Key to the process is the solder—a modified peptide molecule that binds to a human neuron protein on one side, while the other sticks to a microscopic particle of semiconductor material, or nanocrystal.

"We make particles in a solution," and stick the modified peptides onto the surface of the particles, said Brian Korgel, an assistant professor of chemical engineering at the University of Texas at Austin. "The particles [then] stick to the cells in specific locations."

The specialized peptide molecule connects a particle to a biological receptor that sticks out of the nerve cell membrane. The connection brings the particle to within 20 nanometers of the electrically-active cell membrane, which is closer than previous methods. Nerve cells can be grown on semiconductor materials, but do not stick as closely to the electronics; previous research efforts have grown nerves on electronics, but leave a 50-nanometer gap.

As cells go, nerve cells are relatively large—the cells the researchers used were 60 microns in diameter, which is a little over 10 times the size of a red blood cell, and nearly the diameter of human hair. At 5 nanometers, the particles were more than 5,000 times smaller. A micron is one thousandth of a millimeter; a nanometer is one thousandth of a micron.

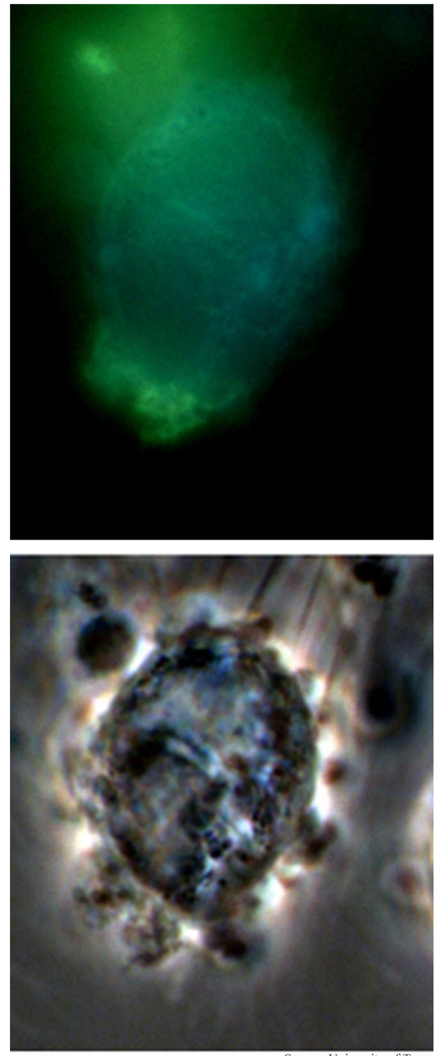
Although it may be difficult to get them together, nerve cells and electrical components do have something in common—they both communicate using electrical signals.

Nerve cells use changes in their electrical fields to create specific nerve firing patterns. "This is one of the underlying properties that enable brain functions like memory," said Korgel. Electrical field-effect transistors turn on and off based on the amount of electricity flowing through a gate electrode.

This mutual responsiveness to electrical fields makes intermaterial communications possible, said Korgel. "The nerve cell can effectively function as a gate on a field-effect transistor if the two materials are interfaced properly," he said.

Nerve-semiconductor links could eventually be used to allow nerves to directly control prosthetics, said Christine Schmidt, an assistant professor of biomedical engineering at the University of Texas at Austin.

"Bioprosthetic devices [like] retinal implants [and] mechanical prosthetics could be connected to the nervous system and brain using semiconductor materials such as those we are investigating." In addition, existing devices like cochlear implants may be improved using these materials, she said.



Source: University of Texas

These photos show nerve cells studded with semiconductor nanocrystals. The nanocrystals appear yellow in the photos.

Cells and nanocrystals could also be combined to detect tiny quantities of chemicals that are toxic to cells, said Korgel. There's also potential for using nerve cells to boost computer memory devices, he said. "One idea that I find particularly exciting is the prospect of combining nanocrystals, nerves and conventional microelectronics to create nerve-cell memory devices," he said.

It is also theoretically possible to use optically-activated nanocrystals to probe cells to study their internal electrochemical reactions, according to Korgel.

The researchers are currently looking into mechanisms that will allow the semiconductor nanocrystals to communicate with the nerve cells, said Korgel. "If we stimulate the nanocrystal with light... will the nerve feel the stimulus? Normally a nerve cell would not be affected by light, but with the nanocrystals attached, could we [change] the function of the nerve cells? These are the questions that we are currently trying to answer," he said.

The idea and methods are excellent, said Shuguang Zhang, associate director of the Center for Biomedical Engineering at the Massachusetts Institute of Technology. "Such direct linkage will likely find application in understanding the nerve connections and the strength of the connections through the fine adjustment of the electric input. This is a significant step forward to interface nerve cells with conducting and semiconducting materials," he said.

The method may eventually be useful in repairing damaged nerve systems; it could also serve to "interface the 40-year young semiconducting industry with biology that has evolved over billions of years. It is one step closer to... Star Trek," Zhang said. However, because dry computers and water-based cells are so inherently different "there still remains a big gap and challenge to be worked out," he said.

The researchers have extended the use of luminescent nanocrystals in biological applications, but it remains to be seen how useful the interface will be because the nanocrystals may still not be close enough to the membrane of the cell to interface with it electrically, said Peter Fromherz, a professor of biophysics at the Max Planck Institute of Biochemistry. "These particles are so far from the membrane that they feel little of the electrical field across the membrane," he said.

There are many hurdles to overcome before cells and semiconductor nanocrystals will combine in practical products, said Korgel. "This is really an unexplored area and we have much to learn," he said. Practical uses are probably a decade away, "but this is only a guess," he said.

Korgel and Schmidt's research colleagues were Jessica O. Winter and Timothy Y. Liu. They published the research in the October 30, 2001 issue of Advanced Materials. The research was funded by the National Science Foundation (NSF), the Welsh Foundation, DuPont, the Petroleum Research Fund, the Gilson Longenbaugh Foundation and the Whittaker Foundation.

Timeline: 10 years
Funding: Government, Corporate, Private
TRN Categories: Biological, Chemical, DNA and Molecular Computing;
Materials Science and Engineering; Semiconductors
Story Type: News
Related Elements: Technical paper, "Recognition Molecule Directed Interfacing between Semiconductor Quantum Dots and Nerve Cells," Advanced Materials, October 30, 2001.

Implant Links Nerve Cells to Electronics

By Kimberly Patch, Technology Research News
July 24/31, 2002

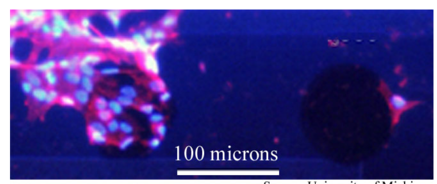
Although nerve cells and electronics work quite differently, scientists have for decades been trying to connect them.

Researchers from the University of Michigan have improved the connection by coating electrodes with a carefully controlled mix of plastic and protein.

The work is a step toward achieving finer control of prosthetic limbs, restoring damaged senses like sight and hearing, and providing direct connections between the brain and equipment like cameras.

Electronics and neurons both communicate using electrical pulses, but the physics involved in each is different. Electronics use currents of electrons, which carry a negative charge, while neurons use ions, which have an imbalance of electrons, and so can carry a negative or positive charge.

Electronics generate pulses by using energy to move electrons along metal wires. A nerve cell generates a pulse from one end of its oblong shape to the other by using energy to pump positively-charged sodium, potassium, or calcium ions across the membrane at one end, then allowing the ions to flow back all at once. Nerve cells range in length from 10 microns to 1 meter. A micron is one millionth of a meter.



Source: University of Michigan

The two black spots are electrodes. The left spot, which has a coating that encourages cell growth, is covered with nerve cells.

The researchers' electrode coating is a mix of a polymer that conducts electrical current and proteins that allow similar interactions with ions. The polymer also contains a substance that encourages neurons to grow around the metal electrode.

The key to the method is that the coating is not smooth, but has a fuzzy surface that increases the contact area between electrode and brain tissue, said David C. Martin, an associate professor of materials science and engineering, and biomedical engineering at the University of Michigan. "A large surface area can be packed into a small volume if the structure is fuzzy," he said.

The fuzzy structure "makes it possible to accommodate the dramatic difference in mechanical properties between the soft brain tissue and the hard silicon devices," Martin said. More surface area provides more places where neurons can bind to the electrode, gives neurons better access to the growth substance contained in the polymer, and makes it easier for electric charge to move across the interface, he said.

To test their method, the researchers coated electrodes a few millimeters long, 100 microns wide and 15 microns deep, and encouraged rat glial cells to grow on the devices. Glial cells provide support for neurons, which conduct the impulses of biological communications systems. A micron is one thousandth of a millimeter, and a human hair is about 75 microns in diameter. They also implanted the probes into living guinea pigs.

The fuzzy surface allowed electric current to flow more easily at the 1-kilohertz frequency that corresponds to the 1-millisecond width, or wavelength, of a neural pulse. The surface decreased the impedance, or resistance to electron flow by one or two orders of magnitude, according to Martin. “The lower the impedance at 1 kilohertz, the easier it is for electrical information to transfer from the probe to cells or vice versa,” he said.

As the coating gets thicker, its surface becomes more and more rough, increasing the surface area, but there’s also a point at which it becomes too thick and begins to increase resistance, said Martin. “There is an optimal film thickness at which the [charge] transport is the easiest,” he said.

The experiments also showed that the coatings encouraged neurons to bind to the electrodes. “We see neural cells attached to the surface of the probe after it is removed, whereas uncoated probes come out clean,” said Martin.

The researchers are currently working on coating electrodes with softer films of hydrogel materials that swell when they come in contact with water, and could increase the surface area further, according to Martin. They are also preparing to do tests with implanted probes in in rats.

The coatings could eventually be used for other types of devices, like pacemakers, that interface electronically with living tissue, Martin said.

Eventually the technique could be used to make new connections between electronics and tissue to restore sight or hearing, said Martin. “Another possible outcome may be the ability to control robotic equipment, or prosthetic... limbs.”

This type of interface could also allow for direct brain interfaces to cameras that detect things humans cannot, like infrared light or magnetic fields, Martin said.

Martin’s research colleagues were Yinghong Xiao, Junyan Yang, David Lin, Donghwan Kim, and Xinyan Cui. They presented the research at the 34th central regional meeting of the American Chemical Society on June 27 at Eastern Michigan University. The research was funded by the the National Institutes of Health (NIH) and the National Science Foundation (NSF).

Timeline: Unknown
Funding: Government
TRN Categories: Biotechnology; Materials Science and Engineering;
Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, “Surface Modification of Neural Recording Electrodes with Conducting Polymer/biomolecule Blends,” *Journal of Biomedical Materials Research*, August, 2001.

TRN

Index

TRN's Making The Future Report is published 12 times a year by Technology Research News, LLC. Each 20- to 40-page package assesses the state of research in a field like biochips, data storage or human-computer interaction.

Single reports are $300 to $500. A one-year subscription is $1,800. To buy a report or yearly subscription, go to www.trnmag.com/email.html.

We welcome comments of any type at feedback@trnmag.com. For questions about subscriptions, email mtfsubs@trnmag.com or call (617) 325-4940.

Technology Research News is an independent publisher and news service dedicated to covering technology research developments in university, government and corporate laboratories.

Kimberly Patch
Editor
kpatch@trnmag.com

Eric Smalley
Editor
esmalley@trnmag.com

Ted Smalley Bowen
Contributing Editor
tbowen@trnmag.com

Chhavi Sachdev
Contributing Writer
csachdev@trnmag.com