

Feeling Sounds, Hearing Sights

 cacm.acm.org/magazines/2018/1/223884-feeling-sounds-hearing-sights/fulltext

By Gregory Mone

Communications of the ACM, Vol. 61 No. 1, Pages 15-17

10.1145/3157075

Comments

Chieko Asakawa, who is blind, uses the NavCog app, which she helped develop, to find her way on the campus of Carnegie Mellon University.

Credit: Pittsburgh Post-Gazette. All rights reserved.

In a 2016 video, Saqib Shaikh, a Microsoft Research software engineer, walks out of London's Clapham Station Underground stop, turns, and crosses a street, then stops suddenly when he hears an unexpected noise. Shaikh, who lost his sight when he was seven years old and walks with the aid of the standard white cane, reaches up and swipes the earpiece of his glasses.



The video then shifts to the view from his eyewear, a pair of smart glasses that capture high-quality still images and videos. That simple swipe instructed the glasses, an experimental prototype designed by a company called Pivothead, to snap a still photo. Microsoft software analyzed the picture, then translated the findings into auditory feedback. Through the smart glasses, which include a small speaker, Shaikh hears the results from an automated voice: "I think it's a man jumping in the air doing a trick on a skateboard."

The Pivothead smart glasses and Microsoft AI technology belong to a broader class of what have become known as sensory substitution technologies, apps and devices that collect visual, auditory, and in some cases haptic stimuli, and feed the information to the user through another sensory channel. While the utility of these devices has long been debated in the vision- and hearing-impaired communities, recent advances suggest that sensory substitution technologies are finally starting to deliver on their promise.

A Rich History

The first sensory substitution devices originated long before computing machines and smartphone apps. The white cane widely used by people who are blind or have low vision alerts users to the presence of obstacles through tactile feedback. Similarly, Braille is a way of converting visual text to felt or tactile text. But a major technological shift has resulted from the efforts of neuroscientist Paul Bach-y-Rita, who developed a prototype device that converts video into tactile feedback.

Today, sensory substitution devices come in a variety of forms. The Brain-Port, for example, translates visual information from a forehead-mounted camera into tactile feedback, delivering stimuli through 400 electrodes on a thumbprint-sized pad that users place on their tongue. Other aids include the vOICe, which translates camera scans of an environment into audible soundwaves, allowing users to hear obstacles they cannot see.

Although these devices use different approaches, they are capitalizing on the same general principle. "Most of the hard computing work is being done in the brain," explains experimental psychologist Michael Proulx of the University of Bath. "What we're doing is relying on the brain's ability to take information in any sensory format and make sense of it independent of that format."

Practical Uses

Neuroscientist David Eagleman and his colleagues at Neosensory, a Silicon Valley startup, are developing a new device, the Buzz, that translates ambient sounds such as sirens or smoke alarms into distinct patterns of vibrations that pulse through and across the device's eight motors. A smartphone microphone picks up the sound, then passes it through an app that mimics the role of the inner ear. One algorithm separates the sound into its component frequencies (as our own ears do) while others cancel out unrelated noise, such as the hum of an air conditioner. The app then transforms this change in frequencies over time into a pattern of vibrations that alters every 20 milliseconds, rolling through or pulsing on the Buzz. "With a siren, you feel it going back and forth on your wrist because there are different frequencies involved," Eagleman explains. "A knock on the door is easy. You feel the knock on your wrist."

The Buzz and its predecessor, a more robust wearable vest, are also designed to be affordable: the projected price of the wrist-worn version should be less than \$400. Cost is a major concern because sensory substitution devices are not reimbursed through health insurance in the U.S., and studies have found that people with disabilities often have lower rates of employment and income, and may not be able to afford technologies like the BrainPort, which retails for \$10,000. "For people with sensory disabilities, none of these technologies are covered" by insurance, says Deborah Cook, Washington Assistive Technology Act Program technical advisor and director of the Older Blind Independent Living Program at the University of Washington. "You can get a wheelchair paid for, but you can't get a new visual or auditory device reimbursed."

Cook also argues that many sensory substitution devices are too focused on navigation. But IBM computer scientist Chieko Asakawa believes there is still an unmet need in this space, and that such technologies have the potential to allow people who are blind to explore unfamiliar areas such as schools, train stations, airports, and more. "It's not fun if I go to a shopping mall by myself, for example," says Asakawa, who lost her sight at age 14. "If there are many people in the mall, it's very difficult to move around with the white cane."

Asakawa and her collaborators at IBM Tokyo and Carnegie Mellon University have developed a new system, NavCog, that deploys bluetooth beacons throughout interior spaces such as academic buildings and, in one case, a public shopping mall. The beacons

connect to a smartphone app, which guides the user via voice assistance. "In the mall," she explains, "I can find out which shop is next to me while I'm walking, such as a coffee shop on the left or a sushi restaurant on the right. That's useful information."

Siri's Shortcomings

Devices that help individuals enhance their productivity in the workplace are also critical. Computer scientist Matt Huenerfauth of the Linguistic and Assistive Technologies Laboratory at the Rochester Institute of Technology (RIT) is working with researchers from the National Technical Institute for the Deaf (NTID) to see if Automatic Speech Recognition (ASR) technology of the sort that powers Siri, Alexa, and Cortana could be used to generate captions in real time during meetings. Often, people who are deaf or hard of hearing either skip business meetings and wait for summaries from other attendees, or sit through the conferences and miss numerous side conversations. However, ASR technology is imperfect, and a real-time captioning system with errors in the text can be confusing. Huenerfauth's team is investigating whether highlighting words that the ASR is not confident it recognized correctly—using italicized fonts—will help users understand which fragments of a transcript they can trust.

Computer scientist Raja Kushalnagar of Gallaudet University, along with colleagues from the University of Rochester, Carnegie Mellon University, and the University of Michigan, has pursued a slightly different approach to the same problem. Instead of attempting to build an automated system, Kushalnagar and his collaborators are developing a crowdsourcing technology that transcribes lectures or business meetings into reliable captions with only a five-second delay. Professional stenographers can cost \$100 per hour, so they are not a viable option for most users. The group's system recruits multiple untrained, lower-paid workers to remotely transcribe brief portions of a given meeting or lecture in real time. Their technology, Legion Scribe, then applies an algorithm that merges the three fragmented transcriptions into one sensible, readable text for people who are deaf or hard of hearing.

Feedback Loop

The incorporation of remote human assistants is not entirely unique in the sensory substitution field. The University of Washington's Cook touts a new service, Aira, which connects blind users with remote human agents through a pair of smart glasses. The agents, who act as assistants, see through the glasses, but they can also tap into GPS data and other information. If the user enters a Target store, for example, Aira software immediately brings up a map of that store's layout to help the agent guide the individual. Individuals can engage the service for a variety of tasks; one Aira user tapped into the service at a family funeral because he didn't want to bother relatives and ask for their help.

"Helping people identify objects and texts is good," says Aira co-founder and CEO Suman Kanuganti, "but we want to help them gain experiences in life, doing things faster and more efficiently."

Kanuganti says there is an additional advantage to keeping human agents in the loop. Since the service launched in early 2017, roughly 3,600 hours' worth of interactions between agents and users have been stored in the cloud. Aira has begun training machine learning algorithms on all that video, audio, GPS, and other data to develop automatic tools that provide some of the same functionality that remote human agents deliver now, such as automatic facial recognition and reading.

While these technical advances are welcome, experimental psychologist Proulx, who tests and compares several sensory substitution devices, including the BrainPort, in his laboratory, believes the devices will also improve as researchers learn more about how the brain processes sensory information flowing in through different channels.

"There have been findings showing that as people learn to use these devices, the brain is starting to use the visual parts of the cortex to process the information it's receiving through sound or touch," Proulx says. "Over the next 10 years is probably when we'll see a nice feedback loop between improvements in the technology and greater findings about the human user and how the brain is actually functioning. As that comes together, I'm really excited to see what sort of advances we're able to make."

Author

Gregory Mone is a Boston-based science writer and the author, with Bill Nye, of *Jack and the Geniuses: At the Bottom of the World*.

©2018 ACM 0001-0782/18/1

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or fee. Request permission to publish from permissions@acm.org or fax (212) 869-0481.

The Digital Library is published by the Association for Computing Machinery.
Copyright © 2018 ACM, Inc.
