

Low-power special-purpose chip could make speech recognition ubiquitous in electronics

February 13 2017, by Larry Hardesty



Researchers at MIT's Microsystems Technology Laboratories have built a low-power chip specialized for automatic speech recognition. With power savings of 90 to 99 percent, it could make voice control practical for relatively simple electronic devices. Credit: Jose-Luis Olivares/MIT

The butt of jokes as little as 10 years ago, automatic speech recognition

is now on the verge of becoming people's chief means of interacting with their principal computing devices.

In anticipation of the age of voice-controlled electronics, MIT researchers have built a low-power [chip](#) specialized for automatic [speech recognition](#). Whereas a cellphone running speech-recognition software might require about 1 watt of power, the new chip requires between 0.2 and 10 milliwatts, depending on the number of words it has to recognize.

In a real-world application, that probably translates to a [power savings](#) of 90 to 99 percent, which could make voice control practical for relatively simple electronic devices. That includes power-constrained devices that have to harvest energy from their environments or go months between battery charges. Such devices form the technological backbone of what's called the "internet of things," or IoT, which refers to the idea that vehicles, appliances, civil-engineering structures, manufacturing equipment, and even livestock will soon have sensors that report information directly to networked servers, aiding with maintenance and the coordination of tasks.

"Speech input will become a natural interface for many wearable applications and intelligent devices," says Anantha Chandrakasan, the Vannevar Bush Professor of Electrical Engineering and Computer Science at MIT, whose group developed the new chip. "The miniaturization of these devices will require a different interface than touch or keyboard. It will be critical to embed the speech functionality locally to save system energy consumption compared to performing this operation in the cloud."

"I don't think that we really developed this technology for a particular application," adds Michael Price, who led the design of the chip as an MIT graduate student in [electrical engineering](#) and computer science and now works for chipmaker Analog Devices. "We have tried to put the

infrastructure in place to provide better trade-offs to a system designer than they would have had with previous technology, whether it was software or hardware acceleration."

Price, Chandrakasan, and Jim Glass, a senior research scientist at MIT's Computer Science and Artificial Intelligence Laboratory, described the new chip in a paper Price presented last week at the International Solid-State Circuits Conference.

The sleeper wakes

Today, the best-performing speech recognizers are, like many other state-of-the-art artificial-intelligence systems, based on neural networks, virtual networks of simple information processors roughly modeled on the human brain. Much of the new chip's circuitry is concerned with implementing speech-recognition networks as efficiently as possible.

But even the most power-efficient speech recognition system would quickly drain a device's battery if it ran without interruption. So the chip also includes a simpler "voice activity detection" circuit that monitors ambient noise to determine whether it might be speech. If the answer is yes, the chip fires up the larger, more complex speech-recognition circuit.

In fact, for experimental purposes, the researchers' chip had three different voice-activity-detection circuits, with different degrees of complexity and, consequently, different power demands. Which circuit is most power efficient depends on context, but in tests simulating a wide range of conditions, the most complex of the three circuits led to the greatest power savings for the system as a whole. Even though it consumed almost three times as much power as the simplest circuit, it generated far fewer false positives; the simpler circuits often chewed through their energy savings by spuriously activating the rest of the chip.

A typical [neural network](#) consists of thousands of processing "nodes" capable of only simple computations but densely connected to each other. In the type of network commonly used for voice recognition, the nodes are arranged into layers. Voice data are fed into the bottom layer of the network, whose nodes process and pass them to the nodes of the next layer, whose nodes process and pass them to the next layer, and so on. The output of the top layer indicates the probability that the voice data represents a particular speech sound.

A voice-recognition network is too big to fit in a chip's onboard memory, which is a problem because going off-chip for data is much more energy intensive than retrieving it from local stores. So the MIT researchers' design concentrates on minimizing the amount of data that the chip has to retrieve from off-chip memory.

Bandwidth management

A node in the middle of a neural network might receive data from a dozen other nodes and transmit data to another dozen. Each of those two dozen connections has an associated "weight," a number that indicates how prominently data sent across it should factor into the receiving node's computations. The first step in minimizing the new chip's memory bandwidth is to compress the weights associated with each node. The data are decompressed only after they're brought on-chip.

The chip also exploits the fact that, with speech recognition, wave upon wave of data must pass through the network. The incoming audio signal is split up into 10-millisecond increments, each of which must be evaluated separately. The MIT researchers' chip brings in a single node of the neural network at a time, but it passes the data from 32 consecutive 10-millisecond increments through it.

If a node has a dozen outputs, then the 32 passes result in 384 output

values, which the chip stores locally. Each of those must be coupled with 11 other values when fed to the next layer of nodes, and so on. So the chip ends up requiring a sizable onboard memory circuit for its intermediate computations. But it fetches only one compressed node from off-chip memory at a time, keeping its power requirements low.

Provided by Massachusetts Institute of Technology

Citation: Low-power special-purpose chip could make speech recognition ubiquitous in electronics (2017, February 13) retrieved 26 April 2024 from

<https://techxplore.com/news/2017-02-low-power-special-purpose-chip-speech-recognition.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.