

What AI Can Tell From Listening to You

Artificial intelligence promises new ways to analyze people's voices—and determine their emotions, physical health, whether they are falling asleep at the wheel and much more

By *John McCormick*

April 1, 2019 9:43 p.m. ET

Are you depressed? In danger of a heart attack? Dozing at the wheel of your car?

Artificial intelligence promises to figure that out—and more—by listening to your voice.

A range of businesses, health-care organizations and government agencies are exploring new systems that can analyze the human voice to determine a person's emotions, mental and physical health, and even height and weight.

The technology is already used by call centers to flag problems in conversations. Now doctors are testing it as a way to spot mental and physical ailments, and companies are starting to tap it to help them vet job applicants.

Making all this possible: increasingly powerful machine-learning methods. AI systems can measure tone, tempo and other voice characteristics and compare them with stored speech patterns that have been identified as happy, sad, mad or any number of other emotions. While the science of vocal analysis has been developing for decades, cheaper computing power and new AI tools such as Google's TensorFlow have made it possible to build more-ambitious projects.

The technology can get even more powerful when used in combination with computer vision in a field known as emotion AI or affective computing. For instance, a voice system in a car might be able to tell if a driver is yawning, while a visual system could see if the driver is nodding off.

Research firm [Gartner](#) Inc. thinks that emotion AI may even spread into consumer products. By 2022, Gartner predicts, 10% of personal devices, up from less than 1% now, will have emotion AI capabilities, such as wearables that are able to monitor a person's mental health and videogames that adapt to a user's mood.

- Read more at [WSJ.com/AIreport](https://www.wsj.com/AIreport)

- [Debate: AI's Effect on Jobs](#)
- [Analytical Tools for the Nonexpert](#)
- [Simulations Teach Caregivers Empathy](#)
- [The Card Game AI Can't Win—Yet](#)

But emotion AI must overcome a big hurdle before it goes mainstream: People are uncomfortable with it. In survey findings released last year, Gartner reported that 52% of more than 4,000 respondents in the U.S. and U.K. said they didn't want their facial expressions to be analyzed by AI. And 63% said they didn't want AI to be constantly listening to get to know them.

Consumers are also concerned about their privacy. About two-thirds, 65%, of the Gartner respondents believe AI will destroy their privacy.

“People are very skeptical in general about AI,” says Annette Zimmermann, a Gartner analyst who wrote the firm's report on emotion AI. “Talking about feelings in AI, I think that's even more personal and more reason for [people] to be skeptical.”

And the systems aren't perfect, says Ms. Zimmermann, with the best systems achieving little more than 85% accuracy.

“It's not exact. And we don't know whether we will ever be exact,” says Rita Singh, a speech scientist at Carnegie Mellon University. “But it's getting closer.”

With those caveats, here's a look at some of the areas where AI voice analysis is having an impact today and the ones it may transform tomorrow.

Treating Mental Illness

In the U.S., mental illnesses affect one in five adults, or 46.6 million people in 2017, according to the National Institute of Mental Health, which estimates only half of those needing treatment receive it. [Emerging voice technology](#) may be able to make problems easier to spot.

At the end of last year, CompanionMx Inc., a company spun off from the behavioral-analysis firm Cogito Corp., launched a mobile mental-health monitoring system called Companion. The tool was developed with funding from the Defense Advanced Research Projects Agency, the U.S. Department of Veterans Affairs and the National Institute of Mental Health.

With the CompanionMx system, patients who are being treated for depression, bipolar disorders and other conditions download an app and create audio logs on their smartphones. The patients are asked to regularly talk about how they're feeling, and the information is automatically transmitted to an AI for analysis.

Read more of our in-depth coverage of the ways in which professionals extract value from the interplay of AI and business.



- [Visit WSJ Pro Artificial Intelligence](#)
- [AI Helps Auto-Loan Company Handle Industry's Trickiest Turn](#)
- [Chief Artificial Intelligence Officers Enter the C-Suite](#)
- [Nevada Casino Resort Uses AI to Manage Water Consumption](#)
- [European Companies Deploy AI With a Continental Twist](#)

Using the emotion AI technology developed by Cogito, CompanionMX analyzes patients' voices along with certain behavioral data for changes in mood, affect or behavior. For instance, CompanionMX monitors smartphone activity to see if the patient is withdrawing from contact with others. Caregivers can then reach out if they see indications of a problem.

The National Institute of Mental Health funded a study of the app from May 2015 to August 2017.

“The results are very encouraging,” says David Ahern, co-principal investigator of the study and director of the digital behavioral health and informatics research program at the Brigham and Women's Hospital and Harvard Medical School.

Mr. Ahern says the app can work as an early-detection system for caregivers, a much-needed tool when many of those in need of treatment don't seek it out until their problems are acute.

Fighting Heart Disease

More than 600,000 Americans die annually of heart disease, according to the Centers for Disease Control and Prevention. Researchers are trying to use voice AI to spot warning signs—and get people help quickly.

The Mayo Clinic conducted a two-year study that ended in February 2017 to see if voice analysis was capable of detecting coronary-artery disease. Every person's voice has different frequencies that can be analyzed, explains Amir Lerman, director of the Cardiovascular Research Center at Mayo.

Mayo, in collaboration with voice-AI company Beyond Verbal, used machine learning to identify what it thought were the specific voice biomarkers that indicated coronary artery disease. The clinic then tested groups of people who were scheduled to get angiograms.

Everyone in the study recorded their voices on a smartphone app, and the recordings were analyzed by Beyond Verbal. The finding: Patients who had evidence of coronary-artery disease on their angiograms also had the voice biomarkers for the disease.

Dr. Lerman says Mayo is hoping to deploy the technology in the near future. “I think it’s just an amazing area that opens new doors into how we treat patients,” he says.

Keeping Drivers Awake

More than 800 Americans died falling asleep behind the wheel in 2015, according to October 2017 statistics from the National Highway Traffic Safety Administration, and more than 30,000 people were injured in crashes involving drowsy drivers.

Now, many major car companies and artificial-intelligence companies are designing AI that uses voice analysis, along with facial recognition, to assess the alertness and emotional state of a driver.

At last year’s Consumer Electronics Show, [Toyota Motor Corp.](#) TM +0.36% displayed its Concept-i demonstration vehicle, which can read facial expressions and voice tones. The car is equipped with an infrared camera on the steering column, a pair of 3-D sensors on the instrument panel and an onboard speech-recognition and conversation system.

The systems work together to assess the state of the driver. For instance, a sagging head, slumping posture and a sleepy or low voice (or the sound of a yawn) would indicate drowsiness.

If the system notices drowsiness, it will react to the problem.

For instance, the car’s voice assistant could engage in a conversation with the driver to improve his or her alertness level. And, over time, the conversation system will know which topics are most likely to engage the driver.

In September, two AI companies, Affectiva and [Nuance Communications](#) NUAN +0.18% Inc., said they would work together to put emotional intelligence into Nuance’s conversational automotive assistant, which can understand and respond to requests.

The assistant, Dragon Drive, can be found in its current form in more than 200 million cars with name plates such as Audi, BMW, Daimler, Fiat, Ford, GM, Hyundai and Toyota, according to

Nuance.

The new technology from Affectiva and Nuance will use cameras to detect facial expressions such as a smile and microphones to pick up vocal expressions such as anger. The company's algorithms then use deep learning, computer vision and speech technology to identify emotions and indicators of drowsiness.

If drivers exhibit signs that they're tired, the voice assistant can address them by saying something as simple as: "You seem tired. Do you want to pull over for a break?"

These technologies are still in development, but according to Nuance Chief Technology Officer Joe Petro, they could be on the road in just a couple of years.

Humanizing Call Centers

Despite the move by many companies to offshore their customer-service operations, there are 7,400 call centers in the U.S. employing more than three million people, according to Site Selection Group, a real-estate advisory service.

A number of these companies, including insurers [Humana Inc.](#) [HUM -1.13%](#) and [MetLife Inc.](#), [MET +0.32%](#) have deployed Cogito's AI software as a way to keep their agents sharp and customers happy.

The system analyzes conversations between agents and customers, tracking in real time the way they interact.

As calls come into a center, they are streamed to Cogito's system, which evaluates hundreds of data points—speech rate, tone and more. If agents are pausing before answering questions, it could indicate they're distracted. If customers raise their voices, it could be a sign of frustration.

When the Cogito system detects a possible issue with a call, it sends a notification in the form of an icon or short message to the staffer's screen. It is a suggestion that the agent recognize and acknowledge the caller's feelings.

The system's main goal, says Joshua Feast, Cogito's chief executive officer, is to coach the agents, to get them to be more confident, more engaged and more empathetic. "Learning to speak to different customers is a real skill," Mr. Feast says. "You're not born with it. You have to learn it."

Cogito says the accuracy of its call-center product varies by where it's used, such as a customer-service center, sales department or claims-management unit, and what behaviors it's monitoring in

each of those areas. Overall, the company says its product has an average accuracy rate of 82%. It says it validates the results by human reviews of call outcomes, customer feedback and machine-learning analysis.

MetLife deployed Cogito's system about 15 months ago in its customer-service center, according to Kristine Poznanski, the insurer's head of global customer solutions.

While the system provides customer-service reps instant feedback on calls and real-time coaching, it also shows managers the status of calls. The data allows the center's manager to monitor a call in progress or to spend time with an agent reviewing a call once it's ended.

Ms. Poznanski says that, since deploying the system, the call center has seen an increase of 10% in both its first-call resolution and net promoter scores, which track customer sentiment to understand how likely they are to recommend a brand.

Hiring the Right Candidates

More than 80% of business owners and managers say they have hired the wrong person, according to [Robert Half International](#) Inc. Often, the problem is that the new employee has difficulty fitting in with the corporate culture.

Voicesense is one of the speech-based AI systems that says it can make applicant screening more effective.

Employers upload video or audio interviews to Voicesense's cloud and the company's system analyzes 200 speech parameters, such as intonation and pace, says Yoav Degani, Voicesense's chief executive officer. The system builds a behavioral model of the applicant's temperament, ambition, dependability and creativity, among other characteristics.

An employer can then use the scores the system generates to tell if an applicant is a good match for a job. For instance, if an organization was looking to hire a salesperson, the system would identify as a possible match someone who was highly active and engaged in the conversation, says Mr. Degani. But he acknowledges that the company's models provide probabilities rather than certainty.

In terms of privacy safeguards, Mr. Degani says that Voicesense doesn't store any of the data and that its tool doesn't analyze the content of a conversation, only the speech patterns.

AdventHealth Orlando, part of the AdventHealth health-care system, is using another analysis system, HireVue, to help with its recruitment efforts. The organization, which operates eight hospitals across central Florida and employs more than 25,000 people, hires 8,000 people each year.

That means reviewing more than 350,000 applications, according to Karla Muniz, AdventHealth's human-resources director.

Candidates who meet basic job requirements are invited to take an online interview using HireVue. Its algorithm evaluates applicants' responses to interview questions, such as tone of voice and word clusters. It also incorporates visual analysis, examining very quick facial movements called microexpressions.

The information from these assessments is then matched against data points that correspond with each job. Applicants who score high for a position are called in for interviews.

Since using HireVue, AdventHealth has decreased the time it takes to fill a job to 36 days from 42, Ms. Muniz says

Fighting Fraud

Property and casualty fraud amounts to about \$30 billion each year, according to statistics posted by the Insurance Information Institute, an industry trade group.

Insurer Allianz-SP Slovakia, a subsidiary of [Allianz](#) ALIZF 0.58% group, handles claims using Nemesysco's voice-stress analysis technology. The tool picks up people's reactions to a set of scripted questions asked by the claims handler. The system looks for a combination of markers, such as tiny pauses when a person is speaking, that may indicate the speaker is providing false information, according to Allianz-SP Slovakia.

"The aim is to pay a claim without any problems immediately and to prevent any fraud-like exaggeration of a claim," says Jaroslava Zemanová, head of control and special activities at Allianz-SP Slovakia.

Allianz-SP Slovakia notes that the voice analysis isn't proof of any wrongdoing, and that it is just the first stage in detecting possible fraud. To reject a claim, the company's investigative team needs additional evidence. Still, the company says the system is saving it time and money.

Investigating Crimes

In some cases, voice analytics doesn't just produce information about someone's health or emotional state—but also about their appearance.

In 2014, the U.S. Coast Guard was trying to track down a person who had placed 28 false distress calls. The emergency responses to these calls cost an estimated \$500,000.

But it is more than the cost, says Marty Martinez, special agent in charge for the Chesapeake Region of the Coast Guard Investigative Service. “It drags away resources from mariners who are actually in distress.”

Coast Guard investigators had little to go on other than the recordings of the emergency calls. Then they went to see Ms. Singh at Carnegie Mellon University, who had been working on computer speech recognition.

Ms. Singh, with just the voice recording, was able to determine the hoax caller’s age, height and weight. The case is ongoing, says Mr. Martinez.

The technology has been used in about a dozen other cases, he adds. “It has helped us narrow down and focus our investigative effort,” he says.

How? The human voice, Ms. Singh explains, carries information that can be linked to the physical, physiological, demographic, medical, environmental and other characteristics of the speaker.

Researchers are uncovering those microsignatures and using them for profiling.

“I call it the science of profiling humans from their voice,” says Ms. Singh.

Ms. Singh admits the technology isn’t perfect. Age, for instance, isn’t exact: It can be predicted only to within a three-year range. But research is improving its accuracy and taking it into new areas.

Ms. Singh and her team recently demonstrated a system that could reconstruct 60% to 70% of a person’s face just from their voice, she says.

Ms. Singh says voice-analysis technology still has a long way to go, but its potential is enormous. “It would enable machines to understand humans a lot better than perhaps even humans can,” she says.

Mr. McCormick is deputy editor of WSJ Pro Artificial Intelligence in New York. He can be reached at john.mccormick@wsj.com.

Appeared in the April 2, 2019, print edition.