

ALEXA, DO I HAVE COVID-19?

Researchers are exploring ways to use people's voices to diagnose coronavirus infections, dementia, depression and much more. **By Emily Anthes**

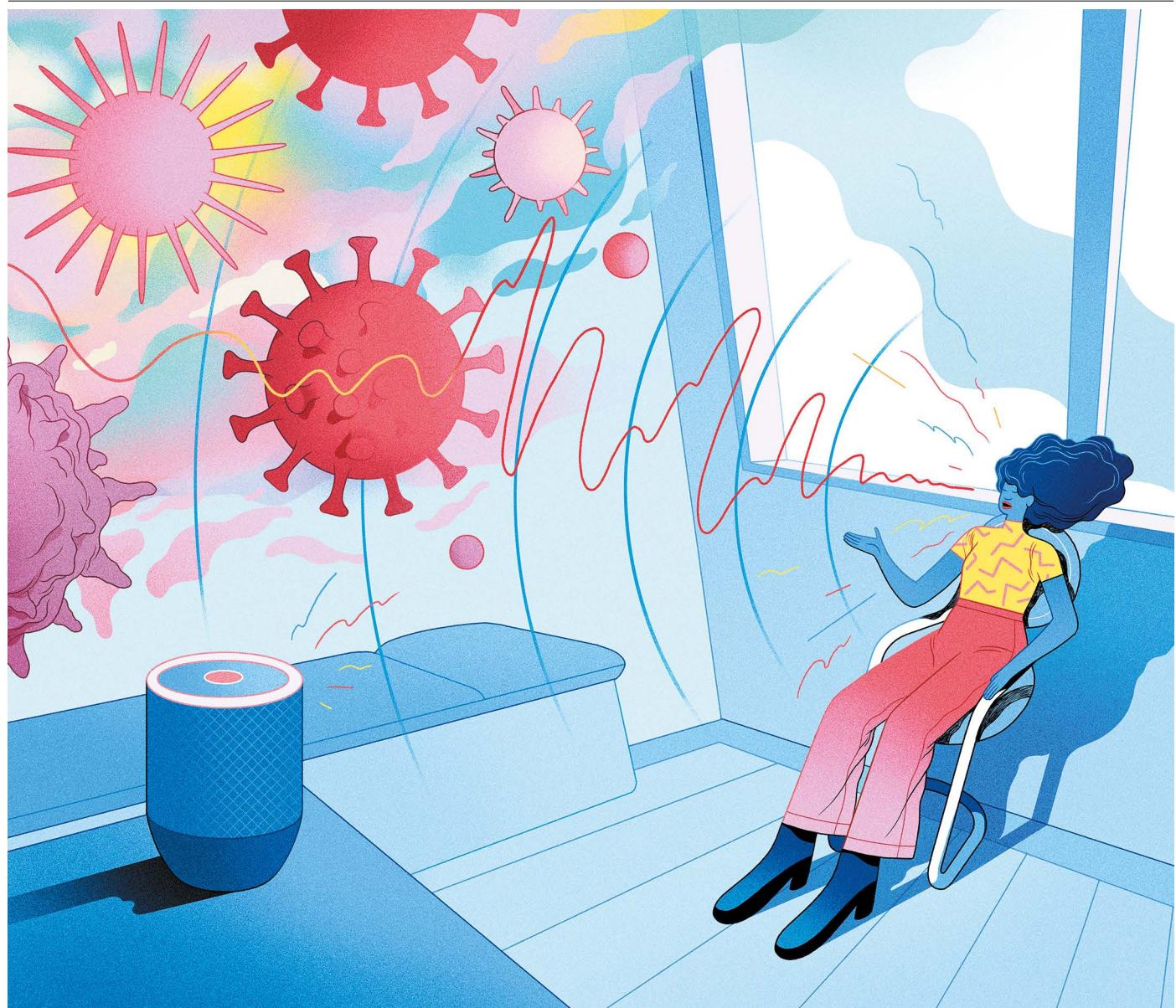


ILLUSTRATION BY RUNE FISKER

In March, as the staggering scope of the coronavirus pandemic started to become clear, officials around the world began enlisting the public to join in the fight. Hospitals asked local companies to donate face masks. Researchers called on people who had recovered from COVID-19 to donate their blood plasma. And in Israel, the defence ministry and a start-up company called Vocalis Health asked people to donate their voices.

Vocalis, a voice-analysis company with offices in Israel and the United States, had previously built a smartphone app that could detect flare-ups of chronic obstructive pulmonary disease by listening for signs that users were short of breath when speaking. The firm wanted to do the same thing with COVID-19. People who had tested positive for the coronavirus could participate simply by downloading a Vocalis research app. Once a day, they fired up the app and spoke into their phones, describing an image aloud and counting from 50 to 70.

Then Vocalis began processing these recordings with its machine-learning system, alongside the voices of people who had tested negative for the disease, in an attempt to identify a voiceprint for the illness. By mid-summer, the firm had more than 1,500 voice samples and a pilot version of a digital COVID-19 screening tool. The tool, which the company is currently testing around the world, is not intended to provide a definitive diagnosis, but to help clinicians triage potential cases, identifying people who might be most in need of testing, quarantine or in-person medical care. “Can we help with our AI algorithm?” asks Tal Wenderow, the president and chief executive of Vocalis. “This is not invasive, it’s not a drug, we’re not changing anything. All you need to do is speak.”

They’re not the only ones racing to find vocal biomarkers of COVID-19 – at least three other research groups are working on similar projects. Other teams are analysing audio recordings of COVID-19 coughs and developing voice-analysis algorithms designed to detect when someone is wearing a face mask.

It’s a sign of how hungry the young field of vocal diagnostics is to make its mark. Over the past decade, scientists have used artificial intelligence (AI) and machine-learning systems to identify potential vocal biomarkers of a wide variety of conditions, including dementia, depression, autism spectrum disorder and even heart disease. The technologies they have developed are capable of picking out subtle differences in how people with certain conditions speak, and companies around the world are beginning to commercialize them.

For now, most teams are taking a slow, step-wise approach, designing tailored tools for use in doctors’ offices or clinical trials. But many dream of deploying this technology

more widely, harnessing microphones that are ubiquitous in consumer products to identify diseases and disorders. These systems could one day allow epidemiologists to use smartphones to track the spread of disease, and turn smart speakers into in-home medical devices. “In the future, your robot, your Siri, your Alexa will simply say, ‘Oh you’ve got a cold,’” says Björn Schuller, a specialist in speech and emotion recognition with a joint position at the University of Augsburg in Germany and Imperial College London, who is leading one of the COVID-19 studies.

But automated vocal analysis is still a new field, and has a number of potential pitfalls, from erroneous diagnoses to the invasion of personal and medical privacy. Many studies remain small and preliminary, and moving from proof-of-concept to product won’t be easy. “We are at the early hour of this,” Schuller says.

Some ailments cause obvious vocal



THIS IS NOT INVASIVE, IT'S NOT A DRUG, WE'RE NOT CHANGING ANYTHING. ALL YOU NEED TO DO IS SPEAK."

distortions; consider the telltale stuffiness of someone afflicted by allergies. But many scientists think that vocal analysis could help to identify an enormous range of disorders, thanks to the complexity of human speech.

Speech signals

Speaking requires the coordination of numerous anatomical structures and systems. The lungs send air through the vocal cords, which produce sounds that are shaped by the tongue, lips and nasal cavities, among other structures. The brain, along with other parts of the nervous system, helps to regulate all these processes and determine the words someone is saying. A disease that affects any one of these systems might leave diagnostic clues in a person’s speech.

Machine learning has given scientists a way to detect aberrations, quickly and at scale. Researchers can now feed hundreds or thousands of voice samples into a computer to search for features that distinguish people with various medical conditions from those without them.

Much of the early work in the field focused on Parkinson’s disease, which has well-known

effects on speech – and for which there is no definitive diagnostic test. The disorder causes a variety of motor symptoms, including tremors, muscle stiffness and problems with balance and coordination. The loss of control extends to the muscles involved in speech; as a result, many people with Parkinson’s have weak, soft voices. “It’s one of those things that you can hear with the human ear,” says Reza Hosseini Ghomi, a neuropsychiatrist at EvergreenHealth in Kirkland, Washington, who has identified vocal features associated with several neurodegenerative diseases. “But if you can get 10,000 samples and a computer, you can get much more accurate.”

More than a decade ago, Max Little, a researcher in machine learning and signal processing now at the University of Birmingham, UK, began to investigate whether voice analysis might help doctors to make difficult diagnoses. In one study, Little and his colleagues used audio recordings of 43 adults, 33 of whom had Parkinson’s disease, saying the syllable “ahhh”. They used speech-processing algorithms to analyse 132 acoustic features of each recording, ultimately identifying 10 – including characteristics such as breathiness and tremulous oscillations in pitch and timbre – that seemed to be most predictive of Parkinson’s. Using just these 10 features, the system could identify the speech samples that came from people with the disease with nearly 99% accuracy¹.

Little and others in the field have also demonstrated that certain vocal features correlate with the severity of Parkinson’s symptoms. The systems are not yet robust enough for routine use in clinical practice, Little says, but there are many potential applications. Vocal analysis might provide a quick, low-cost way to monitor individuals who are at high risk of the disease; to screen large populations; or possibly even to create a telephone service that could remotely diagnose people who don’t have access to a neurologist. Patients could use the technology at home – in the form of a smartphone app, say – to track their own symptoms and monitor their response to medication. “This kind of technology can allow a high-speed snapshot, an almost continuous snapshot of how someone’s symptoms are changing,” Little says.

Researchers are now working to identify speech-based biomarkers for other kinds of neurodegenerative disease. For instance, a trio of scientists in Toronto, Canada, used voice samples and transcripts from more than 250 people to identify dozens of differences between the speech of people with possible or probable Alzheimer’s disease and that of people without it². Among the participants, those with Alzheimer’s tended to use shorter words, smaller vocabularies and more sentence fragments. They also repeated themselves and used a higher ratio

Feature

of pronouns, such as ‘it’ or ‘this’, to proper nouns. “It can be a sign that they’re just not remembering the names of things so they have to use pronouns instead,” says Frank Rudzicz, a computer scientist at the University of Toronto, who led the study.

When the system considered 35 of these vocal features together, it was able to identify people with Alzheimer’s with 82% accuracy. (This has since improved to roughly 92%, Rudzicz says, noting that the errors tend to be more or less evenly split between false negatives and false positives.) “Those features add up to sort of a fingerprint of dementia,” Rudzicz says. “It’s a very intricate hidden pattern that’s hard for us to see on the surface, but machine learning can pick it out, given enough data.”

Because some of these vocal changes occur in the early stages of neurodegenerative diseases, researchers hope that voice-analysis tools could eventually help clinicians to diagnose such conditions earlier and potentially intervene before other symptoms become obvious.

For now, however, this idea remains largely theoretical; scientists still need to do large, long-term, longitudinal trials to demonstrate that voice analysis can actually detect disease earlier than standard diagnostic methods can.

And some clinicians note that voice analysis alone will rarely yield definitive diagnoses. “I learn a lot by listening to someone’s voice,” says Norman Hogikyan, a laryngologist at the University of Michigan in Ann Arbor. “I do it for a living. But I put it together with a history and then my exam. All three parts of that assessment are important.”

Researchers in the field stress that the goal is not to replace doctors or create standalone diagnostic devices. Instead, they see voice analysis as a tool physicians can use to inform their decisions, as another ‘vital sign’ they can monitor or test they can order. “My vision is that collecting speech samples will become as common as a blood test,” says Isabel Trancoso, a researcher in spoken-language processing at the University of Lisbon.

Expanding applications

A number of voice-analysis start-ups – including Winterlight Labs, a Toronto firm co-founded by Rudzicz, and Aural Analytics in Scottsdale, Arizona – are now providing their software to pharmaceutical companies. Many are using the technology to help assess whether people enrolled in their clinical trials are responding to experimental treatments. “Using speech as a more subtle proxy for changes in neurological health, you can help push drugs across the finish line or at the very least identify those that are not promising early on,” says Visar Berisha, the co-founder and chief analytics officer at Aural Analytics.

Neurodegenerative disorders are just the beginning. Scientists have identified distinct speech patterns in children with neurodevelopmental disorders. In one small 2017 study, Schuller and his colleagues determined that algorithms that analysed the babbling of ten-month-old infants could identify with some accuracy which children would go on to be diagnosed with autism spectrum disorder³. The system correctly classified roughly 80% of children with autism and 70% of neurotypical children.

Researchers have also found that many children with attention deficit hyperactivity disorder speak louder and faster than their neurotypical peers, and show more signs of vocal strain. The firm PeakProfiling in Berlin is now developing a clinical speech-analysis tool that it hopes can help doctors to diagnose the condition.

But some clinicians are sceptical about how much useful information such systems will really provide. “Some of it is a little overblown,” says Rhea Paul, a specialist in communication disorders at Sacred Heart University in Fairfield, Connecticut. Children with neurodevelopmental disorders often have many easily observable behavioural symptoms, she notes.

Moreover, it’s not yet clear whether the algorithms are really identifying specific markers for, say, autism spectrum disorder, or just picking up on general signs of atypical brain development – or even just transient aberrations in speech. “Development is a meandering path and not every kid who starts out looking like they have autism grows up to be an adult with autism,” Paul says. Even if scientists do identify

a highly reliable, specific vocal biomarker, she adds, it should only be used to identify children who might benefit from a more thorough evaluation. “It shouldn’t be sufficient in and of itself to label a child, especially so early in life.”

Scientists are also turning the technology to mental illnesses. Numerous teams around the world have developed systems that can pick up on the slow, pause-heavy, monotonous speech that tends to characterize depression, and others have identified vocal biomarkers associated with psychosis, suicidality and bipolar disorder.

“Voice is enormously rich in terms of carrying our emotion signals,” says Charles Marmor, a psychiatrist at New York University. “The rate, the rhythm, the volume, the pitch, the prosody [stress and intonation] – those features, they tell you whether a patient is down and discouraged, whether they’re agitated and anxious, or whether they’re dysphoric and manic.”

In his own work, Marmor has used machine learning to identify 18 vocal features associated with post-traumatic stress disorder (PTSD) in 129 male military veterans. By analysing these features – which were mainly indicators of slow, flat, monotonous speech – the system could identify, with nearly 90% accuracy, which of the veterans had PTSD⁴.

Marmor and his colleagues are now expanding their research to women and civilians; if the team can generalize the findings, Marmor thinks that the technology could be a useful way to quickly identify people who might need a more thorough psychiatric assessment. “The first real-world application would be for

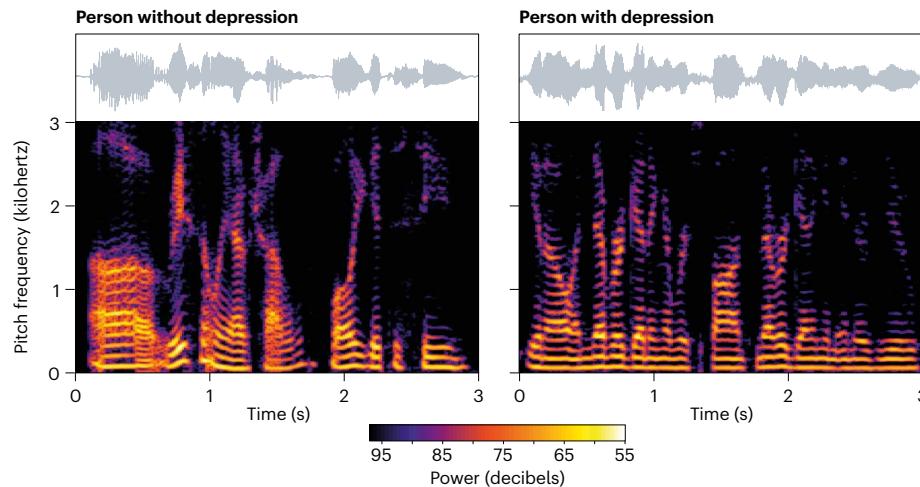


A man with Parkinson’s disease works on his speech with his wife. Vocal changes associated with the disorder could help doctors to diagnose it and assess treatments.

DON KELSEN/LOS ANGELES TIMES VIA GETTY

DEPRESSED TONES

Scientists and clinicians can identify differences between the voices of individuals with and without depression. Although the standard waveform of each person's speech (grey) shows few obvious distinctions, a spectrogram of speech intensity at different frequencies reveals less variation in pitch (frequency) and vocal energy (power) for the individual with depression. This matches subjective observations of flat and monotonous speech in depressed people.



high-throughput screening of PTSD," he says. "You can do 4,000 voice screens in a matter of hours."

Similar consumer applications are already beginning to make their way into the world. The US Department of Veterans Affairs is studying whether an app that monitors mental health can identify service members experiencing psychological distress. The smartphone app, developed by Cogito, a conversational guidance and analytics company in Boston, Massachusetts, collects metadata on users' habits – such as how frequently they call or text other people – and analyses voice memos they leave on their phones.

There might even be vocal biomarkers for conditions that seem to have nothing to do with speech. In one study from 2018, scientists analysing speech samples from 101 people who were scheduled to undergo coronary angiograms discovered that certain vocal frequency patterns were associated with more severe coronary artery disease⁵.

It's not clear what explains these differences. "We struggle with the mechanism because it's not obvious," says Amir Lerman, a cardiologist at the Mayo Clinic in Rochester, Minnesota, who led the research. Coronary artery disease could theoretically change the voice by reducing blood flow, he says. But it's also possible that it's not the disease itself that causes the vocal changes, but other associated risk factors, such as stress or depression.

Tricky translation

That study demonstrates both the promise and the limitations of this technology. It's one thing for a computer to pick out vocal patterns, but it's another, harder task to understand what they mean and whether they're clinically meaningful. Are they fundamental characteristics of the disease in question?



VOICE IS ENORMOUSLY RICH IN TERMS OF CARRYING OUR EMOTION SIGNALS: THE RATE, THE RHYTHM, THE VOLUME, THE PITCH.

Or merely markers of some other difference between groups, such as age, sex, body size, education or fatigue, any of which could be a confounding factor? "We're trying to move away from just shoving data into an algorithm, and really diving into the data sets, coming up with a model of the disease first and then testing that with machine learning," Ghomi says.

Most studies so far have identified potential biomarkers in just a small, single population of patients. "Reproducibility is still a question," Lerman says. "Is my voice today and tomorrow and the day after tomorrow the same?" To ensure that the results can be generalized – and to reduce the possibility of bias, a problem known to plague medical algorithms – researchers will need to test their classification systems in larger, more diverse samples and in a variety of languages. "We don't want to validate a speech model just with 300 patients," says Jim Schwoebel, vice-president of data and research at Sonde Health, a Boston-based voice-analysis company. "We think we need 10,000 or more."

The company runs SurveyLex, an online platform that allows researchers to easily create and

distribute voice surveys, as well as the Voiceome project, an effort to collect voice samples and health information from up to 100,000 people, across a wide variety of speech tasks, locations and accents. "You might be depressed in New York, and sound differently depressed in Houston, Texas," Schwoebel says.

For many of the applications that researchers have in mind, voice-analysis systems will have to not only distinguish sick people from healthy controls, but also differentiate between a variety of illnesses and conditions. And they'll need to do this outside the lab, in uncontrolled everyday situations and on a wide variety of consumer devices. "You've got smartphones which have a limited range of sensors, and people are using them everywhere in very uncontrolled environments," says Julien Epps, a researcher who studies speech-signal processing at the University of New South Wales in Sydney, Australia.

When Epps and his colleagues, including a researcher at Sonde Health, analysed voice samples recorded with high-quality microphones in a lab, they were able to detect depression with roughly 94% accuracy (see 'Depressed tones'). When using speech samples that people recorded in their own environments on their smartphones, the accuracy dropped to less than 75%, the researchers reported in a 2019 paper⁶.

And just because the technology is non-invasive doesn't mean that it is without risks. It poses serious privacy concerns, including the possibility that individuals could be identified from anonymous speech samples, that the systems might inadvertently capture private conversations, and that sensitive medical information could be sold, shared, hacked or misused.

If the technology isn't regulated properly, there's a danger that insurers or employers could use these systems to analyse speech samples without explicit consent or to obtain personal health information, and potentially discriminate against their customers or employees.

And then there's the perennial risk of false positives and overdiagnosis. "We have to be real and realize that a lot of this is still research," Rudzicz says. "And we need to start thinking about what's going to happen when we put it into practice."

Emily Anthes is a science journalist in New York City.

- Tsanas, T., Little, M. A., McSharry, P. E., Spielman, J. & Ramig, L. O. *IEEE Trans Biomed. Eng.* **59**, 1264–1271 (2012).
- Fraser, K. C., Meltzer, J. A. & Rudzicz, F. J. *Alzheimers Dis. Res.* **49**, 407–422 (2016).
- Pokorny, F. B. et al. *Proc. 18th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)* **2017**, 309–313 (2017).
- Marmar, C. R. et al. *Depress. Anxiety* **36**, 607–616 (2019).
- Maor, E. et al. *Mayo Clin. Proc.* **93**, 840–847 (2018).
- Huang, Z., Epps, J. & Joachim, D. *IEEE Trans. Affect. Comput.* <https://doi.org/10.1109/taffc.2019.2944380> (2019).