"HEY SMARTPHONE, AM I ILL?" DETECTING DISEASES FROM THE VOICE

Katrin D. Bartl-Pokorny [1,2], Adria Mallol-Ragolta [1,3], Anika A. Spiesberger [1,3], Anastasia Semertzidou [1], Johanna Löchner [4,5], Florian B. Pokorny [1,2,3*] and Björn W. Schuller [1,3,6,7]*

[1] CHI – Chair of Health Informatics, TUM University Hospital, Munich, Germany

[2] Division of Phoniatrics, Medical University of Graz, Graz, Austria

[3] MCML – Munich Center for Machine Learning, Munich, Germany

[4] Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital of Tübingen, Tübingen, Germany

[5] German Center for Mental Health (DZPG), Tübingen, Germany

[6] MDSI – Munich Data Science Institute, Munich, Germany

[7] GLAM – Group on Language, Audio, & Music, Imperial College London, London, United Kingdom

YOUNG REVIEWERS:

HRDAYA
AGE: 11

ZICHEN
AGE: 13

As humans, we learn from what we perceive with our senses in our daily lives. Computers can have similar learning capabilities, allowing them to learn from what they "see" and "hear" and to use the knowledge they learn to solve future tasks. This ability is called artificial intelligence (AI). Devices equipped with AI, such as smartphones, smartwatches, or smart speakers, have now become our everyday companions. Among other things, they can listen to us and answer our questions. This type of technology is also playing a growing role in medicine. In this article, we explain how a computer

can figure out whether the sound of a person's voice or the way they speak indicates a certain disease. We demonstrate this using the example of detecting COVID-19, and discuss both problems and opportunities that arise when using AI for diagnosis.

COMPUTER AUDITION

Have you ever heard your parent say, "Oh dear, this cough does not sound right!" when you have a cough and sore throat, for instance? Or maybe your parent realized you had a cold after just a short conversation? What this fascinating ability tells us is that your health condition is audible from the sound of your cough or your voice.

In the last few years, we have started to talk not only with humans but also with computers, such as smartphones, smartwatches, and smart speakers. We ask them, for example, to turn on the light or to give us information about the weather. Smart devices can hear and understand what we say, so, in the future, might they also be able to detect when we are ill, similar to or even better than the human ear can do? Could they potentially even detect the kind of disease a person has?

To do this, computers need the capability to learn and to use the learnt knowledge to solve future tasks. This capability is called **artificial intelligence** (AI). Using AI for sound data is called **computer audition** (CA).

For computers to learn, they usually require both example data and the information they should learn from the data—in our case, whether a disease is present (Figure 1). As humans, we learn similarly. We have heard the voices of healthy and ill people many times, so we have learnt how healthy and ill voices sound with regard to specific voice characteristics, such as roughness.

Doctors have used **auditory** information for many years to diagnose diseases. A well-known example is using a stethoscope—a metal disc with tubes connected to earpieces—to listen to lung and heart sounds. You might wonder why we even need the help of a computer. AI has several advantages over humans, which could make CA an important aid for **diagnosis**. Doctors see and hear only a limited number of patients with any specific disease during their professional lives. In contrast, a computer can learn from a huge amount of example data within just a few hours and without getting tired. Besides, a computer can safely store learnt knowledge and will not forget it over time.

In the future, using CA in medicine could reduce the need for medical examinations, which can be uncomfortable or even frightening for some patients. Additionally, it would be very convenient for people to

ARTIFICIAL INTELLIGENCE

Technology that gives computers the capability to solve tasks which were previously expected to be only solvable by humans.

COMPUTER AUDITION

The use of artificial intelligence to listen to and understand sounds.
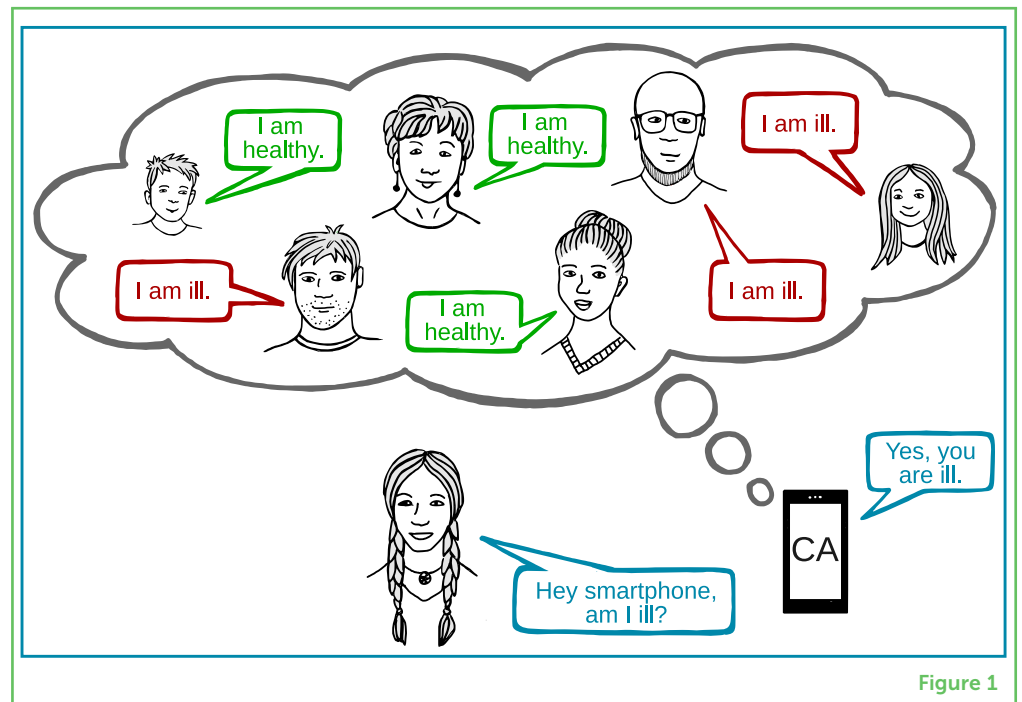
AUDITORY

Detectable by the sense of hearing.

DIAGNOSIS

A doctor's decision about which disease a person has, based on the reported symptoms and a medical examination.

Figure 1

just cough three times or speak a few words into their smartphone—or into any other microphone—from their bed at home, instead of visiting a doctor's office. With this data, a CA system could provide a doctor with the information needed to make a diagnosis.

But how can a computer find out whether the sound of a person's voice or the way they speak (e. g., emotional, slow) indicates a certain disease? We will explain this using the example of detecting COVID-19, one of the most widespread infectious diseases of the last few years, and discuss some problems and opportunities that arise when using AI for diagnosis.

VOICE RECORDING AND ANALYSIS

How can computers hear the human voice when they do not have ears? The microphones in our computers and smart devices act like human ears. They capture the human voice as it travels through the air. The voice recording is then saved on the device in a way that can be analyzed.

Everyone's voice has several characteristics that depend on the size, shape, and functionality of all body parts involved in the production of the voice, namely the lungs, throat, mouth, and nose. The unique combination of these characteristics in each individual is the reason why we usually know who is talking to us even if we cannot see the person, such as on the phone. Based on these characteristics, we can describe and compare the voices of different people. For instance, the voice of a young child likely sounds higher in pitch than the voice

of their parent. Characteristics called **acoustic features** are used to describe the voice in detail. Computers can extract hundreds or even thousands of acoustic features from a voice signal. In the example above, certain features that describe the pitch of a voice will have higher values for a young child than for their parent.

Computers can also generate a picture from a voice signal that combines the information from several acoustic features. This picture is called a **spectrogram** (Figure 2). A spectrogram shows how the voice sounds at which point in time, for example, in terms of pitch, roughness, and loudness. The spectrogram even reflects which speech sound, such as *a*, *e*, or *s*, is produced at a certain time and when exactly the speaker pauses. A CA system can use these acoustic features or spectrograms to learn. The latest systems can even learn directly from the audio recordings.
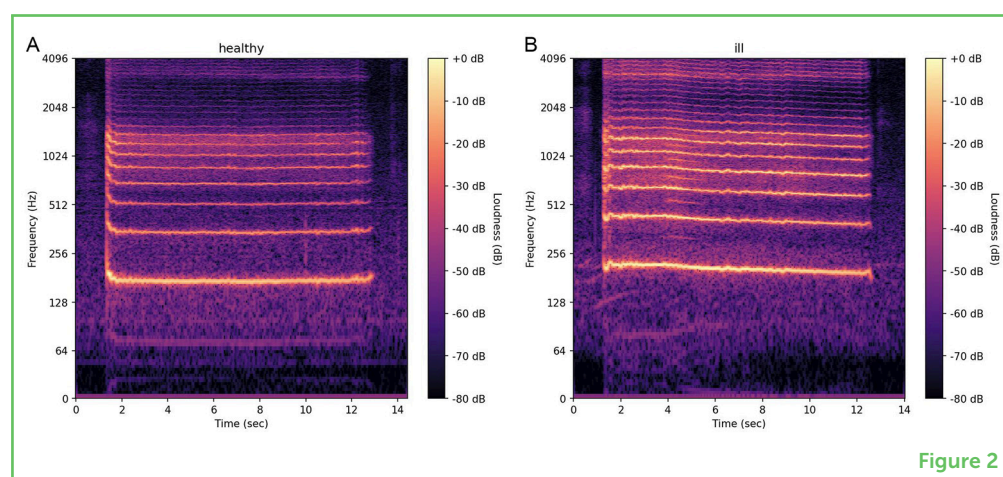
Figure 2

Spectrograms from **(A)** a healthy person and **(B)** an ill person show the loudness (measured in decibels; dB) for each frequency (measured in Hertz; Hz) present in the voice over time. Both spectrograms illustrate the production of the vowel *a* for about 12 s. The produced vowel is represented by the light parallel lines. The more the lines are blurred/distorted, or the more additional light components exist in between the parallel lines, the rougher the vowel sounds. Can you see the differences?

Figure 2

COVID-19 DETECTION WITH COMPUTER AUDITION

At the end of 2019, COVID-19 quickly spread worldwide, causing flu-like symptoms such as cough, shortness of breath, sore throat, headache, and fever. COVID-19 is usually diagnosed with tests that analyze liquids from the nose or throat. However, these tests can be uncomfortable, need to be bought in a drugstore, and produce waste, which is not good for the environment. COVID-19 symptoms can impact the patients' voice, breath, and cough sounds [1]. Therefore, CA could be a helpful, eco-friendly alternative to existing tests for COVID-19 and similar diseases [2].

Several CA studies have already aimed to detect COVID-19 [3]. To collect learning data for the CA systems, researchers asked participants with and without COVID-19 to produce various sounds: single vowels like *a*, *e*, or *o* produced as long as possible, text reading, taking deep breaths, or coughing into a microphone. The researchers used several CA systems to figure out which systems and sounds are best suited to

detect COVID-19. The studies yielded very different results. In some cases, the CA system correctly identified the COVID-19 status only in 6 out of 10 people, whereas other studies reported hardly any mistakes [3].

Although these results are very promising, we need to consider that some of the studies were based only on recordings of severely ill COVID-19 patients and healthy people without COVID-19. Therefore, more research is needed to build a CA system that is really capable of detecting whether a patient actually has COVID-19 or instead has some other illness, instead of simply telling us whether a patient is ill or healthy [4].

PROTECTING PRIVATE VOICE DATA

So, in the future, CA could become a valuable aid for diagnosing infectious diseases such as COVID-19 and other conditions that impact the voice. It could, for example, detect depression by identifying unusual sadness in a person's voice. Maybe simultaneous detection of multiple diseases in the same patient, such as COVID-19 *and* depression, will be possible as well, assuming that example data from people with multiple diseases are available for the CA system to learn from.

It should be clear to you now that, to build CA systems for disease detection, we first need to collect data. People with and without a disease of interest will need to provide voice, cough, or breath samples from which the CA system can learn. Everyone who agrees to participate in a study must sign a document in which all planned investigations and the use of the data are explained. This document is called an **informed consent** form. The human voice is a highly sensitive type of data, because other people, like family, friends, colleagues, teachers, and neighbors, can recognize a person from their voice. Since the recording might contain information about a person's health, feelings, or attitudes, it is very important to handle that data in a secure and responsible way, to protect the privacy of the participants.

INFORMED CONSENT

First, study participants need to be informed about all planned measurements and the intended use of their data. Then, they formally express their agreement by signing a respective document.

CA works best if it not only learns from many recordings in total but also from recordings of people of different ages, genders, languages, professions, and family backgrounds. Imagine there are two types of CA systems. System 1 has learnt only from recordings of female adults, whereas System 2 has learnt from a similar total number of recordings, but the data included recordings of both female and male adults and children. If both systems decide whether a 9-year-old boy has a certain disease, System 2 has a better chance of correctly identifying the child's health status, as System 1 has never learnt how the voice of a 9-year-old boy sounds—neither healthy nor ill.

Even if we provide a CA system with an adequate amount of diverse data to learn from, its results will not be 100% accurate. Occasionally, the CA system will predict a certain disease despite the person being healthy or having a different disease, or it will fail to detect a disease that is actually present. Thus, doctors should be aware that AI—similar to all other tools they use for diagnosis—can be wrong. CA should be used as one of several tools helping them to make a diagnosis. So, AI will not replace doctors. Instead, the benefits of such technologies lie in providing doctors with improved information that can potentially increase the accuracy of their diagnoses, reduce incorrect diagnoses, help them to diagnose diseases earlier, and reduce inconvenient examinations for patients. Overall, using CA and other forms of AI as part of the diagnostic "toolbox" could improve patients' care and their health outcomes.

ACKNOWLEDGMENTS

REFERENCES

1. Bartl-Pokorny, K. D., Pokorny, F. B., Batliner, A., Amiriparian, S., Semertzidou, A., Eyben, F., et al. 2021. The voice of COVID-19: acoustic correlates of infection in sustained vowels. *J. Acoust. Soc. Am.* 149:4377–83. doi: 10.1121/10.0005194
2. Schuller, B. W., Schuller, D. M., Qian, K., Liu, J., Zheng, H., and Li, X. 2021. COVID-19 and computer audition: an overview on what speech & sound analysis could contribute in the SARS-COV-2 corona crisis. *Front. Digit. Health* 3:564906. doi: 10.3389/fdgth.2021.564906
3. Kosar, A., Asif, M., Ahmad, M. B., Akram, W., Mahmood, K., and Kumari, S. 2024. Towards classification and comprehensive analysis of AI-based COVID-19 diagnostic techniques: a survey. *Artif. Intell. Med.* 151:102858. doi: 10.1016/j.artmed.2024.102858
4. Coppock, H., Jones, L., Kiskin, I., and Schuller, B. 2021. COVID-19 detection from audio: Seven grains of salt. *Lancet Digit. Health* 3:e537–8. doi: 10.1016/S2589-7500(21)00141-2

Detecting Diseases From The Voice. Front. Young Minds 13:1509215. doi: 10.3389/frym.2025.1509215

YOUNG REVIEWERS

HRDAYA, AGE: 11
I am an avid book reader, and I am fond of writing critical reviews of books that I read. English and Science are my favorite subjects. I like to play chess when I am not reading. I am also curious to know more about things around me. I am an animal lover, and I love to play with my pet birds and feed them.

ZICHEN, AGE: 13
I am currently 13 years old. I enjoy reading and playing interscholastic sports. I am curious and like to learn. I am also volunteering in a non-profit organization that helps kids with autism.

AUTHORS

KATRIN D. BARTL-POKORNY
I am a senior lecturer at the Medical University of Graz, Austria, and a postdoctoral researcher at the TUM University Hospital, Germany. I studied applied linguistics and completed my Ph.D. in medical science in Graz, Austria. I like to do research on the language and communicative development of children and on diseases that impact speech. My favorite part of the job is generating new research ideas and discussing research findings with my colleagues. *katrin.bartl-pokorny@tum.de

ADRIA MALLOL-RAGOLTA
I received my bachelor's degree in audiovisual telecommunication systems engineering from the Universitat Pompeu Fabra in Barcelona, Spain, and my master's degree in electrical engineering from the University of Colorado in Colorado Springs, USA. Currently, I am a doctoral student at the Technical University of Munich, Germany, conducting research on the development of AI algorithms for the detection and recognition of diseases using wearable-based data.

ANIKA A. SPIESBERGER

I hold a bachelor's in psychology from Ludwig-Maximilians-University Munich, Germany, and a master's in human factors engineering from the Technical University of Munich (TUM), Germany. Since 2022, I have been pursuing my doctoral studies. Currently, I work as a research assistant at TUM University Hospital's Chair of Health Informatics. My research is primarily focused on enhancing accessibility and equity in healthcare, with a particular emphasis on addressing the needs of understudied diseases and marginalized groups.

ANASTASIA SEMERTZIDOU

I have been a research assistant since 2012 and currently work at the Chair of Health Informatics at the TUM University Hospital, Germany. My academic background includes a bachelor's degree in informatics focusing on digital media and a master's degree in medical informatics, both of which were attained from the Aristotle University of Thessaloniki, Greece. My research is centered around medical data acquisition and analysis, especially of brain signals, based on studies of elderly people and children.

JOHANNA LÖCHNER

I am a psychologist and assistant professor working at the Department of Child and Adolescent Psychiatry at the University of Tübingen, Germany. I hold a doctorate from the Ludwig-Maximilians-University Munich, Germany, where I focused on the prevention of depression in children and adolescents. Ever since, my research has focused on how children, adolescents, and parents may be empowered to cope with challenges. In addition, I work with patients with mental health disorders, e.g., depression, as a psychotherapist.

FLORIAN B. POKORNY

I graduated from the Graz University of Technology, Austria, in electrical engineering and audio engineering and received my doctoral degree in electrical and computer engineering at the Technical University of Munich, Germany. Currently, I am a senior scientist at the Medical University of Graz, Austria, and a postdoctoral researcher at the TUM University Hospital, Germany. My main research interest is in the early detection of diseases in babies on the basis of babies' vocal sounds. *florian.pokorny@tum.de

BJÖRN W. SCHULLER

I received my doctoral degree from the Technical University of Munich (TUM), Germany, in 2006. After stops at the University of Passau and the University of Augsburg, both in Germany, I am now a professor of health informatics at the TUM University Hospital and artificial intelligence at the Imperial College London, UK. My main research interests are machine learning, speech processing, and emotion recognition from audio. I am also the co-founder of audEERING, a company specializing in intelligent audio analysis.