This article has been accepted for publication in IEEE Pervasive Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MPRV.2024.3429779

Dimitris Spathis dimitrios.spathis@nokia-bell-labs.com
FROM THE EDITOR The new "Research Brief" section in IEEE Pervasive Computing aims to rethink science communication through ultra-compact papers. In the inaugural article of the series, Prof. Cecilia Mascolo discusses the state of the art in AI-powered audio for health drawing on examples from her research group's work.

Sounding out the Power of Audio for Health

Cecilia Mascolo cm542@cam.ac.uk
Professor, Department of Computer Science and Technology, University of Cambridge

The pervasive computing research community has successfully explored and deployed techniques that use sensors in wearable devices and smartphones to understand our behaviour, fitness, and health. Examples are sensors such as inertial measurement units and photoplethysmography that are now incorporated in many existing commercial wearables (e.g., smartwatches or bands) to gather data about our movement, gait, and vital signs, such as heart rate and respiration rate.

A sensor which has perhaps had less use in wearables for behaviour and health monitoring is the microphone. Microphones are cheap sensors which are incorporated in many wearable devices already for the purpose of allowing communication or music listening. However, their ability to shed light on our health has only reasonably recently started to be explored in conjunction with wearable devices.

The power of audio for the detection and monitoring of disease is long known. A glaring example of its widespread use is the stethoscope, used for centuries to auscultate our cardiac and respiratory health. It is also not only body sounds which have diagnostic power. More recently the research community has linked voice biomarkers to conditions such as Parkinson's disease, post-traumatic stress disorder and even cardiovascular health.

The pervasive research community has started to explore the power of audio in various ways. Examples include the use of smartphone microphones as a source of information to contextualize the environment, or to understand emotions from the voice recorded on a microphone on a phone. In other research, the phone microphone detects sounds emitted while the user is sleeping and the analysis is able to signal sleep apnea episodes. Smartphones and wearables have also used sound to detect asthma symptoms and wheezing. Sound has also been explored in a contactless fashion where Smart Speaker devices would record users' voices and explore their health. All the above examples have investigated the use of audio in a passive fashion, meaning that the microphone would record the sounds produced in their environment. Some approaches have used human inaudible sounds instead in an active fashion: the device would use a speaker to produce a sound and a microphone would record the results of the sound bouncing around in an environment with diagnostics purposes.

While not as extensive as the existing image datasets, audio-based datasets do exist which can be used to train machine learning algorithms. Examples include heart sound data collected with both digital stethoscopes and smartphones, as well as smaller-scale datasets of asthma sounds or sleep sounds collected with external microphones.

This article has been accepted for publication in IEEE Pervasive Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MPRV.2024.3429779

I will now describe two example projects we have run which have used the microphones of wearable devices in different ways.

COVID-19 Sounds

During the COVID-19 pandemic, we developed a smartphone-based app to collect users' data about their medical history, current symptoms, and current COVID-19 test results, as well as coughs, breathing sounds, and voice while reading a sentence on the smartphone's screen [1]. We collected a total of 75,200 recordings from about 36,364 participants. The data was released to the academic community under controlled access (due to the sensitivity of the audio sounds which could reveal individuals' identity) and has been downloaded by more than 600 institutions to date. Algorithms and machine learning models were developed to show that audio could, at the time at which a) COVID-19 was the only circulating respiratory disease and b) was a lower respiratory infection disease, detect the disease reasonably accurately. Notably, the app asked participants to return to it to repeat the recordings every couple of days, and a few hundred participants have given more than two recordings. While the value of the one-off prediction is high, a much more interesting and perhaps more realistically accurate task is the prediction of the longitudinal progression of a disease in an individual. Unfortunately, there is a general lack of longitudinal audio respiratory data to explore the power of audio in terms of disease progression. We are in the process of collecting more of this data in the context of general respiratory infections.

In-ear Microphones for Vital Sign Monitoring and Activity

Some earpods such as Airpods Pro have inward-facing microphones, mainly for noise cancelling. However, the insulating component placed at the entrance of the ear canal to block the ear also generates occlusion which helps emphasize certain sounds related to foot strikes as well as heartbeats and respiration, enabling new ways of monitoring activity and well-being. We have explored this in the context of activity recognition, and heart rate detection [2] and are currently looking into respiratory rate detection, further cardiac health parameters, as well as running performance.

Together these studies demonstrate the potential of audio in the context of health diagnostics, progression, and fitness monitoring. They also attest to its versatility. However, working with this signal in wearables comes with challenges:

Longitudinal studies: there are very few data collections which follow users and have more than one sample per individual: as mentioned above, longitudinal data would allow the audio models to leverage the user baseline and learn more useful details of the changes in a user's health. This is a very promising avenue for audio and wearables.

Privacy: audio is prone to be considered a very private input. While some studies have considered passive sensing via audio, many have resorted to active audio sampling when the users actively record themselves (like in the COVID-19 study described above). Aids for more private passive sensing however exist that confine the data on the device or preprocess to avoid raw data being sent to the cloud. Generally, on-device machine learning solutions for this sort of input are, in my opinion, an excellent match.

Scarce labels: even if large audio datasets could be (and are) collected, the issue of scarce labels would hinder model development. As for many wearable devices or health studies, the generation of meaningful labels is time-consuming and sometimes hard. Very recently, unsupervised or semi-supervised models have started to emerge, including foundation models specific to this type of input [3].

In summary, audio is a rich input signal for health diagnostics and it couples well with wearable devices as microphones (and speakers) are generally cheap. The challenges that the use of audio imposes are substantial but, at the same time, offer interesting opportunities for this research community.

References

[1] https://www.covid19-sounds.org
[2] K. Butkow, T. Dang, A. Ferlini, D. Ma, Y. Liu, C. Mascolo. An evaluation of heart rate monitoring with in-ear microphones under motion. In Pervasive and Mobile Computing. 2024
[3] Zhang, Y., Xia, T., Han, J., Wu, Y., Rizos, G., Liu, Y., ... & Mascolo, C. (2024). Towards Open Respiratory Acoustic Foundation Models: Pretraining and Benchmarking. arXiv preprint arXiv:2406.16148.

Cecilia Mascolo is a Professor of Mobile Systems in the Department of Computer Science and Technology, University of Cambridge, UK. She is director of the Centre for Mobile, Wearable Systems and Augmented Intelligence. She is also a Fellow of Jesus College Cambridge and the recipient of an ERC Advanced Research Grant. Prior to joining Cambridge in 2008, she was a faculty member in the Department of Computer Science at University College London. She holds a PhD from the University of Bologna. Her research interests are in mobile systems and machine learning for mobile and wearable health. She has published in several top-tier conferences and journals in the area and her investigator experience spans projects funded by Research Councils and industry. She has served as steering, organizing and programme committee member of mobile and sensor systems, signal processing and machine learning conferences.