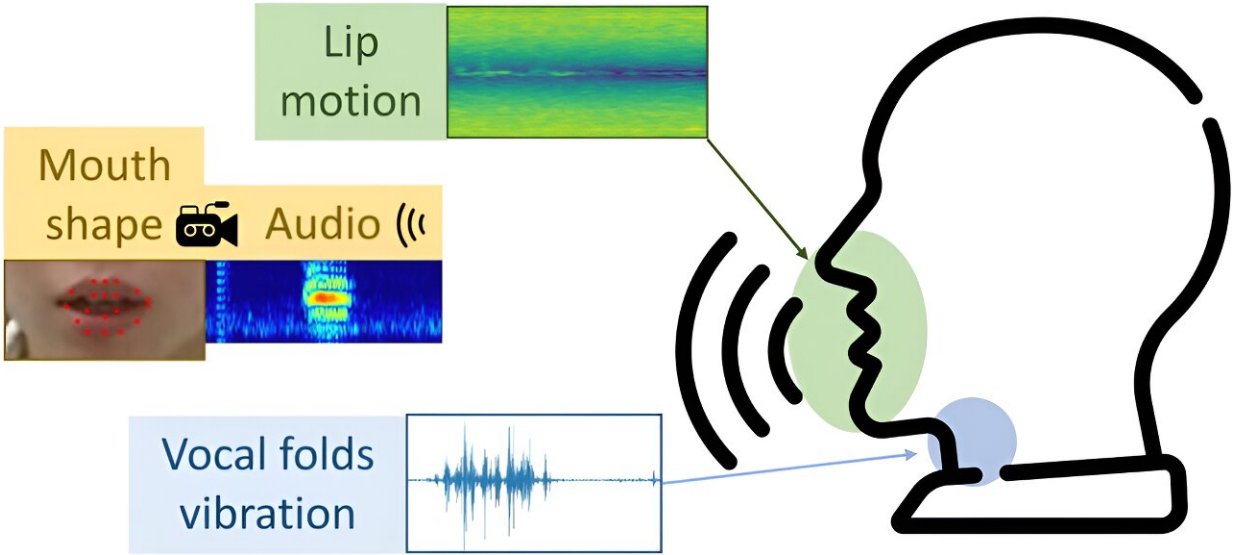


Comprehensive dataset for contactless lip reading and acoustic analysis holds potential for speech recognition tech

December 14 2023



Schematic diagram of the multimodal signals. Credit: *Scientific Data* (2023). DOI: 10.1038/s41597-023-02793-w

Sophisticated new analysis of the physical processes which create the sounds of speech could help empower people with speech impairments and create new applications for voice recognition technologies, researchers say.

Engineers and physicists from the University of Glasgow led the

research, which scrutinized the internal and external muscle movements of volunteers as they talked using a wide spectrum of wireless sensing devices.

The Glasgow team is making the data from 400 minutes of analysis freely available to other researchers to aid the development of new technologies based on [speech](#) recognition.

Those future technologies could help people suffering from speech impairment or voice loss by using sensors to read their lips and facial movements and provide them with a synthesized voice.

The dataset could enable voice-controlled devices like smartphones to read users' lips as they speak silently, enabling silent speech recognition. It could also help to enhance [voice recognition](#), improving the quality of video and [phone calls](#) in noisy environments.

It could even help improve security for banking or confidential transactions by analyzing users' unique facial movements like a fingerprint before unlocking sensitive stored information.

The researchers discuss how they conducted their multi-modal analysis of speech formation in a paper titled "[A comprehensive multimodal dataset for contactless lip reading and acoustic analysis](#)," in the journal *Scientific Data*.

To gather their data, the researchers asked 20 volunteers to speak a series of vowel sounds, single words and entire sentences while complex scans of their facial movements and recordings of their voices were collected.

The team used two different radar technologies—in the protocol of impulse radio ultra wideband (IR-UWB) and frequency modulated

continuous wave (FMCW) to image the movement of the volunteers' facial skin as they spoke, along with the movements of their tongue and larynx.

Meanwhile, vibrations on the surface of their skin was scanned with a laser speckle detection system, which used a [high speed camera](#) to capture the vibration of emitted laser speckle. A separate Kinect V2 camera capable of measuring depth read the deformations of their mouths as they shaped different sounds.

The University of Glasgow researchers collaborated with colleagues at the University of Dundee and University College London to synchronize and compile the dataset, which they call RVTALL for the [radio frequency](#), visual, text, audio, laser and lip landmark information it contains.

The data was validated by [signal processing](#) and machine learning techniques, building a uniquely detailed picture of the physical mechanisms which allow people to form sounds.

Professor Qammer Abbasi, of the University of Glasgow's James Watt School of Engineering, is the paper's corresponding author. Professor Abbasi has previously led research on speech recognition which used multimodal sensing to read lip movements through masks.

Professor Abbasi said, "This type of multimodal sensing for speech recognition is still a relatively new field of research, and our review of existing public data found that there wasn't much available to help support future developments.

"What we set out to do in collecting the RVTALL dataset was create a much more complete set of analyses of the visible and invisible processes which create speech to enable new research breakthroughs,

and we're pleased that we're now able to share it."

Professor Muhammad Imran, leader of the University of Glasgow's Communications, Sensing and Imaging hub, is a co-author of the paper. He said, "Contactless sensing has huge potential for improving [speech recognition](#) and creating [new applications](#) in communications, health care and digital security.

"We're keen to explore in our own research group here at the University of Glasgow how we can build on previous breakthroughs in lip-reading using multi-modal sensors and find new uses everywhere from homes to hospitals."

More information: Yao Ge et al, A comprehensive multimodal dataset for contactless lip reading and acoustic analysis, *Scientific Data* (2023). [DOI: 10.1038/s41597-023-02793-w](https://doi.org/10.1038/s41597-023-02793-w)

Provided by University of Glasgow

Citation: Comprehensive dataset for contactless lip reading and acoustic analysis holds potential for speech recognition tech (2023, December 14) retrieved 27 April 2024 from <https://techxplore.com/news/2023-12-comprehensive-dataset-contactless-lip-acoustic.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.