University of Nottingham
UK | CHINA | MALAYSIA

The COG-MHEAR project

Towards cognitively-inspired, 5G-IoT enabled, multi-modal Hearing Aids

Some highlights selected by
Michael.Akeroyd@nottingham.ac.uk

Why? (& who pays?)

Engineering and Physical Sciences Research Council

Transformative Technologies Theme
2021 – 2025
EP/T021063/1

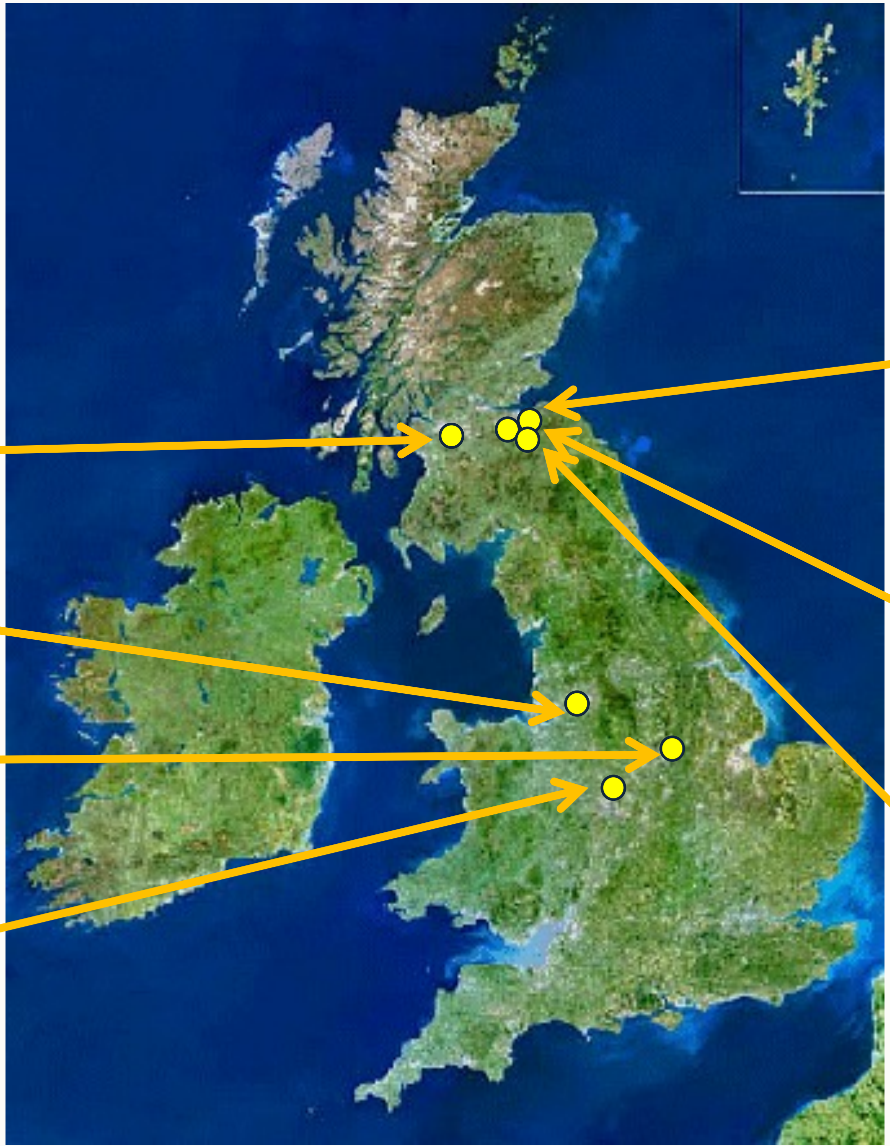
Our overall aim is to create "multi-modal” (**MM**) aids which not only amplify sounds but contextually use simultaneously collected information from a range of sensors to improve speech intelligibility.

Who? Where?

Colour coding …

- Red = CompSci, AI, IoT, HCI, signals
- Blue = wireless, 5G, flexible electronics
- Black =speech, hearing, neurobiology

+ experts, user-groups and external board
(inc. Peter Derleth, Sonova, John Hansen, Dallas)



Qammar Abbasi
Muhammad Ali Imran

Glasgow

Alex Casson

Manchester

Michael Akeroyd

Nottingham

Ahsan Adeel

Wolverhampton

Amir Hussain (PI)
Emma Hart
Ahmed Al-Dubai
William Buchanan

Edinburgh Napier

Peter Bell
Steve Renals
Tughrul Arlsan
Tharmalingam Ratnarajah

Edinburgh

Lynne Baillie
Mathini Sellathurai

Heriot-Watt

What?

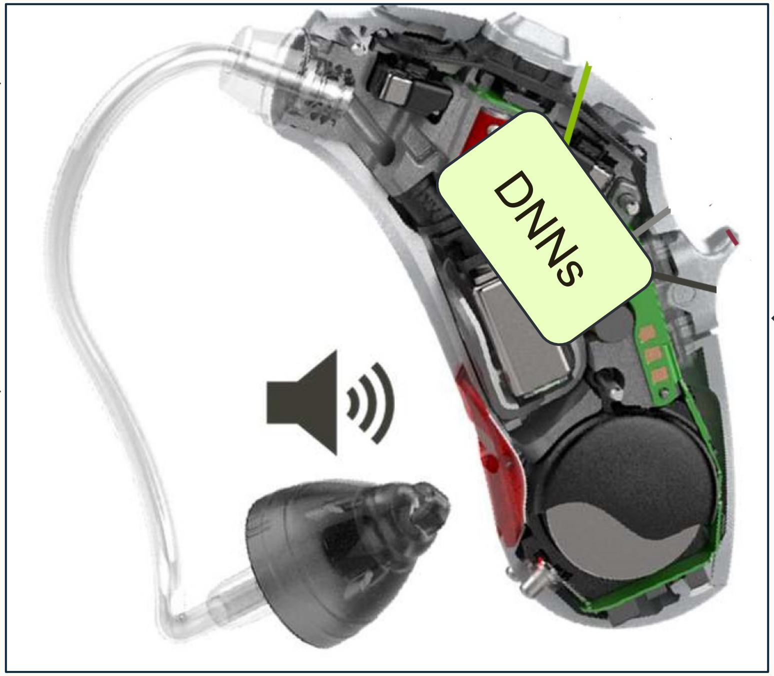
Our overall aim is to create "multi-modal” (**MM**) aids which not only amplify sounds but contextually use simultaneously collected information from a range of sensors to improve speech intelligibility.

Audio

Cognitive-load sensors

Video
RF sensing
AVSEC challenge
“Wireframe” lipreading

Constraints ..
.. Privacy
.. Wearability



“IoT” data

cloud

Transmission architectures
Two-point neural models

Constraints
.. Devices need to be very small
.. Tiny LI-ION cell, power budget: <1-5 mW
.. Very low audio processing delay: < 10 ms

https://cogmhear.org/

Codec Frame Structures

2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom)

A Novel Frame Structure for Cloud-Based Audio-Visual Speech Enhancement in Multimodal Hearing-aids

Abhijeet Bishnu*, Ankit Gupta†, Mandar Gogate‡, Kia Dashtipour‡, Ahsan Adeel§, Amir Hussain‡, Mathini Sellathurai†, and Tharmalingam Ratnarajah*
* *School of Engineering, University of Edinburgh*, Edinburgh, United Kingdom
Email: {abishnu,t.ratnarajah}@ed.ac.uk
† *School of Engineering & Physical Sciences, Heriot-Watt Watt University*, Edinburgh, United Kingdom
Email: {ankit.gupta,m.sellathurai}@hwu.ac.uk
‡ *School of Computing, Edinburgh Napier University*, Edinburgh, United Kingdom
Email: {m.gogate, k.dashtipour, a.hussain}@napier.ac.uk
§ *School of Mathematics & Computer Science, University of Wolverhampton*, Wolverhampton, United Kingdom
Email: a.adeel@wlv.ac.uk

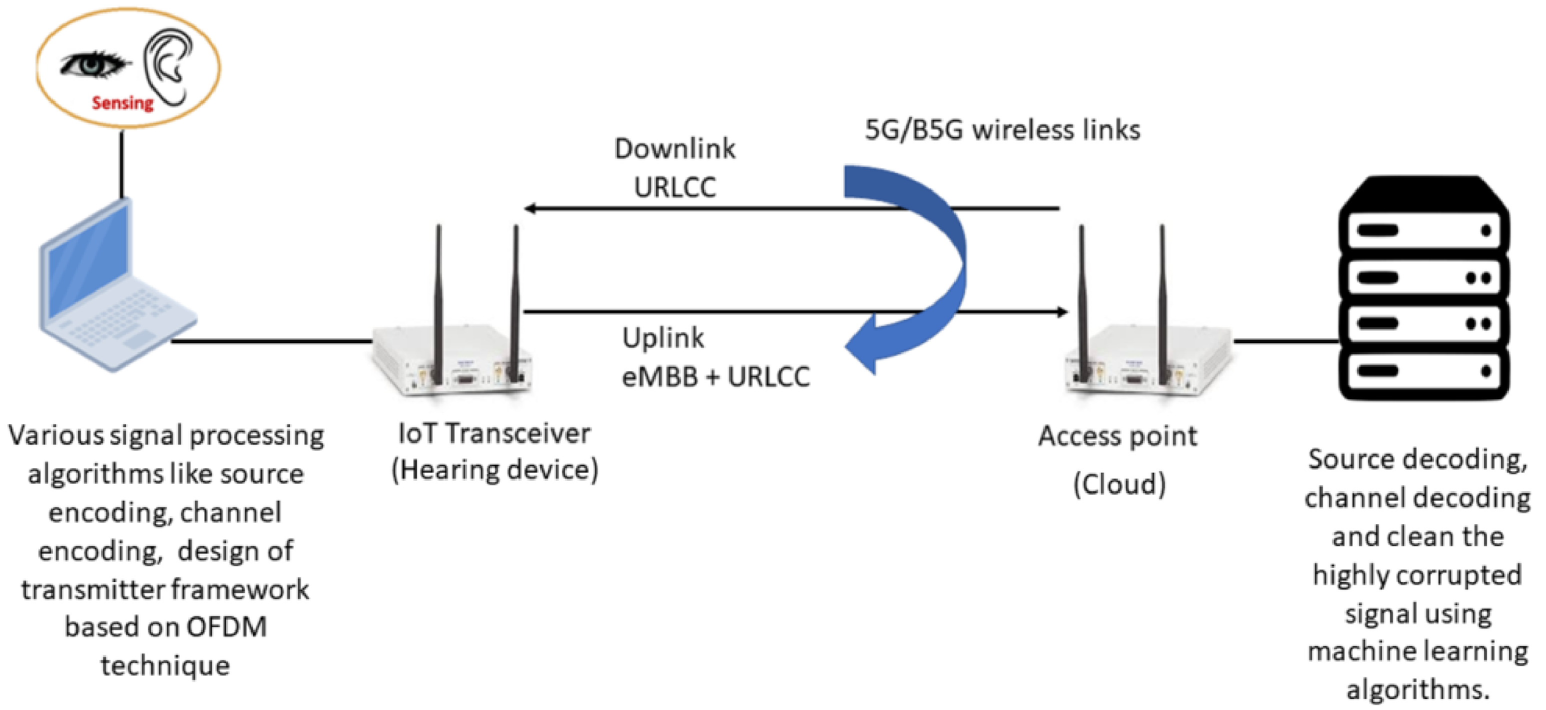


Fig. 1. Model of cloud-based audio-visual speech enhancement hearing aid

Codec Frame Structures

TABLE I
COMPARING STATE-OF-THE-ART AUDIO CODECS

Parameters	OPUS codec	EVS codec
Signal Bandwidth	4 kHz to 24 kHz	4 kHz to 20 kHz
Supported Bit-rates	6 kbps to 510 kbps	5.9 kbps to 128 kbps
Standardized By	IETF (in 2012)	3GPP (in 2016)
Used By	YouTube, Skype, Zoom, MS Teams	Voice over LTE (VoLTE)
Performance Comparison	EVS outperforms OPUS at low bit- rates	
Latency Comparison	26.5 ms	32 ms

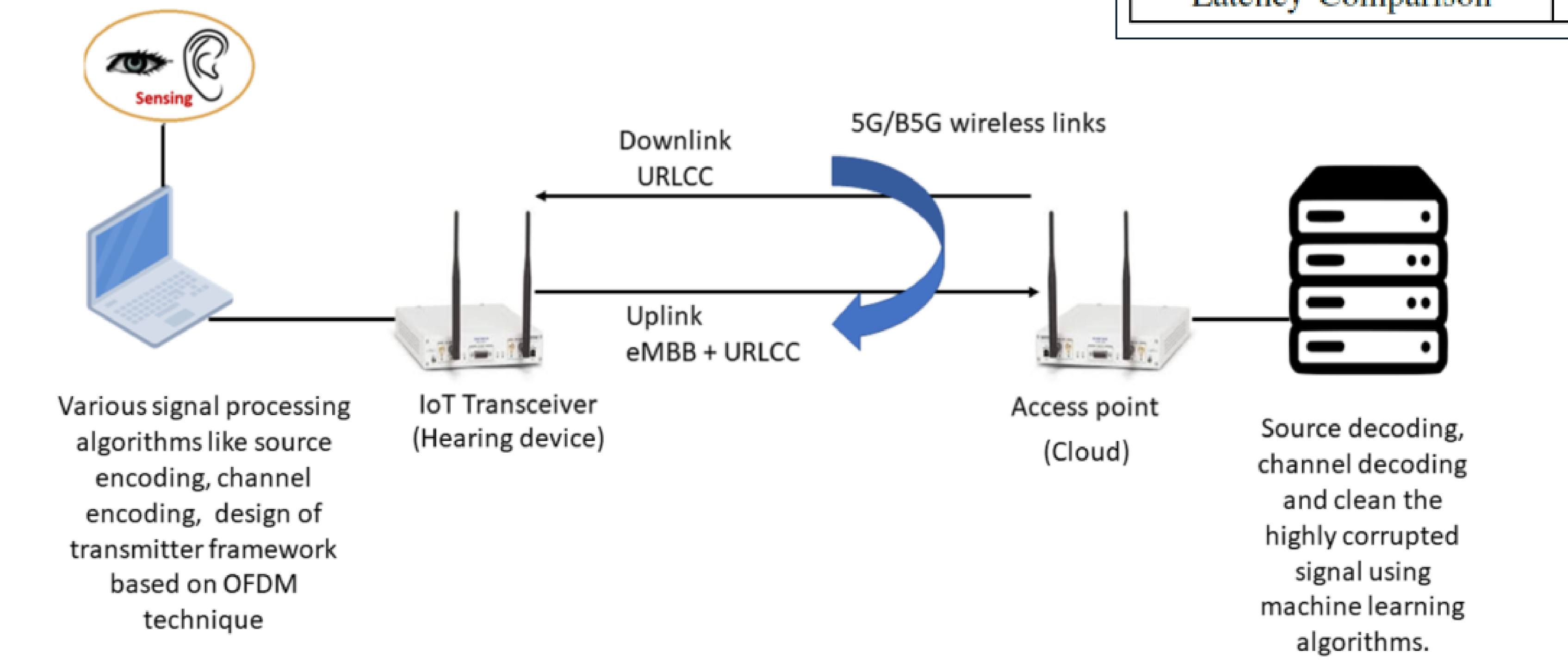


Fig. 1. Model of cloud-based audio-visual speech enhancement hearing aid

Codec Frame Structures

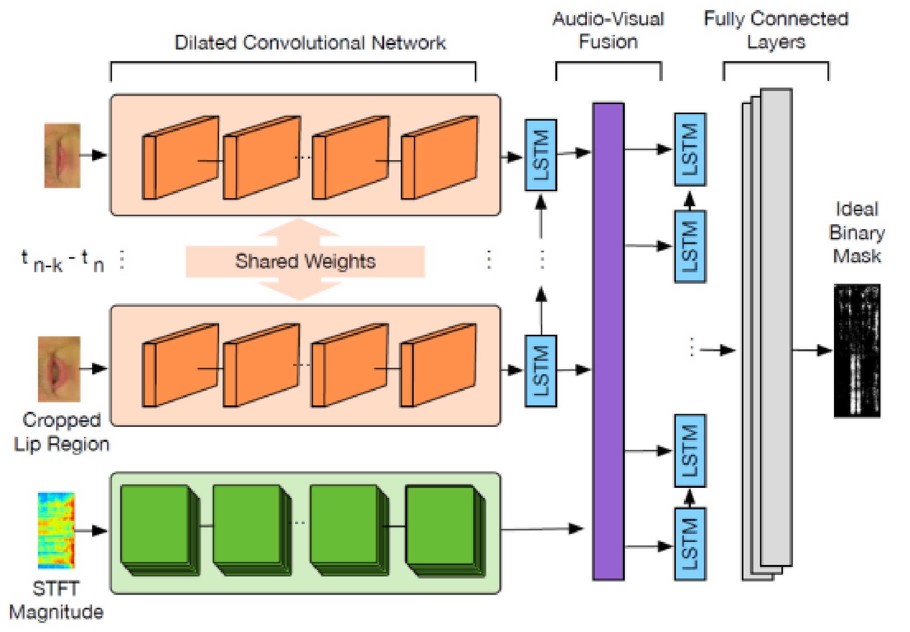
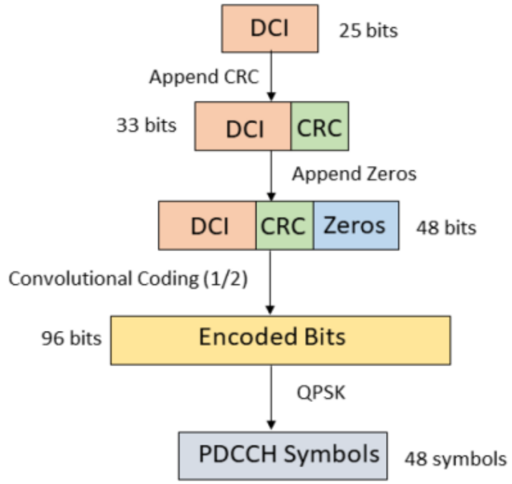


Fig. 2. Proposed AV Speech Enhancement model

TABLE III
BITS OF DOWNLINK CONTROL INFORMATION

MIB Parameters	# of Bits
Frame Number	10
Code Rate	1
Modulation	1
# of Frames in a Transport Block	4
End of Payload	1
Uplink SS ID	3
Reserved	5



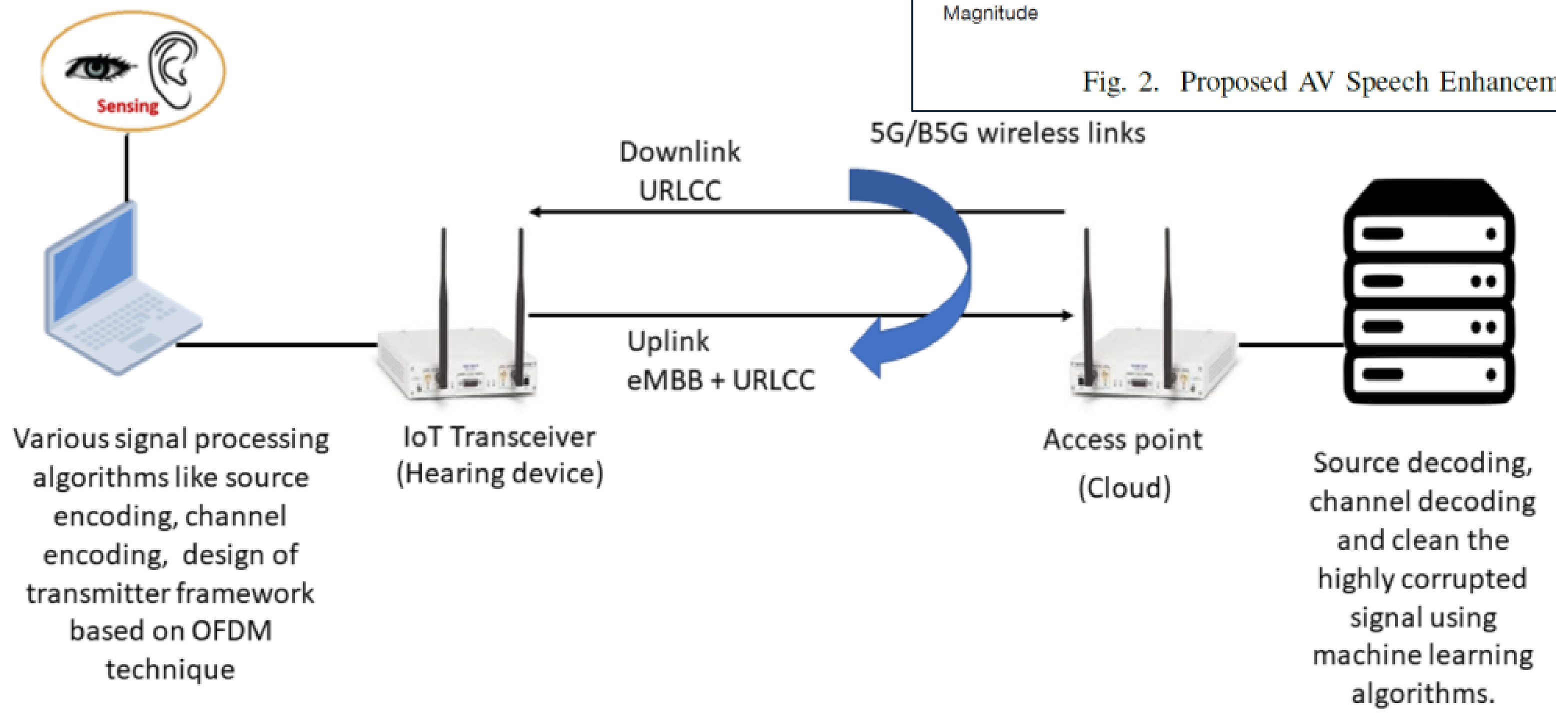


Fig. 1. Model of cloud-based audio-visual speech enhancement hearing aid

Radio-frequency lip-reading

Article

https://doi.org/10.1038/s41467-022-32231-1

Pushing the limits of remote RF sensing by reading lips under the face mask

Received: 1 April 2022

Accepted: 21 July 2022

Published online: 07 September 2022

Check for updates

Hira Hameed[1], Muhammad Usman[1,2], Ahsen Tahir[1,3], Amir Hussain[4], Hasan Abbas[1], Tie Jun Cui[5], Muhammad Ali Imran[1] & Qammer H. Abbasi[1]

The problem of Lip-reading has become an important research challenge in recent years. The goal is to recognise speech from lip movements. Most of the ip-reading technologies developed so far are camera-based, which require deo recording of the target. However, these technologies have well-known mitations of occlusion and ambient lighting with serious privacy concerns. urthermore, vision-based technologies are not useful for multi-modal hearing ds in the coronavirus (COVID-19) environment, where face masks have ecome a norm. This paper aims to solve the fundamental limitations of amera-based systems by proposing a radio frequency (RF) based Lip-reading amework, having an ability to read lips under face masks. The framework nploys Wi-Fi and radar technologies as enablers of RF sensing based Lip-ading. A dataset comprising of vowels A, E, I, O, U and empty (static/closed ps) is collected using both technologies, with a face mask. The collected data used to train machine learning (ML) and deep learning (DL) models. A high assification accuracy of 95% is achieved on the Wi-Fi data utilising neural etwork (NN) models. Moreover, similar accuracy is achieved by VGG16 deep arning model on the collected radar-based dataset.

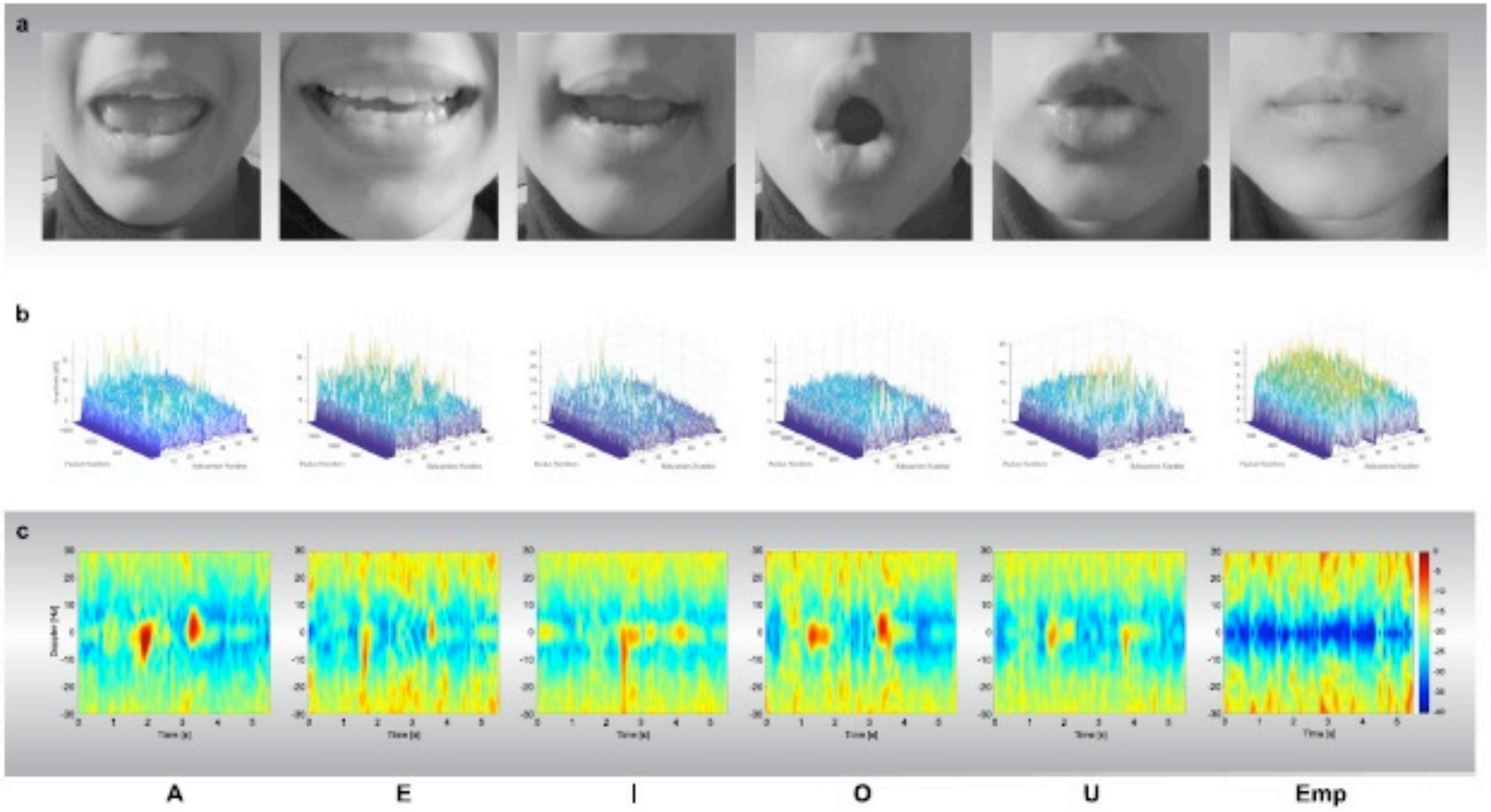
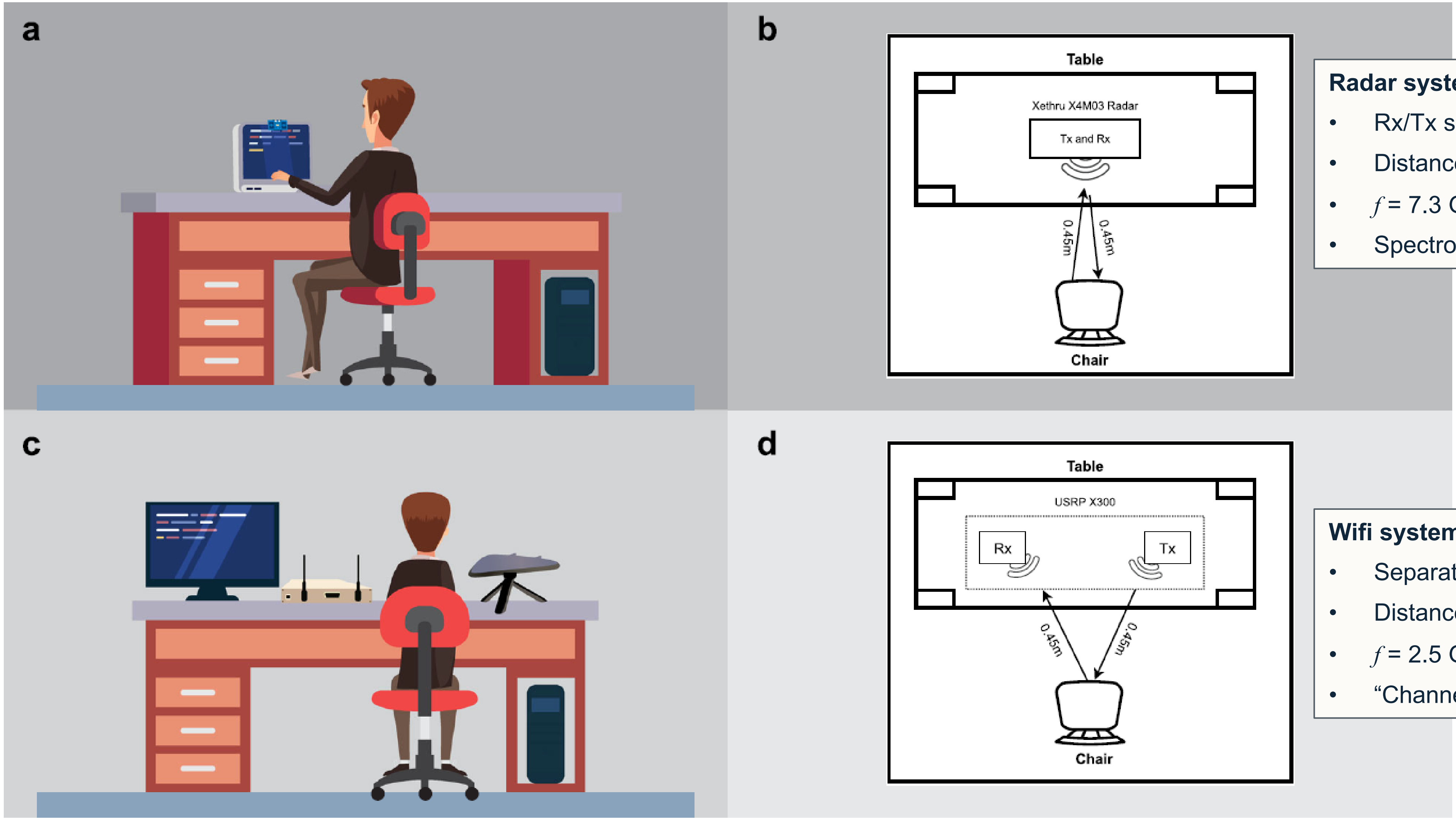


Fig. 2 | Pronounced vowels with their representation in Wi-Fi and radar signal. **a** A visual illustration of the pronounced vowels. **b** Wi-Fi data samples with mask representing various vowels classes. **c** Radar data samples with mask representing various vowels classes.

Radio-frequency lip-reading

“Fig. 3 | Experimental setup of the data collection through radar and Wi-Fi.
A Front view of the data collection setup using XethruUWB radar.
B Top view of the radar-based data collection.
C Front view of Wi-Fi based data collection.
D **Top view of the Wi-Fi-based data collection setup.”**



Radar system

- Rx/Tx sensor on top of laptop screen
- Distance = 0.45 m
- f = 7.3 GHz ∴ λ = 4 cm
- Spectrograms of Doppler shifts

Wifi system

- Separate Tx/Rx on desk.
- Distance = 0.45 m
- f = 2.5 GHz ∴ λ = 12 cm
- “Channel-state-information” amplitude

Radio-frequency lip-reading

Accuracy of best deep-learning model classifying the radio-freq data

- 5 vowels + blank
- x 3 talkers
- x with/without facemask

Radar .. 73% without facemask
86% with ..

Wifi .. 61% without facemask
73% with ..

Radar system

- Rx/Tx sensor on top of laptop screen
- Distance = 0.45 m
- f = 7.3 GHz $\therefore \lambda$ = 4 cm
- Spectrograms of Doppler shifts

Wifi system

- Separate Tx/Rx on desk.
- Distance = 0.45 m
- f = 2.5 GHz $\therefore \lambda$ = 12 cm
- “Channel-state-information” amplitude

Radio-frequency BSL

Accuracy of best deeplearning model at classifying 15 BSL gestures x 4 presenters (about 3:1 training-testing ratio of data) = 90%

2090 IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 10, NO. 4, AUGUST 2023

Recognizing British Sign Language Using Deep Learning: A Contactless and Privacy-Preserving Approach

Hira Hameed, *Student Member, IEEE*, Muhammad Usman, *Senior Member, IEEE*, Ahsen Tahir, *Member, IEEE*, Kashif Ahmad, *Senior Member, IEEE*, Amir Hussain, *Senior Member, IEEE*, Muhammad Ali Imran, *Senior Member, IEEE*, and Qammer H. Abbasi, *Senior Member, IEEE*

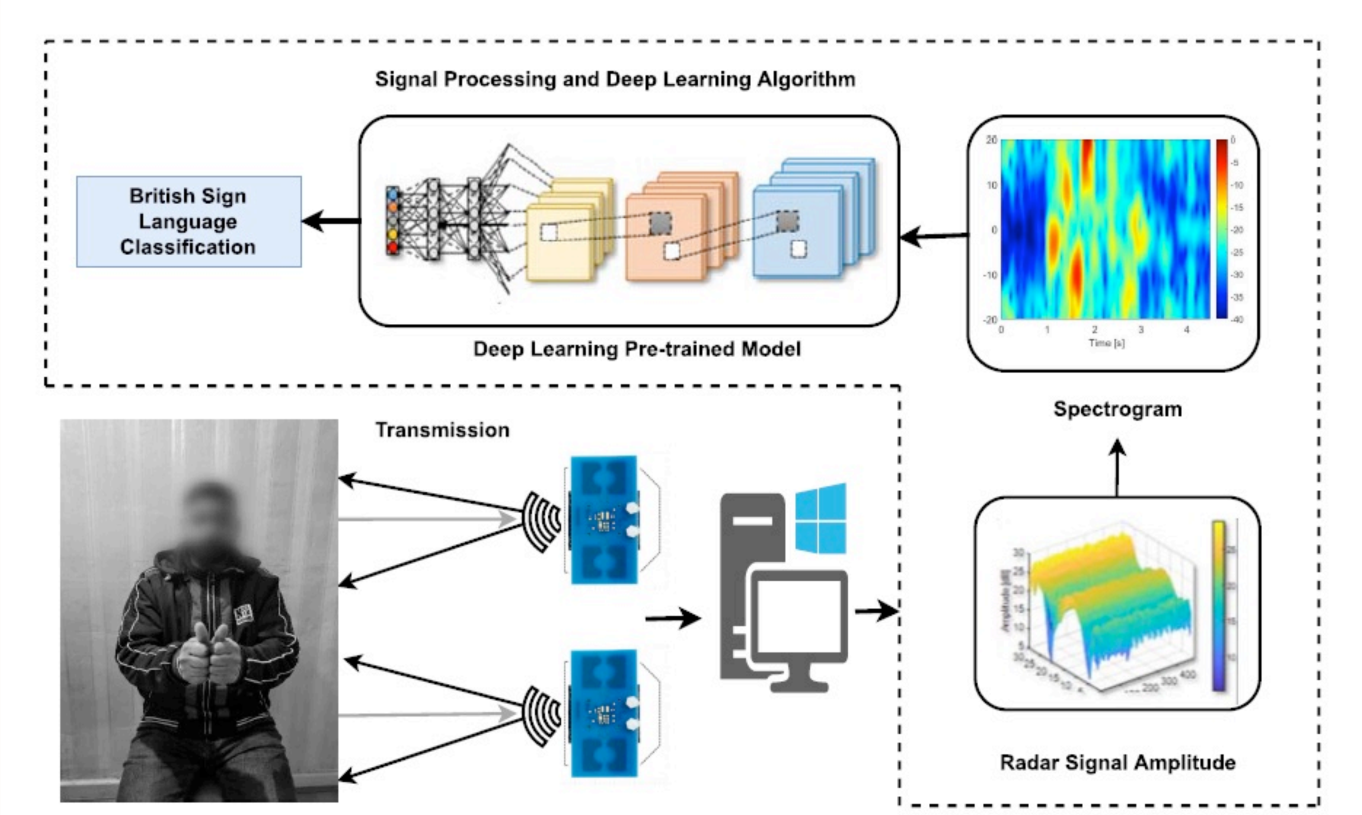


Fig. 3. Visual illustration of the pronounced BSL. (a) *Brother*. (b) *Sister*. (c) *Mother*. (d) *Father*. (e) *Family*. (f) *Confuse*. (g) *Depress*. (h) *Happy*. (i) *Hate*. (j) *Sad*. (k) *Walk*. (l) *Eat*. (m) *Help*. (n) *Drink*. (o) *Stop*.

https://cogmhear.org/