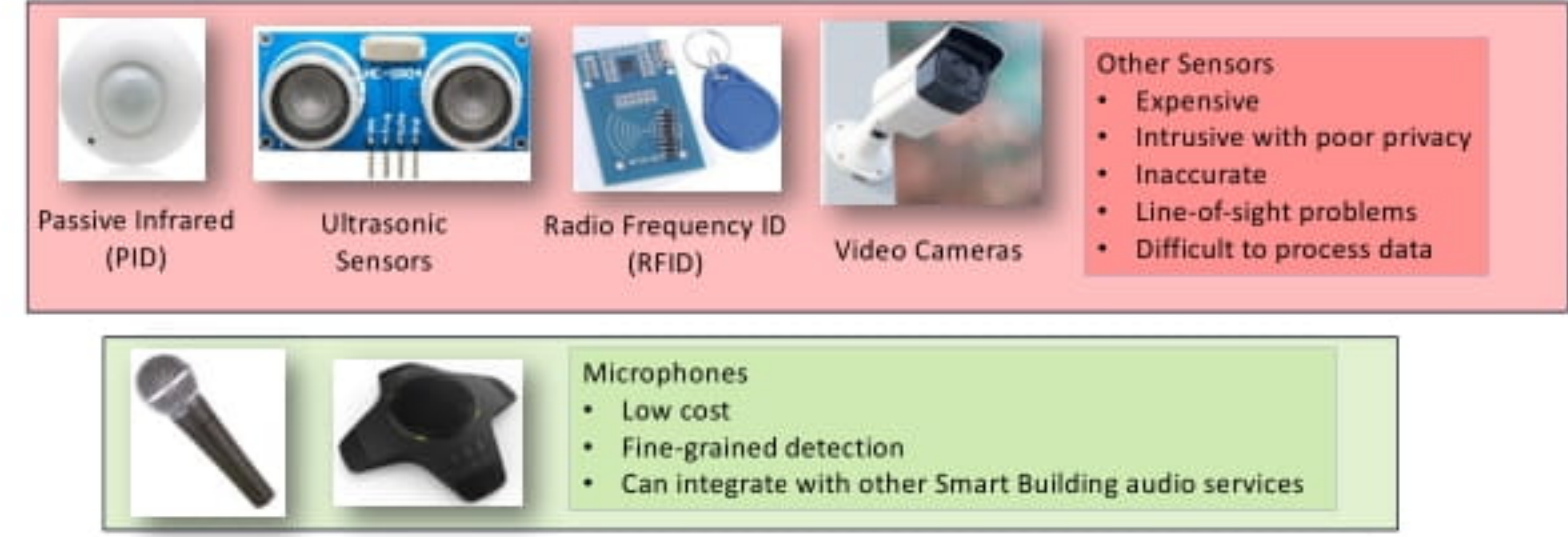# Developing Privacy-Preserving Audio Capture for Smart Buildings

School of Electronics and Computer Science
University of Southampton

Jennifer Williams and Sebastian Stein
j.williams@soton.ac.uk, s.stein@soton.ac.uk

Citizen-Centric AI Systems (CCAIS)

Background – Occupancy Detection

Other Sensors
- Expensive
- Intrusive with poor privacy
- Inaccurate
- Line-of-sight problems
- Difficult to process data

Passive Infrared (PID) · Ultrasonic Sensors · Radio Frequency ID (RFID) · Video Cameras

Microphones
- Low cost
- Fine-grained detection
- Can integrate with other Smart Building audio services

Audio is an under-explored solution for occupancy detection

Huang, Q. (2018). Occupancy-driven energy-efficient buildings using audio processing with background sound cancellation. Buildings, 8(6), 78.
Chen, Z., Jiang, C., & Xie, L. (2018). Building occupancy estimation and detection: A review. Energy and Buildings, 169, 260-270.
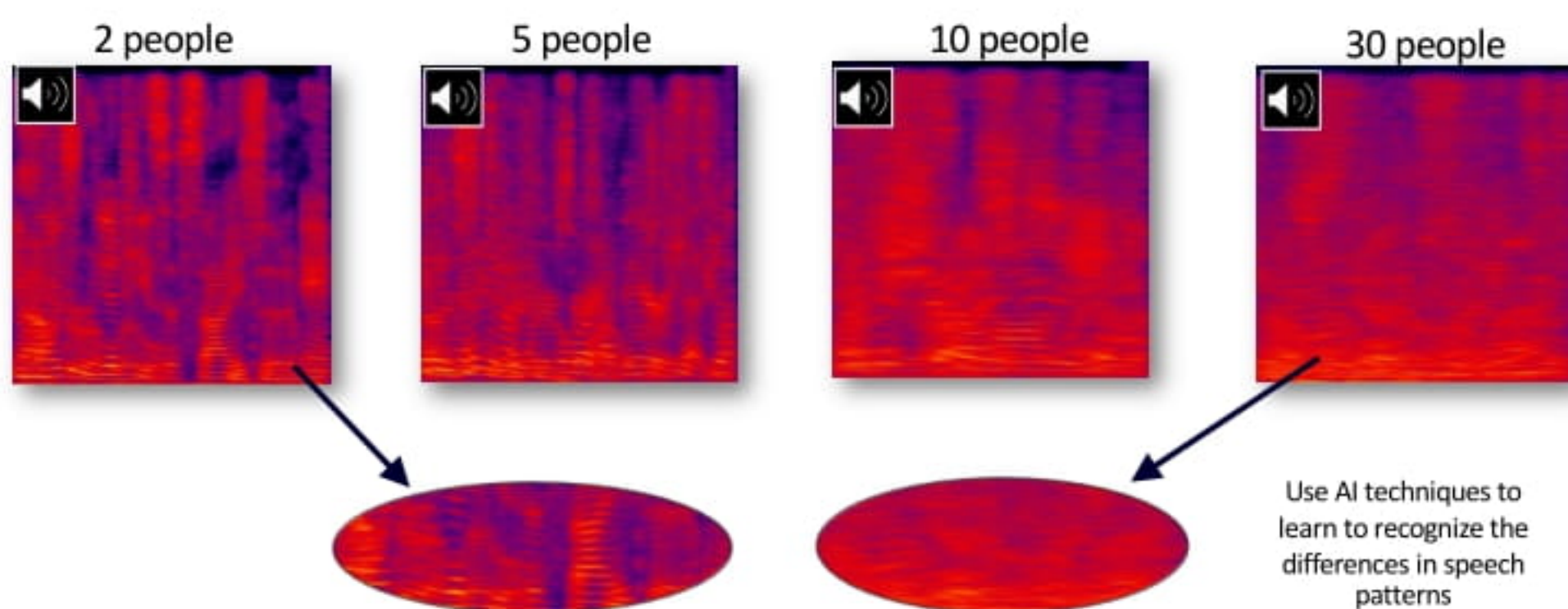
Why Are Audio Services Needed?

Standard occupancy schedules for office buildings are changing

More emphasis on occupant-centric design and occupant comfort

Need accurate and inexpensive solutions for occupant counting and activity detection

Smart buildings (of the future) can offer a variety of services that enhance satisfaction and productivity

Benefits of Audio Services:
Affordable, accurate, privacy-preserving
Occupancy detection, activity detection, accessibility, meeting recording, secure voice-identity access

Summary of Findings and Ongoing Work

- Smart buildings must meet the needs of occupants for **comfort**, **energy** usage, **costs**, and **information** services

- **Audio** is an effective and low-cost solution for occupancy sensing

- Microphone data is **less invasive than video** and there are known techniques to conceal conversation and identity from speech

- Buildings equipped with audio capture can provide **energy** management, **business** services, **emergency** services, and **accessibility** services (e.g., for blind and deaf)

- **Occupancy** and **activity** detection information can enhance building management systems, further reducing carbon footprints

## Method and Approach – "Cocktail Party" Problem

2 people · 5 people · 10 people · 30 people

Use AI techniques to learn to recognize the differences in speech patterns

Goal: find representations of audio that correlate to the number of people

Audio Spectrograms
- The spectrograms are different depending on the **number of people talking**
- Machine learning algorithms **can learn differences** within the spectrograms
- This approach **assumes overlapping talkers**
- Currently the model **does not save 'state' information** (leave/enter the room)
- Predicting 1 to 10 talkers is **more fine-grained than previous work**

Initial Results – Convolutional Neural Network

Occupancy Prediction Accuracy
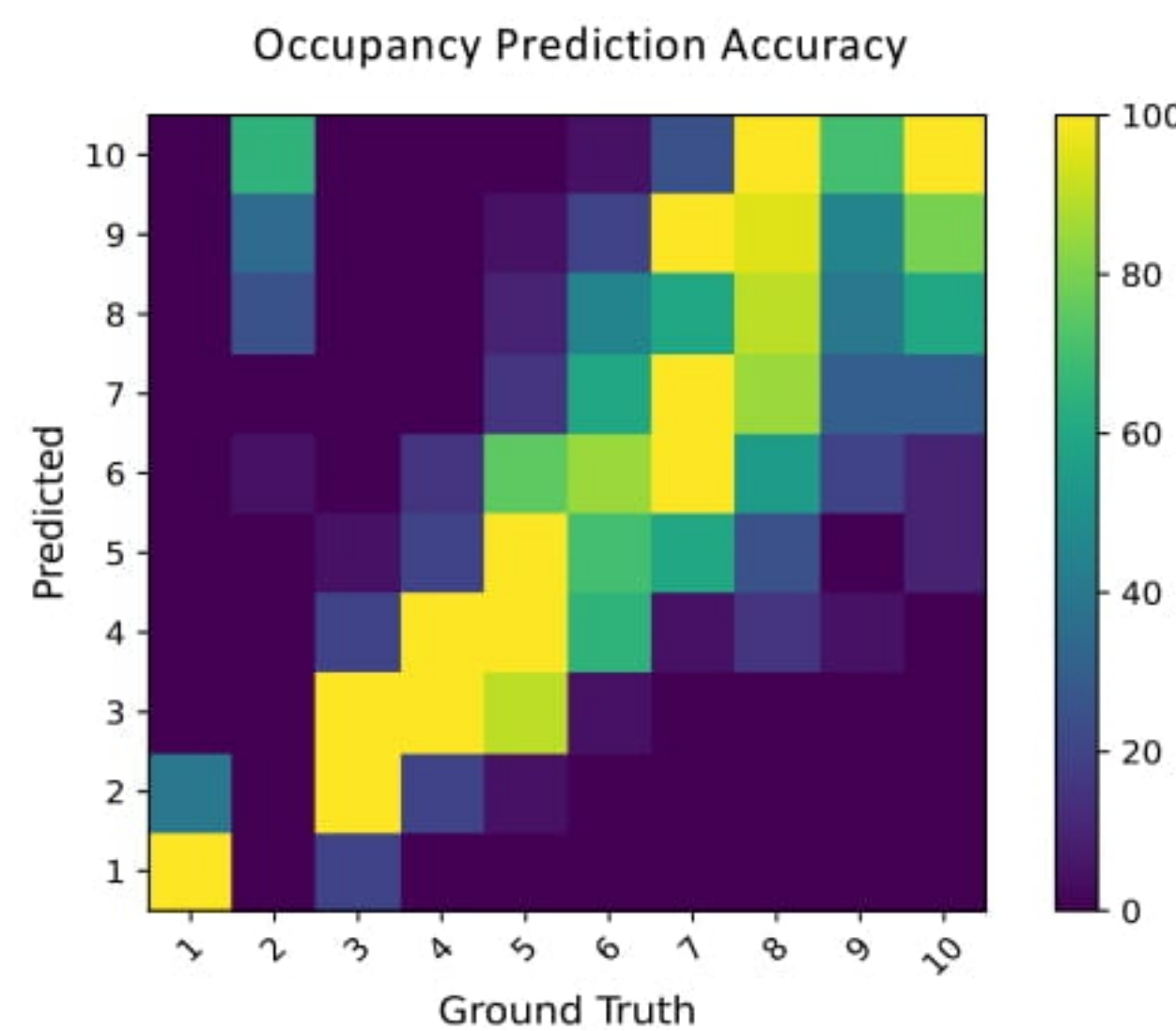
- Detects number of people talking simultaneously

- High accuracy from full waveform representations

- **Finer granularity prediction than previous work**

- Generalizes to people not seen during AI-training

- Some occupancy levels are easily confused with another occupancy level

AI algorithms can learn to detect the number of people speaking in audio

Next Steps: Add Privacy-Preserving Mechanisms
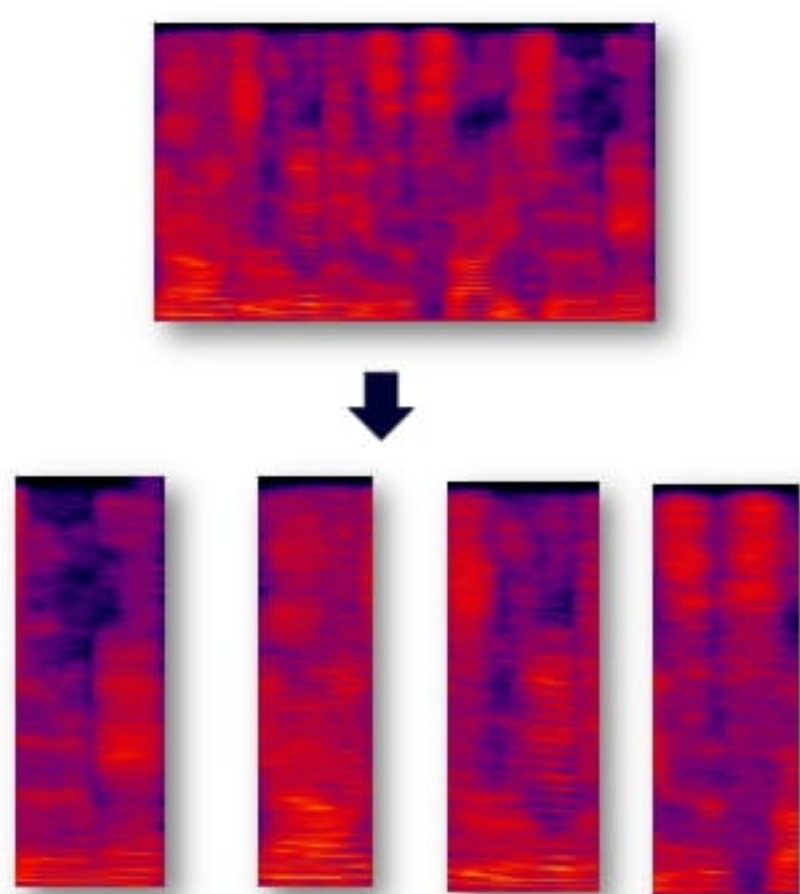
*"Shredding" audio re-arranges frames and prevents words from being recognized*

Limitations
- Assumes that people talk at the same time (known as "cocktail party" scenario)
- Full audio used for the prediction does not preserve privacy
- Model does not keep a memory of state (people enter and then leave a room)
- Model does not account for non-speaking occupants

Next Steps:
- Try the technique of "shredding" to rearrange the audio frames and obscure content
- Explore signal-level feature representations that do not identify individuals
- Develop a model that maintains state as people enter and leave the conversation
- Include non-speech events and different room characteristics

Multiple Uses of Audio in Smart Buildings
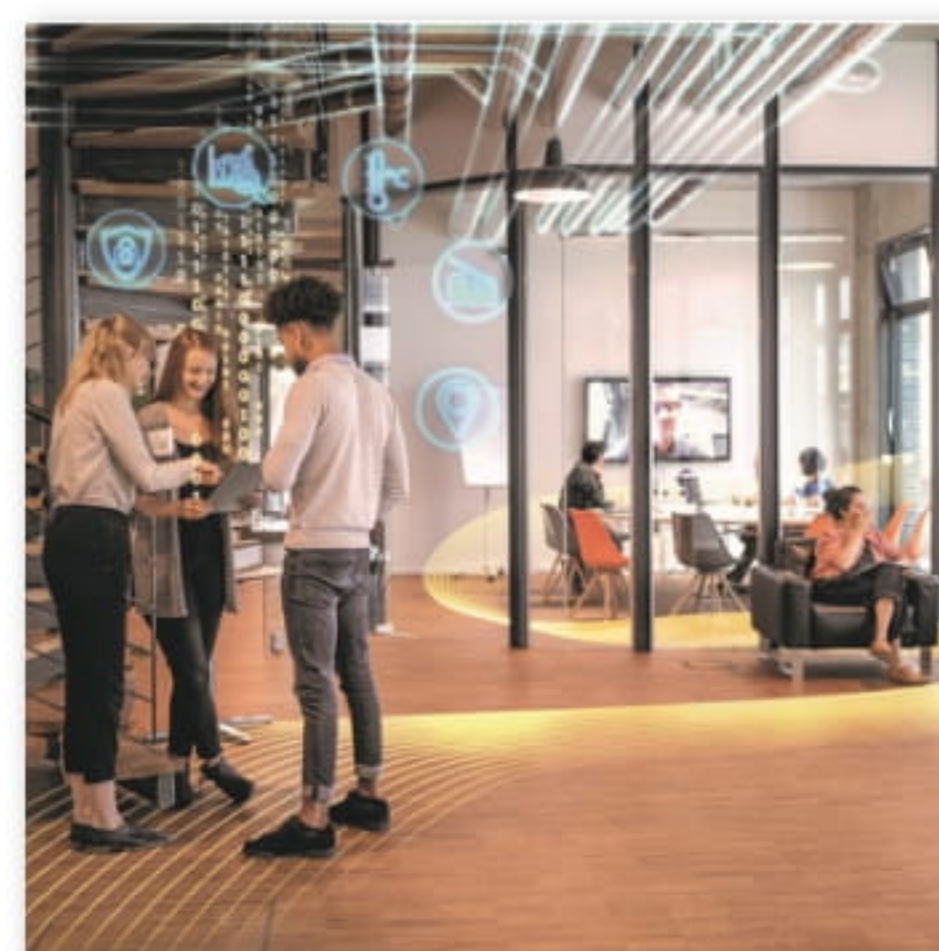
- Occupancy and activity detection for HVAC and electricity management

- Audio scene understanding

- Real-time speech captions from meetings

- Meeting interaction and engagement

- Sign-language recognition

- Voice-activated emergency services

- Voice identity access to secure rooms or storage