

## What makes a particular voice easy or difficult for automatic speaker recognition systems to recognise?

Issues with current work on ASR:

- Focus on error rates/ overall measures of performance
- Systems generally developed for other application; priorities in forensics are different
- One-size fits all approach; same system for every voice
- Work looking at ASR errors often small scale
- Lack of understanding about potential biases
- Admissibility (in the UK)

## Research Questions

1. What systematic properties of speakers make them more or less susceptible to ASR errors, in terms of voice and demographic factors? How do the magnitudes of these effects compare to known technical effects?
2. How consistent are results for individual speakers within and across ASR systems?
3. How do results produced by techniques that combine ASR and linguistic methods on a person-specific basis compare with the current one-size-fits-all approach?
4. How generalisable are methods and results across datasets and languages?

### Workpackage 1: *Controlled input from phonetically trained speakers*

- Read phonetically balanced text on multiple occasions; systematically varying aspects of speech production
- Simultaneous recordings with multiple channels and devices
- Technical manipulation of sound files (e.g. increasing mean f0)
- Analysis of relative effect on ASR scores

### Workpackage 2: *Uncontrolled input from forensically realistic data*

- ASR and linguistic analysis of large-scale databases
- Analysis of ASR scores/ LR as a function of linguistic and demographic information about speakers
- By-speaker performance metrics
- Effects of different ASR systems on output for individual speakers

### Workpackage 3: *Developing person-specific approaches*

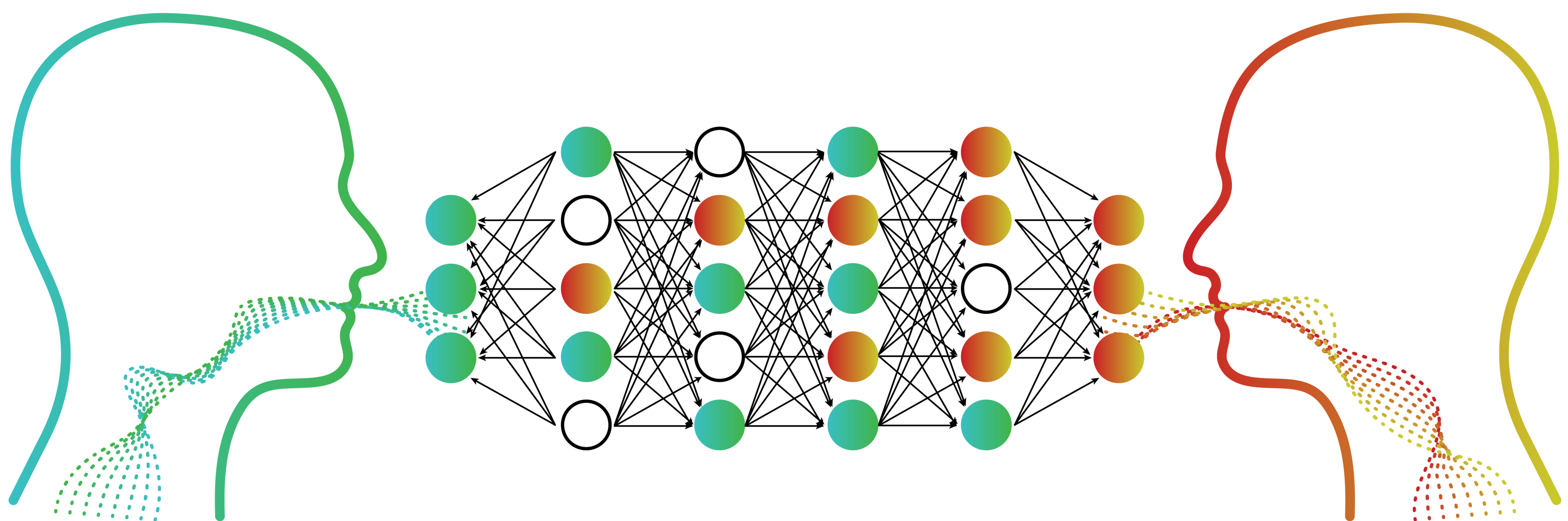
- Identifying *problematic* speakers for ASR
- Integrate knowledge about individuals in the system
- Analysis of *problematic* speakers by forensic experts – validating combination of ASR and human methods
- Build bespoke models for individual speakers

# Person-specific automatic speaker recognition: understanding the behaviour of individuals for applications of ASR

Vincent Hughes, Philip Harrison, Paul Foulkes, Jessica Wormald, Chenzi Xu<sup>1</sup>, David van der Vloed<sup>2</sup> and Finnian Kelly<sup>3</sup>

[vincent.hughes@york.ac.uk](mailto:vincent.hughes@york.ac.uk)

<https://www.pasr.york.ac.uk/>



(ES/W001241/1; £1,012,570)