

Towards linguistically-informed automatic speaker recognition



WRoCAH funded Collaborative Doctoral Award between Department of Language and Linguistic Science at the University of York, and Aculab PLC¹

Lead Academic and Partner Organisation Supervisors

Dr Vincent Hughes

Department of Language and Linguistic Science
University of York

Dr Ladan Baghai-Ravary

Speech Technology Director
Aculab PLC

Project summary

Automatic speaker recognition (ASR) is increasingly used by commercial institutions to identify individuals from their voices, and by forensic labs for voice evidence in legal cases. Despite their very low error rates in experiments, little is known about what information is actually captured by state-of-the-art ASR systems. This project examines how speaker-characterising information captured by ASR systems maps on to linguistic properties of the voice. This study has important implications for the development and improvement of ASR, enhancing public and legal/judicial understanding of ASR, and improving voice analysis in forensic cases.

Project description

Understanding and modelling the human voice in all its complexity is a key issue in both the humanities (linguistics, phonetics) and the sciences (engineering, computer science). Different disciplines approach this problem in fundamentally different ways. Linguistics adopts a componential approach: speech is broken down into its constituent parts (vowels, consonants, pitch, etc.) via analytical listening and acoustic measurements. Speech technology works holistically, extracting abstract features from longer speech samples to build a mathematical model of the voice. Despite the clear overlap in interests, very little work has been situated at the intersection between linguistics and speech technology.

The proposed project focuses on speaker recognition – i.e. identifying individuals from their voice. Banks and other institutions (e.g. HMRC) use automatic speaker recognition (ASR) systems to verify the identity of customers attempting to access accounts. Such systems are also used by forensic scientists (e.g. the Spanish police) as evidence in cases involving

¹ If you are selected by the CDA project team to be put forward to WRoCAH, your application will be considered competitively within the main WRoCAH competition. Being put forward for consideration is not a guarantee of funding.

recordings of unknown criminal voices. It is of central importance that such systems make very few errors, preventing imposters from gaining access to bank accounts, and protecting innocent people from prosecution.

ASR is widely perceived to be a black box, whereby the inner workings of the system are opaque to users, and in some cases even to developers. State-of-the-art ASR systems are now increasingly based on artificial intelligence. This means that the systems themselves learn the best way of structuring and analysing the data. These approaches have shown dramatic improvements in performance, achieving almost perfect recognition under certain conditions. However, as the underlying algorithms get more complex, understanding their inner workings becomes increasingly difficult. This is now arguably the biggest problem for the field of ASR.

This project contributes to the small but emerging body of research investigating how linguistic information can help us better understand ASR systems. The project will be hosted by the University of York and Aculab, and will address three key questions:

1. To what extent do ASR systems capture tangible linguistic properties of a voice?
2. By understanding what information is captured by ASR systems, can we predict which speakers will be problematic for the system?
3. Can linguistic information be used to improve the performance of ASR?

The project will use collections of voice recordings available at York and Aculab. Voices will be processed using Aculab's VoiSentry ASR system and the output will be compared with linguistic data extracted from the same voices. The linguistic approach will involve auditory and acoustic analysis of features commonly used in forensic voice casework, e.g. vowels, consonants, and pitch. The voice samples will also be manipulated systematically to alter certain linguistic features (such as pitch, timbre etc.) and evaluate the effects on the output of the system. As there is no methodological precedent for how to approach this issue, there will be considerable scope for the student to mould the project.

The collaboration between York and Aculab is unique. York is home to a world leading research centre for forensic speech science. The project builds on work on our AHRC-funded Voice and Identity project, which represents one of the only large-scale investigations into how speaker-specific information is encoded in the voice, by comparing the output of ASR systems with linguistic approaches to assess their efficacy and complementarity. Aculab is a leading commercial developer of ASR systems used by institutions such as banks. The student will have unique access to the underlying code of the VoiSentry system, allowing research into the inner workings of ASR in ways that were not possible previously. Aculab will host the student, providing unique employment-related skills via experience in a commercial environment.

About Aculab PLC

Aculab currently offers a range of speech technology products, including a speaker verification system, "VoiSentry". The proposed project will allow us to better understand the operation of our current system, and could potentially provide us with knowledge which will allow us to enhance both VoiSentry and our other products. It will also allow us to

address new forensic applications of our technology, and in general it will increase the uptake of automatic speech and voice analysis within the forensics industry, by providing a legally acceptable interpretation and justification of the evidence which such systems can give.

Engagement, outreach, dissemination and impact opportunities

Aculab organises workshops and similar activities, and is frequently represented at academic conferences and other meetings. We would encourage the student to submit papers to such events, and produce white papers which would publicise their work to both industry and the media. Where deemed appropriate by the academic supervisor(s), we would hope the student would be able to contribute to the creation of metadata for some extensive telephone-based speech databases that we have collected.

Other Partner Organisations

The successful student candidate will benefit from an existing relationship between the University of York and J P French Associates (JPFA), the UK's largest forensic speech and audio laboratory. Members of staff at JPFA hold teaching and research positions within the Department of Language and Linguistic Science. The student candidate would gain real-world forensic insights through consultation with these members of staff and through presentations at our forensic speech science research group meetings, regularly attended by JPFA staff. The partnership has previously delivered grants totalling over £2m, and is the foundation for a 4* rated impact case study in REF 2014.

Financial support and eligibility

Full-time AHRC Competition Studentships for doctoral research are 3 years in duration (or 6 years part time). Awards are subject to satisfactory academic progress. Awards must be taken up in October 2019. No deferrals are possible. Awards will comprise UK/EU fees at Research Council rates and, for eligible students, a maintenance grant (£14,777 in 2017/8).

Please note that all applicants should meet the AHRC's academic criteria and residency requirements (<http://wrocah.ac.uk/new-student/2019-cda/>).

How to apply

Before applying for any WRoCAH Studentship, please first ensure that you have read the WRoCAH webpages about Collaborative Doctoral Awards, the WRoCAH training programme and requirements

<http://wrocah.ac.uk/new-student/2019-cda/>

Please note that the selection of Collaborative Doctoral Award applicants is a two stage process:

STAGE 1

Please send the following information directly to vincent.hughes@york.ac.uk no later than 14 December 2018

- A c.v. detailing your academic record (no more than 3 pages);
- Academic transcript from your university;
- Names and contact details for two referees who will be able to provide a reference by the interview on 3rd December at latest if required;
- 800 word statement of purpose.

As the general scope of the project is already defined, the 800 word statement of purpose should explain why you would be interested in undertaking this particular programme of research and what experience and skills you would bring to the position.

Deadline for STAGE 1 applications

14 December 2018

Date of interviews

8-9 January 2019

STAGE 2

Apply for funding to WRoCAH

Candidates selected after Stage 1 then liaise with the project supervisory team to submit a WRoCAH studentship application via the WRoCAH online application form by the **5pm on Wednesday 23 January 2019 deadline**.

NB: Candidates who have been unsuccessful in Phase One can submit a separate WRoCAH application as a non-CDA applicant.

For more information about this project contact

Dr Vincent Hughes (vincent.hughes@york.ac.uk)