



SPEECH RECOGNITION IN POLICING

Exploring the safe use of AI-based speech and language tools for law enforcement

Research to inform policy from the University of York. Jointly written by the [Forensic Speech Services group](#)¹ in the Language & Linguistic Science Department and the [UKRI AI Centre for Doctoral Training in Safe Artificial Intelligence Systems](#)² in the Department of Computer Science.

Jessica Wormald¹, James Tompkinson¹, Lauren Harrington¹, Ben Gibb-Reid¹, Ibrahim Habli², Philip Harrison¹, Vincent Hughes¹, Colin Paterson², Faisal Alkhatib², Madeleine Nielsen²

Summary

- In January 2026 the Home Office published a White paper: [From Local to National: A New Model for Policing](#). This white paper details i) the introduction of a National Police Service to ensure consistency across all forces, and ii) an announcement about the launch of Police.AI; a £115m investment to “enable the rapid and responsible adoption of AI” with speech and language tools considered a priority. This is echoed in [Part 2 of the Independent Review of Criminal Courts](#), published in February which recommends the use of AI tools, including transcription, to support clearing the backlog in the Criminal Justice System.
- In December 2025, the House of Lords Public Services Committee published a [policy letter](#) identifying a number of opportunities around the transcription of police suspect interviews. These included standardisation around the training, production and checking of human-produced transcripts. Acknowledgement was made within the letter of opportunities around technology, although it highlighted that these opportunities are in part hindered by the general lack of audio and visual investment across policing and within the court system.
- Responses from a University of York Freedom of Information request undertaken in Summer 2025 highlighted that a number of individual police forces are considering or trialling speech recognition tools to support transcription of police interviews. We also know that forces are using speech technologies for a variety of purposes beyond transcription, and these tools can be embedded in other technologies.
- The NPCC has published overarching principles and guidance around the use and adoption of AI in a [Covenant and Strategy](#). The [NPCC Playbook](#) provides the most extensive guidance but there is currently limited detail on the practicalities of integrating tools within existing workflows. Additionally, many tools are procured from developers or industry providers, with few opportunities for users to fine-tune or assess performance by more relevant metrics.

Recommendations for policy

1. Develop a standard national framework for the transcription of police interviews and appropriate assurance metrics for AI-based speech recognition as part of a human-centred process.
2. Build on best practice in safety-critical systems, especially those using AI, by developing safety cases for the deployment in order to improve transparency and support accountable deployment. Each use case (e.g. police interviews v. witness statements) needs a specific safety case to ensure safe deployment.
3. Engagement with academic researchers is critical to provide independent, domain-specific expertise to support different areas. This applies to the use case outlined in 1, and equally to wider deployment of speech and language tools.

Police interviews: a targeted use case?

Producing a written record of a conversation provides an accessible way for information to be reviewed. Speech differs from writing in many ways, especially when compared with standard written language. However, a written document can sometimes be a practical and efficient way for users to access information. This is often the case for police interviews, these are usually long and only certain parts of information need to be accessed by a given individual at any one point. Transcription of police interviews by humans is a labour intensive process, with police forces dedicating significant resources to facilitate it. Thousands of documents are produced annually and support countless investigations. However, there is currently no standardised framework; each of the 43 police forces in England and Wales has its own internal process. The lack of consistency is reflected in the type of transcript produced: for example, some produce verbatim (word-for-word) transcripts, whereas other forces produce only summaries of interviews.

AI-based automatic speech recognition (ASR) could facilitate standardisation and save time. A well designed and safely integrated tool could provide a standardised framework for all forces, speed up transcription, effectively facilitate human checking and potentially support playback in court (provided there is an appropriate technological set up). However, there are increasing examples of the risks involved with poor deployment of AI-based ASR systems in the public sector (e.g. reports from [Ada Lovelace Institute](#) and [Centre for Policing Research and Learning](#)).

Speech is complicated and context is important

Individual speaker patterns can impact the performance of speech technologies. Our research has shown this can happen with both speaker and speech recognition systems. For automatic speech recognition (ASR), speakers who mumble, speak quickly or are particularly disfluent can all be more challenging for speech recognition systems. These patterns are not unusual within the wider population. In law enforcement, the individual is central and understanding how and why ASR systems might vary should be the first step when aiming to safely deploy these systems. Conversational topics and speaking contexts within law enforcement are unlikely to be well represented in how off-the-shelf ASR systems have been trained. Law enforcement should be able to fine tune systems and assess performance within a more appropriate (conversational) context.

When are errors a problem?

The most common metric used to assess ASR systems is word error rate. This is a simple metric which calculates the difference between two transcripts - a reference transcript and the ASR transcript. It is not possible to calculate error rates or meaningfully assess system performance without a reference transcript with which to compare outputs. The word error rate is simply the number of differences (or 'errors') between the two transcripts. These include insertions (ASR transcript inserts content which was not in the reference transcript), deletions (ASR deletes content included in the reference) or substitutions (ASR and reference transcript have transcribed content differently).

Whilst this metric is useful to assess the overall variability between transcripts, it does not distinguish mundane variation from meaningful errors: an 'error' could be formatting difference in the presentation of a contraction (e.g. *could not* v. *couldn't*), or a significant change in meaning (e.g. *gum in the car* v. *gun in the car*). We have been exploring more nuanced methods of assessing performance: considering whether the meaning changes between human and ASR produced transcripts, whether

the transcript still makes sense and the degree of phonetic similarity between the reference and ASR transcripts. We have found that for all speakers, around a third of the errors result in potentially significant changes in meaning. In order to mitigate the risk of these meaningful errors a robust human-centred process needs to be developed.

A whole systems approach: The BIG argument for safe AI in policing

We have found that simply asking humans to review an ASR transcript does not guarantee that meaningful errors will be identified. Further, there are important decisions to be made prior to the ASR transcript being produced. How to keep the human in-the-loop is not a straightforward question. There are opportunities to standardise the process for how police interviews are transcribed whilst also safely taking advantage of ASR technology.

Given the focus of ASR in policing and the severity of potential errors for both individuals and public trust, a high level of assurance is required. We recommend adopting best practices from safety-critical domains, specifically the use of safety cases. These structured, evidence-based arguments, widely used in defence and transport, provide a transparent rationale for safe deployment, clarifying risks and mitigation controls for all relevant stakeholders. [The Balanced, Integrated and Grounded \(BIG\) framework](#) offers a whole system approach to constructing these safety cases for AI systems of varying complexity. This helps move the focus from a narrow, technology-centric narrative toward a sociotechnical solution that accounts for the human and organisational complexities of policing.

Further information

Contact us:

- Jessica Wormald - jessica.wormald@york.ac.uk
- James Tompkinson - james.tompkinson@york.ac.uk
- Lauren Harrington - lauren.harrington@york.ac.uk
- Ibrahim Habli - ibrahim.habli@york.ac.uk

DOI: <https://zenodo.org/records/18926595>