Audio Enhancement & Transcription Interlaboratory Comparisons

Katherine Earnshaw, Bryony Nuttall & Richard Rhodes
THE FORENSIC VOICE CENTRE

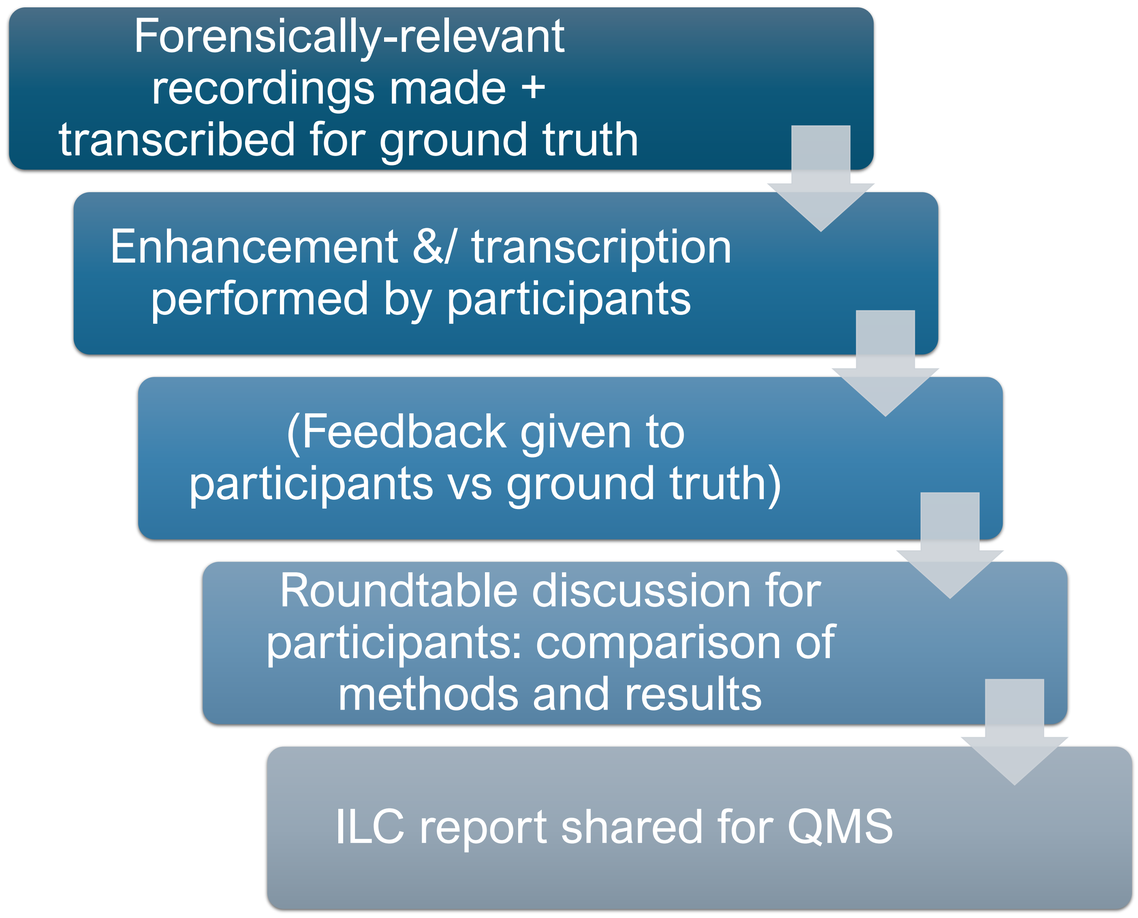
Context

The UK Forensic Science Regulator's Code of Practice - and other international guidance - requires forensic providers in the UK to assure the quality of results, including taking part in proficiency tests (PT), interlaboratory comparisons (ILC) and collaborative exercises (CE).

While there are institutions that provide speech and audio related PT schemes for its members such as ENFSI (e.g. Fröhlich et al. - oral session 1); there are still relatively few opportunities to participate in these exercises across the UK and internationally.

For this reason, we are proposing to run two relatively small-scale ILC exercises: one in audio enhancement and one in forensic transcription. The purpose of these exercises is for practitioners to share methods and results and to discuss best practices for these FSAs. The findings may also assist in demonstrating practitioner competence, method development and method validation.

ILC process



We are designing 2 interlaboratory comparison (ILC) exercises for audio enhancement & transcription

We have collaborated with University of York on a new set of recordings; these are forensically-realistic and have available ground truth data

We invite feedback on the design and assessment criteria

We are advertising for ILC participants [sign up via the QR code]:

Recordings:

- Multi-speaker recordings with a 'forensic feed' and a 'ground-truth' feed from each speaker via head-mics
- From 3 forensically-relevant scenarios:
  - Conversation in a noisy road environment (CCTV/doorbell)
  - Telephone call to emergency services; speech in background
  - Covertly-recorded conversation in a café
- Ground-truth (GT) for speech content and speaker identity will be available in GT transcripts

Questions:

- We invite feedback on how to make this exercise most useful, e.g.:
  - *How do we assess the extent to which the intelligibility of the speech in the enhanced recordings has been improved?*
  - *Should we ask lay listeners or experts to rate the improvement?*
  - *What is the best method to 'grade' transcripts and/or the types of errors that occur? (Given WER is a relatively blunt instrument)*
  - *Could AI be used to rate transcript accuracy?*

Feel free to ask us questions now or later via: katherine.earnshaw@forensicvoicecentre.com