Database selection for forensic voice comparison

Geoffrey Stewart Morrison[*], *Felipe Ochoa*

Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications,
University of New South Wales, Sydney, Australia
*geoff-morrison@forensic-voice-comparison.net

Defining the relevant population to sample is an important issue in data-based implementation of the likelihood-ratio framework for forensic voice comparison. A forensic likelihood ratio is the answer to a specific question which depends on the prosecution and defence hypotheses and the circumstances of the case. If an inappropriate background sample is selected the likelihood ratio produced by the forensic-voice-comparison system will not answer the question asked, and if an inappropriate test sample is selected the results of validity and reliability tests will not be informative as to the performance of the system under conditions reflecting those of the case at trial.

We present a logical argument that because an investigator or prosecutor only submits suspect and offender recordings for forensic analysis if they sound sufficiently similar to each other, the appropriate defence hypothesis for the forensic scientist to adopt will usually be that:

> the suspect is not the speaker on the offender recording but is a member of a population of speakers who sound sufficiently similar to the offender recording that an investigator or prosecutor would submit recordings of these speakers for forensic comparison with the offender recording.

We argue that categories such as speaker gender and dialect spoken are not relevant per se.

We propose a procedure for selecting background, development, and test databases using a panel of human listeners, who select voice recordings on the basis of their perceived similarity to the actual offender recordings. We briefly demonstrate software we have developed to aid in this process (Fig. 1). We argue that for this task:

1. Listeners should as far as possible be the same type of listener as the listener who made the original decision to submit the suspect and offender recording for analysis.
2. Listeners should listen under conditions reflecting those of the case under investigation, including any speaking-style and recording-channel mismatched between the suspect and offender recordings.

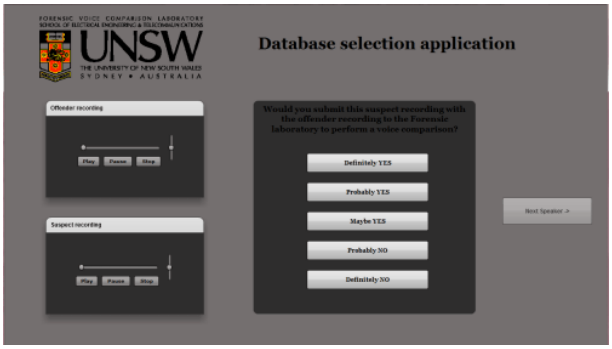

Figure 1: Screenshot of software for selection of offender-recording-similar recordings by human listeners.

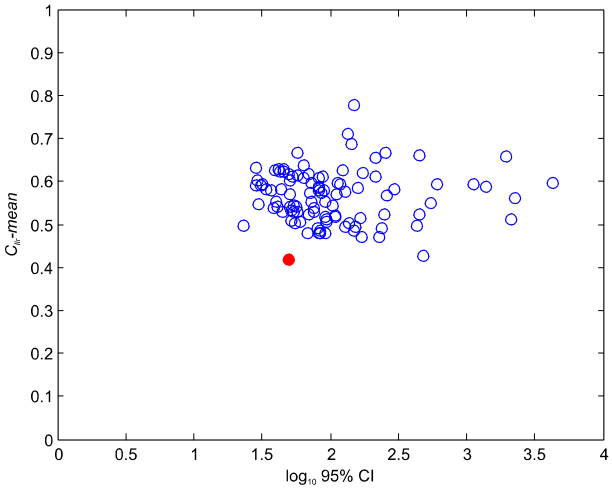

Figure 2: Comparison of the validity and reliability of the forensic-voice-comparison system when the automatic database-selection procedure is used for the background and development databases (red filled circle) and when the speakers in the background and development databases are chosen at random (blue unfilled circles).

We also present results of an empirical test of an automatic procedure inspired by the above. Although the automatic procedure is not entirely consistent with the logical arguments and human-listener procedure, it serves as a proof of concept for the importance of database selection. A forensic-voice-comparison system using the automatic database-selection procedure outperformed systems with random database selection (Fig. 2).

This presentation is based in part on Morrison, Ochoa, & Thiruvaran (2012). To aid in discussion, conference attendees are encouraged to read this before the conference.

Reference

Morrison, G. S., Ochoa, F., & Thiruvaran, T. (2012). Database selection for forensic voice comparison. In *Proceedings of Odyssey 2012: The Language and Speaker Recognition Workshop, Singapore* (pp. 62–77). International Speech Communication Association.