

A game-based approach to eliciting and evaluating likelihood ratios for speaker recognition

Vincent Hughes¹, Carmen Llamas¹, and Thomas Kettig¹

¹Department of Language and Linguistic Science, University of York, UK
{vincent.hughes|carmen.llamas|thomas.kettig}@york.ac.uk

Oral presentation (20 mins) or poster (no preference)

Not eligible for 'Best Student Paper Award'

This presentation describes a bespoke computer game which doubles as a sociolinguistic experiment that elicits and evaluates likelihood ratio (LR)-like scores from human non-expert listeners in a speaker recognition task. Previous work has examined human speaker recognition performance with unfamiliar voices; however, very little research has attempted to compare and combine such results with those of automatic speaker recognition (ASR) systems due to the considerable challenges in extracting judgements from humans that are both logically and empirically comparable with outputs of data-driven systems. Our project, *Humans and Machines: Novel Methods for Assessing Speaker Recognition Performance*, aims to provide a framework for combining LR-like judgements from human listeners with the output of an ASR system. We also explore sources of cognitive bias on human responses. We focus here on our methodology and experimental design.

Over the course of gameplay, participants encounter a series of voice comparisons. In each comparison, participants first listen to one stimulus (the nominal 'criminal' or 'unknown' sample) and rate on a 0–100 scale how typical they consider the voice to be relative to other speakers of the same accent. They are then presented with a second stimulus (the nominal 'suspect' or 'known' sample) and asked to provide a judgement of the similarity between this and the first sample on a 0–100 scale. Finally, participants indicate on a 0–100 scale whether they think the two voices belong to the same speaker.

The stimuli used in the game are 10-second audio samples of the speech of young, white, male British English speakers extracted from two corpora: the *Dynamic Variability in Speech* corpus (Nolan et al. 2009) and *The Use and Utility of Localised Speech* corpus (Llamas, French & Watt 2016-19). In addition to demographic information, participants initially provide judgements on a 0–100 scale to indicate how familiar they are with the three accents represented by the stimuli: Newcastle, Middlesbrough, and Standard Southern British English (SSBE).

The first stimulus in each pair is a far-end, landline telephone recording while the second stimulus is a high-quality studio recording; this channel mismatch replicates common conditions within forensic voice comparison casework. Half of the pairs that each participant encounters are same-speaker (SS) pairs, and half are different-speaker (DS) pairs; DS pairs are always matched for regional accent. We can thus explore the effect of self-identified familiarity with an accent on speaker recognition performance.

Furthermore, in order to probe how listeners' LR scores might be affected by situating the task in a legal context, the game is comprised of several levels in which: 1) no legal context is supplied; 2) participants are immersed in their role on a 'jury of the future'; 3) participants are primed with extralinguistic evidence; 4) participants are given advice from an 'expert witness'.

LR-like scores are calculated by dividing average listener similarity and typicality judgements. Tests of initial prototypes have confirmed that listener judgements about similarity and typicality do produce LR-like scores that can be calibrated and evaluated like any other speaker recognition system.

References

Llamas, C., Watt, D. and French, P., 2016-2019, *The use and utility of localised speech forms in determining identity: forensic and sociophonetic perspectives*. UK Economic and Social Research Council (ES/M010783/1)

Nolan, F., McDougall, K., De Jong, G., & Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech Language and the Law*, 16(1), 31-57.