To errrr is human, for guilt divine

Paul Foulkes & Vincent Hughes

University of York & NZILBB

The voice has long been recognised as a marker of human identity, albeit an imperfect one. Analysis of voice is now widely undertaken in forensic cases. This usually involves comparison of the properties of a criminal's voice, recorded in the course of a crime, with those of a suspect, recorded in police custody. The aim of the comparison is to aid the court in determining the likelihood that the recordings were made by the same person, and thus that the suspect was indeed the criminal.

Two broad methodological approaches are used: (1) automatic speaker recognition (ASR), and (2) 'componential' phonetic-linguistic analysis (CPLA). ASR has been developed largely in engineering and physics. Put simply, analysis of voice tends to be holistic and highly abstract, with little direct mapping between the coefficients used for analysis and phonetically defined properties of the speech signal. By contrast, CPLA is grounded on well-established and uncontroversial methods derived from phonetics and linguistics, such as vowel formant and f0 analysis.

Surprisingly few attempts have been made to compare and contrast ASR and CPLA, both of which have widely recognised advantages and disadvantages. This is precisely the aim of our ongoing project, *Voice and Identity*.[1] We are exploring the performance of the different methods on the same data to assess their relative strengths, the consistency of their results and error patterns, and thus the potential for phonetic and automatic methods to be integrated. The ultimate aim is to improve methods in forensic voice comparison, taking a major step towards the development of a methodology that is more transparent, validated, and replicable.

What both ASR and CPLA need for robust performance is good variables that serve to discriminate reliably between individuals. This means features of speech, voice or language that display low within-speaker but high between-speaker variability. In this presentation I will outline a range of analyses conducted on hesitation markers, *uh* (or *er*) and *um* (*erm*), phonetically ~ [əː(m)]. Hesitations are predicted to be good variables, in that they are frequent, easy to measure, unconscious, and relatively free of coarticulatory effects.

We analysed hesitations from 73 speakers of standard British English using standard acoustic methods (vowel midpoints, formant trajectories, and durations). The acoustic data were then compared with ASR results gathered from the whole recordings. Quantitative data were used in speaker-discrimination tests (paired samples in which the true identity of speakers is known). The analyses support the hypothesis that hesitations are good variables: we found very high rates of successful discrimination, and few errors. However, two key points emerge. First, the errors made by ASR and CPLA were different, and there was no correlation between results. This indicates that the methods are sensitive to different vocal properties. Second, by fusing ASR and CPLA results together, the results improved significantly compared with the results from each method separately. Both observations suggest that integrating traditional linguistic variables into an ASR system is beneficial for establishing robust analysis of forensic recordings.

[1] Funded by the UK Arts and Humanities Research Council, grant # AH/M003396/1, 2015-2018.