Developing the vocal profile analysis scheme for forensic voice comparison

Peter French[1,2], Paul Foulkes[1,2], Philip Harrison[1,2], Vincent Hughes[1] and Eugenia San Segundo Fernandéz[1]

[1]J P French Associates, York, UK.
[2]Department of Language and Linguistic Science, University of York, UK.

Forensic voice comparison (FVC) involves analysing the speech of an unknown offender and a known suspect to aid the court in determining whether the voices belong to the same or different speakers. In the UK, FVC casework is conducted using a combination of auditory and acoustic linguistic-phonetic analysis. For forensic phoneticians voice quality (VQ) is considered one of the most valuable features for distinguishing between speakers (Nolan 2005). In a recent survey of practitioners, 94% report examining VQ, with 61% using a recognised framework such as Laver's VPA (Laver 1980).

However, for several reasons these frameworks are not widely used for full systematic VQ analysis. They require considerable training to use and can be relatively difficult to employ with typical forensic recordings (e.g. with poor technical quality, emotional speech, telephone transmission; Nolan 2005). The modified VPA used by J P French Associates in casework contains 38 separate dimensions, but some form a continuum (e.g. close jaw ~ open jaw) or are correlated (e.g. retracted tongue body ~ pharyngeal constriction). The inter-rater reliability of VPAs conducted using forensic material has not been reported, and the degree of within-speaker variability across multiple recordings has not been tested empirically. The present study addresses these issues and proposes developments for the VPA in FVC.

Data were taken from the 100 young male RP speakers in the DyViS corpus. High quality, near-end recordings of a telephone conversation from DyViS Task 2 were analysed. VQ was assessed by three analysts independently, using a modified version of the Laver (1980) VPA protocol. This had three rather than six scalar degrees for 'present' features, i.e. ignoring degrees which would be considered pathological. Features were scored on a 0 – 3 scale, where 0 means that a setting is absent and values of 1 – 3 mean the setting is present to an increasing degree.

Three sets of tests were performed. Firstly, correlations between each of the 38 VPA features were assessed. Empirical correlations were compared against predictions from phonetic theory to identify features which could be collapsed to simplify the scheme. Secondly, the inter-rater agreement across the three analysts was assessed. Initial results suggest a high degree of convergence across analysts, with relatively little disagreement about presence or absence of certain settings and typical differences of only one scalar degree for 'present' settings. Finally, a subset of 10 VPAs from Task 2 were compared with VPAs for the same speakers assessed from DyViS Task 1 (mock police interview) recordings in order to evaluate within- and between-speaker variation. Pairs of same- and different-speaker VPAs were reduced to a distance measure, and the distances used as speaker discrimination scores from which it is possible to generate error rates. These results represent a step towards formalising VQ analysis in FVC, and modifying the VPA scheme for wider forensic use.

References

Laver, J. (1980) *The Phonetic Description of Voice Quality*. Cambridge: CUP.

Nolan, F. (2005) Forensic speaker identification and the phonetic description of voice quality. In W.J. Hardcastle & J. Mackenzie Beck (eds.) *A Figure of Speech: A Festschrift for John Laver*. Mahwah NJ: Lawrence Erlbaum. pp. 385-411.