

Likelihood ratio-based forensic voice comparison with higher level features: research and reality

Phil Rose

Australian National University Emeritus Faculty

philjohn.rose@gmail.com <http://philjohnrose.net>

Abstract

Examples are given of forensic voice comparison with higher level features in real-world cases and research. A pilot experiment relating to estimation of strength of evidence in forensic voice comparison is described which explores the use of higher-level features extracted over a disyllabic word as a whole, rather than over individual monosyllables as conventionally practiced. The trajectories of the first three formants and tonal F0 of the hexaphonic disyllabic Cantonese word *daihyat* ‘first’ from controlled but natural non-contemporaneous recordings of 23 male speakers are modeled with polynomials, and multivariate likelihood ratios estimated from their coefficients. Evaluation with the log likelihood ratio cost validity metric C_{llr} shows an optimum performance is obtained, surprisingly, with lower order polynomials, with F2 requiring a cubic fit, and F1 and F3 quadratic. Fusion of F-pattern and tonal F0 results in considerable improvement over the individual features, reducing the C_{llr} to ca. 0.1. The forensic potential of the *daihyat* data is demonstrated by fusion with three other Cantonese higher-level features: the F-pattern of /i/, short-term F0, and syllabic nasal cepstral spectrum, which reduces the C_{llr} still further to 0.03. Important pros and cons of higher-level features and likelihood ratios are discussed, the latter illustrated with data from Japanese, and three varieties of English in real forensic casework.

Index Terms: Forensic voice comparison, likelihood ratio, Cantonese, higher-level features, F-pattern trajectories, tonal F0 trajectory, short term F0, segmental cepstrum.