

Beyond the long-term mean: Multivariate likelihood ratio-based FSR using F0 distribution parameters

Yuko Kinoshita¹, Shunichi Ishihara², and Phil Rose³

¹*School of Languages and International Studies, University of Canberra, Australia*

²*Faculty of Asian Studies, The Australian National University, Australia*

shunichi.ishihara@anu.edu.au

³*Phonetics Laboratory, Faculty of Arts, The Australian National University, Australia*

philip.rose@anu.edu.au

Long-term mean F0 (LTF0) is a popular parameter in Forensic Speaker Recognition (FSR). Its popularity probably stems from promising results in early SR research (eg. Atal, 1972, Sambur, 1975), together with its conforming to three of Nolan's (1983) desiderata for FSR parameters, namely: robustness, measurability, and availability. On the other hand, many (eg. French, 1994, Braun, 1995) have noted that a single speaker can show large variation in this parameter from occasion to occasion, and even on the same occasion. Furthermore, Kinoshita (2005) demonstrated that because of its poor variance ratio (of between- to within-speaker variance), LTF0 shows very poor strength of evidence, typically generating Likelihood Ratios of effectively unity.

Forensic analysis of F0 has concentrated so far on its long-term mean and standard deviation. This paper explores the forensic discriminatory potential of combinations of parameters from LTF0 distributions. A multivariate likelihood ratio (Rose et al., 2006) is used as discriminant function on non-contemporaneous spontaneous speech samples of 201 Japanese male speakers from the Corpus of Spontaneous Japanese (Maekawa et al., 2000). Results are promising. With six features (mean, SD, kurtosis, skew, modal F0, probability density at mode) we obtain an EER of 10% (out of 80,400 different-speaker trials, ca. 72,000 are correctly discriminated, with ca. 182 out of 201 same-speaker trials correct). The EER is also well calibrated.

References

- Atal, B. S. (1972). Automatic Speaker Recognition Based on Pitch Contour. *JASA*, 52 6 (part 2), 1687-1697.
- Braun, A. (1995). Fundamental Frequency - How Speaker Specific Is It? *BEIPHOL Studies in Forensic Phonetics* (64), 9-23.
- French, P. (1994). An overview of forensic phonetics with particular reference to speaker identification. *Forensic Linguistics*, 169-181.
- Kinoshita, Y. (2005). Does Lindley's Ir estimation formula work for speech data?: investigation using long-term f0. *International Journal of Speech Language and the Law*, 12(2), 235-254.
- Maekawa, K., Koiso, H., Furui, S. & Isahara, H. (2000). Spontaneous speech corpus of Japanese. Paper presented at the Second International Conference of Language Resources and Evaluation (LREC2000), Athens.
- Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
- Rose, P., Kinoshita, Y. & Alderman, T. (2006) Realistic Extrinsic Forensic Speaker Discrimination with the Diphthong /ai/. *Proc. 11th Australasian Intl. Conf. on Speech Science & Technology*, 329-334.
- Sambur, M. R. (1975). Selection of acoustic features for speaker identification. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-23(2), 178-182.