

Investigating the evidential value of filled pauses in cross-language forensic voice comparison

Filled pauses (FPs) are known for their speaker-specific characteristics, displaying low within-speaker variability and demonstrating promising speaker-discriminatory power in forensic voice comparison (FVC) (e.g., Hughes et al., 2016). However, the evidential value of FPs in cross-language FVC is unclear. This study examines the evidential value of FPs among trilingual Hong Kong Cantonese speakers (i.e., Cantonese, English, Mandarin). 19 tokens of the FP *uh* in three mock police interviews were analysed for 21 female speakers with at least two tokens per speaker per language. Midpoint formant values (F1, F2, F3) of FPs were used to generate LR-like scores for every combination of same- and different-speaker comparisons using the multivariate kernel density (Aitken & Lucy, 2004) implemented with cross-validation. Scores were then calibrated using cross-validated logistic regression (Brümmer et al., 2007). Nine different language mismatch conditions were tested. Results suggest poor performance in general with C_{lr} ranging from 0.8 to 1. The best performance was found in the condition where Mandarin was used as the suspect sample, English as the offender sample and reference data ($C_{lr} = 0.87$). This might be due to the similar vowels the trilingual speakers used in their English and Mandarin filled pauses, resulting in a lower within-speaker variability. A practical implication of this work is that forensic practitioners should take into account different types of bilingualism when handling multilingual cases. In the current case of Cantonese-English-Mandarin trilinguals, the L1-transfer effect was not found in the FPs of their English (L2) and Mandarin (L3).

References

- Aitken, C. G. G., & Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(1), Article 1. <https://doi.org/10.1046/j.0035-9254.2003.05271.x>
- Brümmer, N., Burget, L., Cernocky, J., Glembek, O., Grezl, F., Karafiat, M., van Leeuwen, D. A., Matejka, P., Schwarz, P., & Strasheim, A. (2007). Fusion of Heterogeneous Speaker Recognition Systems in the STBU Submission for the NIST Speaker Recognition Evaluation 2006. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(7), Article 7. <https://doi.org/10.1109/TASL.2007.902870>
- Hughes, V., Wood, S., & Foulkes, P. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech Language and the Law*, 23(1), Article 1. <https://doi.org/10.1558/ijssl.v23i1.29874>