Statistical models for forensic voice comparison from a phonetic acoustic approach.

Fernanda López-Escobedo [1], *N. Sofía Huerta-Pacheco*[2]

[1]flopeze@unam.mx, Escuela Nacional de Ciencias Forenses, Universidad Nacional Autónoma de México

[2]nshuerta@enacif.unam.mx, CONACYT - National School of Forensic Sciences, Universidad Nacional Autónoma de México

In this work, two models developed by Rose et al. (2004) and Morrison (2011) were implemented for acoustic-phonetic data from speech recordings. These methods have been commonly used to assess the strength of evidence from an auditory-acoustic-phonetic approach. The results of both models are numerical values that can be used to evaluate the evidence's strength in the likelihood ratio framework.

In many real-world cases, the amount of voice data is not sufficient to use models commonly used in automatic speaker recognition, such as Gaussian mixture models (GMMs) or Deep Neural Network (DNN). For this reason, we chose to test the Rose et al. (2004) and Morrison (2011) models and explore their performance when only a small number of samples are available. The model proposed by Rose et al. (2004) represents the distribution of the variables with normal curves because, after studying the behavior of the acoustic parameters in a corpus of 60 Japanese speakers, they were found not to have a distribution sufficiently far from normal to warrant non-parametric modeling. In contrast, the model proposed by Morrison (2011) assumes a distribution of variables with normal curves when measured across samples of the same speaker, but unlike Rose et al. (2004), it assumes a nonparametric distribution of variables when measured across different speakers.

In this work, a speech corpus was collected from 27 female speakers with an average recording time of approximately 2:13 minutes. Each vowel was manually segmented, and different metrics of the first four formants (F1, F2, F3, and F4) of the five Spanish vowels /a/, /e/, /i/, /o/, and /u/ were analyzed. After a descriptive analysis of the data and due to the high variability, it was decided to remove outliers with a filter based on 95% interval confidence. The likelihood ratio values obtained in the case of the same speaker pairs of recordings are in agreement in both models. This shows that the results of both models are consistent and can be used when the number of speech samples is limited.

Keywords: Forensic voice comparison, Likelihood ratio, Acoustic-phonetic data

Bibliography

Rose, P., Lucy, D., & Osanai, T. (2004). Linguistic-Acoustic Forensic Speaker Identification with Likelihood Ratios from a Multivariate Hierarchical Random Effects Model-A Non-Idiot's Bayes' Approach. Proceedings of the 10th Australian International Conference on Speech Science and Technology.

Morrison, G. S. (2011). A comparison of procedures for the calculation of forensic likelihood ratios from acoustic–phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model–universal background model (GMM–UBM). Speech Communication, 53(2), 242-256.