Forensic voice comparison: Older sister or younger sister?

Cuiling Zhang[1,2]*, Geoffrey Stewart Morrison*[3,4]*, and Ewald Enzinger*[5]

[1]*School of Criminal Investigation, Southwest University of Political Science & Law, Chongqing, China*

[2]*Chongqing Institutes of Higher Education Key Forensic Science Laboratory, Chongqing, China*

cuiling-zhang@forensic-voice-comparison.net

[3]*Forensic Speech Science Laboratory, Centre for Forensic Linguistics, Aston University, England, United Kingdom*

[4]*Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada*

[5]*Eduworks, Corvallis, Oregon, United States of America*

{geoff-morrison | ewald-enzinger}@forensic-evaluation.net

Introduction

The plaintiff in this case had recorded a mobile telephone conversation. Whether her interlocutor was the respondent or the respondent's younger sister was at issue. Five new telephone conversations with each sister were recorded using the plaintiff's mobile telephone. This provided known-speaker recordings under the same conditions as the questioned-speaker recording. The known-speaker recordings were used to train and test a forensic voice comparison system. The system was then used to evaluate the strength of evidence associated with the questioned-speaker recording: What is the probability of obtaining the acoustic properties of the voice on the questioned-speaker recording if it were produced by the older sister versus if it were produced by the younger sister?

Acoustic and statistical analysis

MFCCs were extracted every 10 ms from the speech of the speaker of interest in each recording. The 1st through 4th coefficients were used for statistical analysis. Fig. 1 shows the smoothed spectra corresponding to these measurements.

The data were transformed using canonical linear discriminant functions (CLDFs). This procedure finds new dimensions which maximize the ratio of between- to within-category variance (between to within-speaker variance). Only the first CLDF dimension was used for subsequent analysis. Even though there was no mismatch in recording conditions, there may still be some between-session variability, and this served as a mismatch compensation technique. Also, with only five data points for each category, we needed to fit a parsimonious model at the next stage of statistical analysis. By only using one dimension, the number of parameters for which values had to be estimated was reduced. Fig. 2 shows the resulting CLDF values and Gaussian distributions fitted with a pooled variance.

With so little data, we were concerned about having poor estimates of parameter values, which could lead to vast overestimation of the strength of evidence. For the actual case, we a priori chose one solution, but in this presentation we present the use of different statistical models which include shrinkage (these include additional analyses compare to those in an earlier publication [1] based on this case). As a point of comparison, we fitted a linear discriminant analysis model (LDA), which includes no shrinkage. We also fitted a Bayesian model with uninformative Jeffreys reference priors, we limited the maximum and minimum values from the LDA model using empirical lower and upper bounds (ELUB) [2], and we fitted a novel regularized logistic regression model (LogReg). The regularization consisted of a uniform distribution with a weight equivalent to 5 data points.

Results

A leave-one-out cross validation procedure was applied to the known-speaker recordings from the two sisters. Table 1 shows preliminary likelihood ratio / Bayes factor results.

The LDA procedure produced ridiculously large and small likelihood ratio values, which are not justifiable given the small amount of training data. The Bayesian analysis produced much more moderate Bayes factor values, but still questionable given the amount of training data. The ELUB procedure gave very conservative values. The regularized logistic regression procedure also gave conservative values, but these values could be above or below the ELUB values.

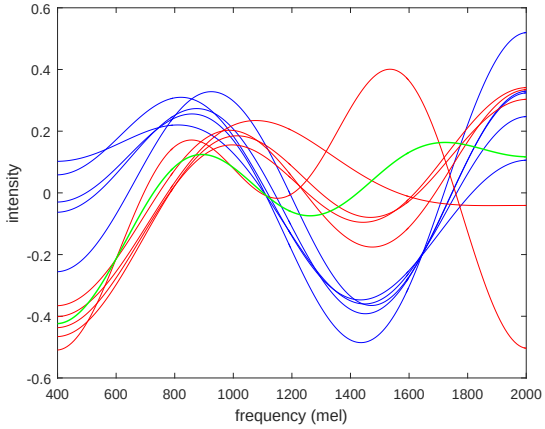

Figure 1 Mean smoothed spectrum for each known-speaker recording (blue curves: older sister. red curves: younger sister), and for the questioned-speaker recording (thicker green curve).

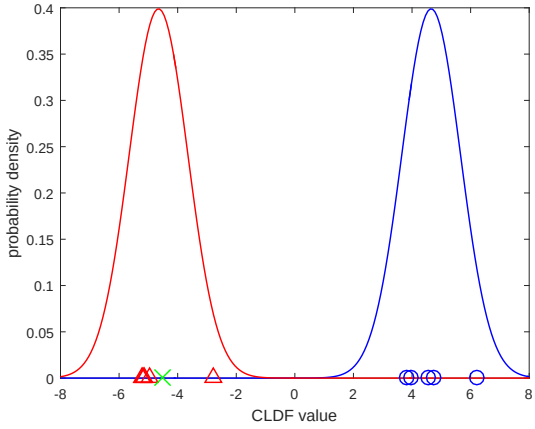

Figure 2 The CLDF values from each known-speaker recording (blue circles: older sister. red triangles: younger sister), and the questioned-speaker recording (green cross).

Table 1 Likelihood ratio values / Bayes factor values, $p(x|H_O)/p(x|H_Y)$, for each known-speaker recording and the questioned-speaker recording. Y: a younger sister recording. O: an older sister recording. Q: the questioned speaker recording.

speaker:	Y	Y	Y	Y	Y	O	O	O	O	O	Q
LDA:	2×10^{-20}	3×10^{-20}	6×10^{-24}	5×10^{-40}	2×10^{-22}	1×10^{12}	2×10^{8}	1×10^{16}	5×10^{16}	3×10^{40}	6×10^{-19}
Bayesian:	$\frac{1}{1{,}369}$	$\frac{1}{1{,}349}$	$\frac{1}{1{,}738}$	$\frac{1}{567}$	$\frac{1}{300}$	130	25	774	878	4,876	$\frac{1}{2{,}439}$
ELUB:	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	5	5	5	5	5	$\frac{1}{5}$
LogReg:	$\frac{1}{5.7}$	$\frac{1}{5.7}$	$\frac{1}{7.4}$	$\frac{1}{26.6}$	$\frac{1}{2.0}$	2.2	1.6	4.2	4.6	9.6	$\frac{1}{4.6}$

References

[1] Zhang, C., Morrison, G.S., Enzinger, E. (2016). Use of relevant data, quantitative measurements, and statistical models to calculate a likelihood ratio for a Chinese forensic voice comparison case involving two sisters. *Forensic Science International*, 267, 115–124. http://dx.doi.org/10.1016/j.forsciint.2016.08.017

[2] Vergeer, P., van Es, A., de Jongh, A., Alberink, I., Stoel, R.D. (2016). Numerical likelihood ratios outputted by LR systems are often based on extrapolation: When to stop extrapolating? *Science & Justice*, 56, 482–491. http://dx.doi.org/10.1016/j.scijus.2016.06.003