Mismatch compensation in the evaluation of evidence under conditions reflecting those of an actual FVC case

Ewald Enzinger
Acoustics Research Institute, Vienna, Austria   University of New South Wales, Sydney, Australia
Austrian Academy of Sciences

Introduction

This work demonstrates the application of mismatch compensation techniques in the evaluation of forensic evidence under conditions reflecting those of an actual FVC case. There is a mismatch in recording conditions:

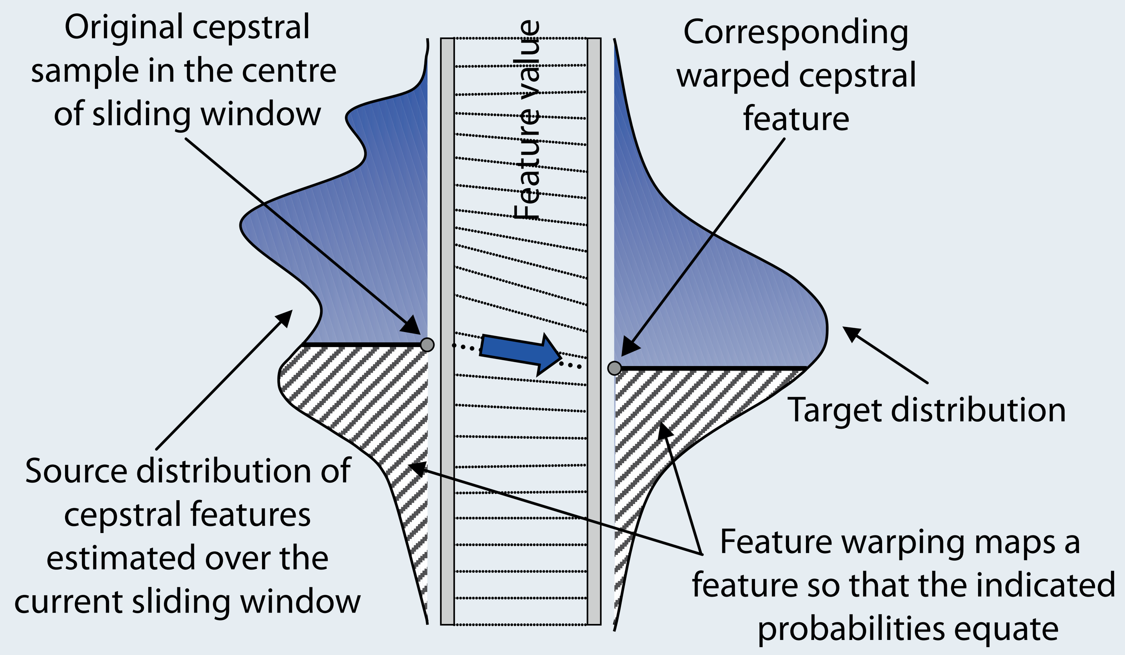
- Offender: landline telephone call made to a call centre (Telephone transmission, background noise, compression)
- Suspect: recording of police interview (reverberant room, ventilation noise, compression)

Forensic-voice-comparison system

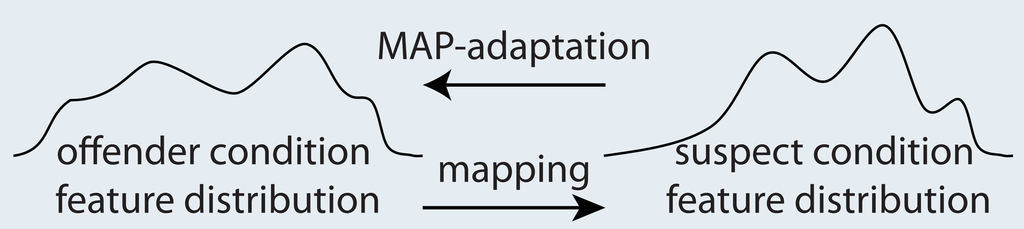
- 14 Mel frequency cepstral coefficients + Deltas
- Gaussian mixture model–universal background model (GMM-UBM), logistic-regression calibration [1, 2]

Recording-condition mismatch compensation

Feature warping [3]: Sequential feature streams are warped to more closely follow a predetermined distribution over a sliding time window. The aim is to compensate for slowly changing noise and channel influences that distort the feature distribution.



Probabilistic feature mapping [4]: Feature vectors from offender sample are transformed according to mapping between suspect- and offender-condition feature distributions.

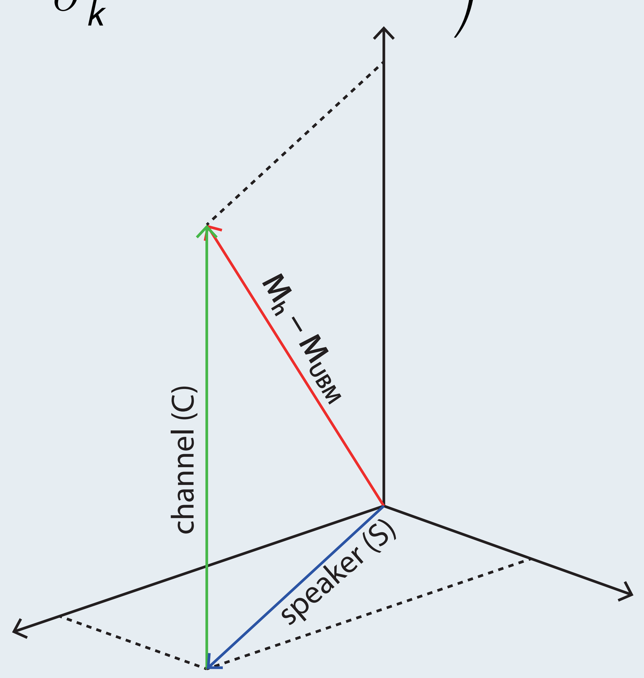


$$\widehat{\mathbf{x}}_t = \sum_{k=1}^{K} p(k|\mathbf{x}_t, \lambda^{\text{offender}}) \left((\mathbf{x}_t - \boldsymbol{\mu}_k^{\text{offender}}) \frac{\sigma_k^{\text{suspect}}}{\sigma_k^{\text{offender}}} + \boldsymbol{\mu}_k^{\text{suspect}} \right)$$

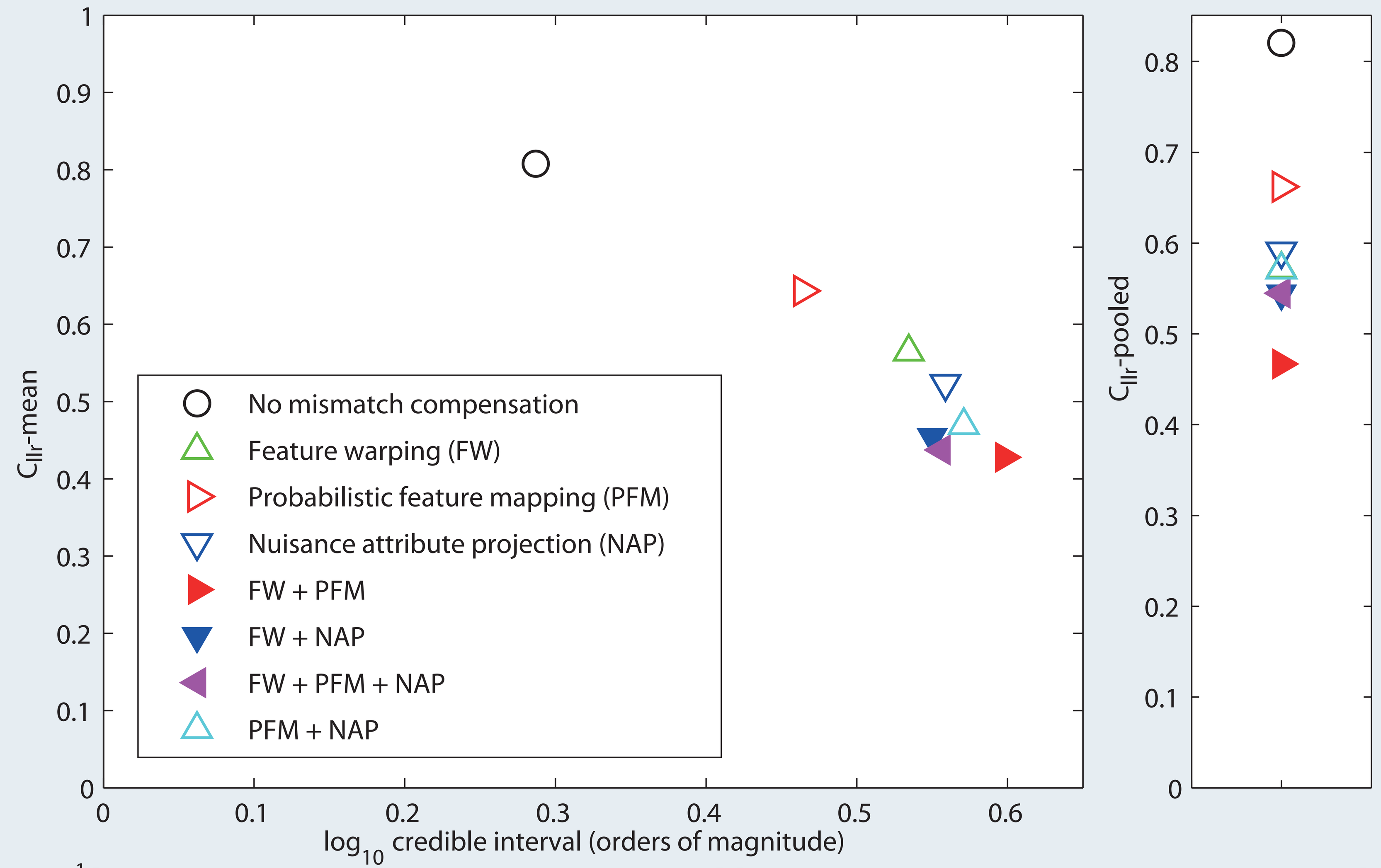
Nuisance attribute projection [5, 6]: Decomposition into speaker and channel space. The channel component $\mathbf{C}$ of a sample is estimated and used to transform each feature vector $\mathbf{x}_t$:

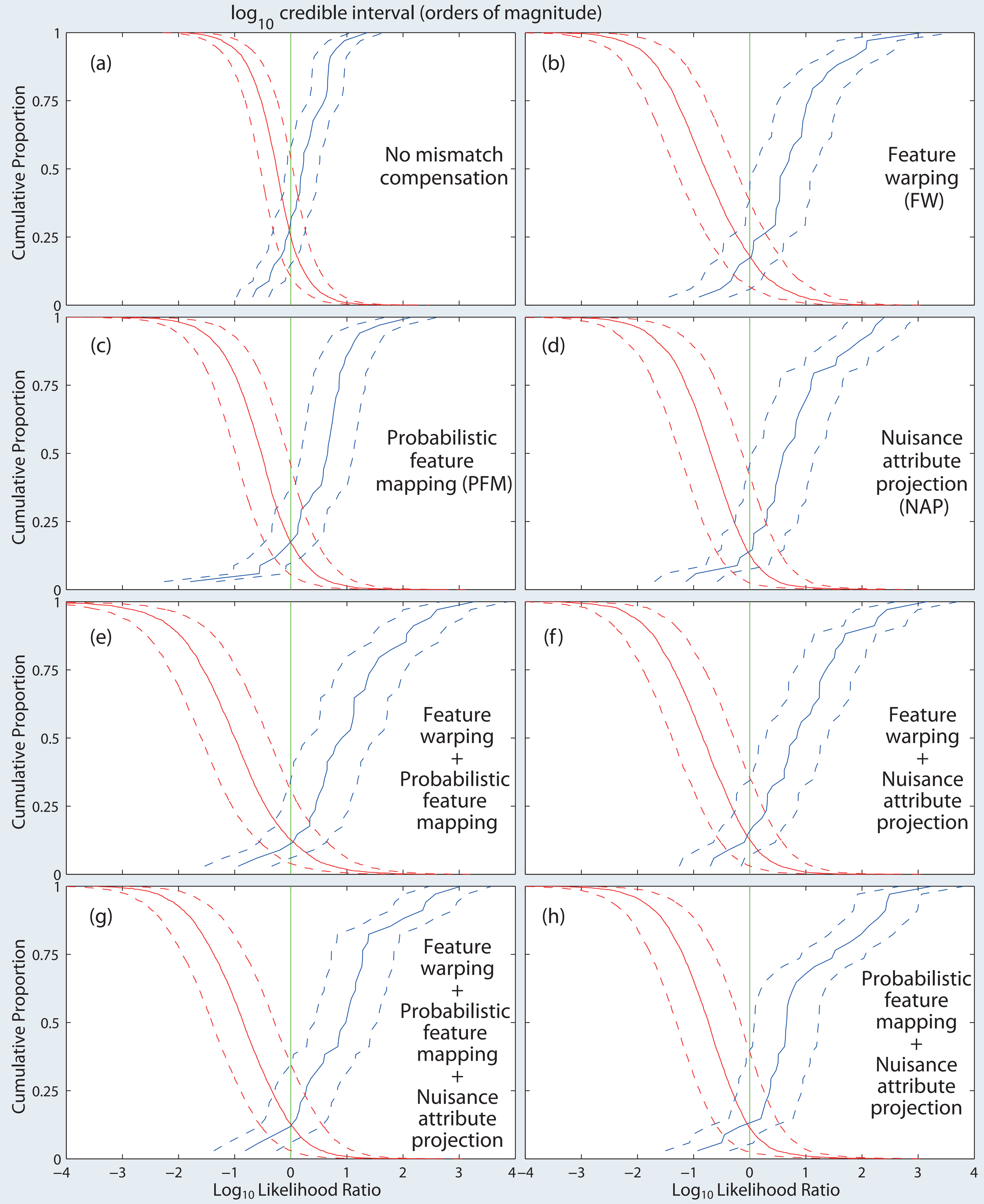
$$\mathbf{C} = \mathbf{U}\mathbf{U}^T(\mathbf{M}_h - \mathbf{M}_{\text{UBM}})$$

$$\widehat{\mathbf{x}}_t = \mathbf{x}_t - \sum_{k=1}^{K} p(k|\mathbf{x}_t, \lambda_h)\mathbf{c}_k$$



Comparison of validity and reliability – development set



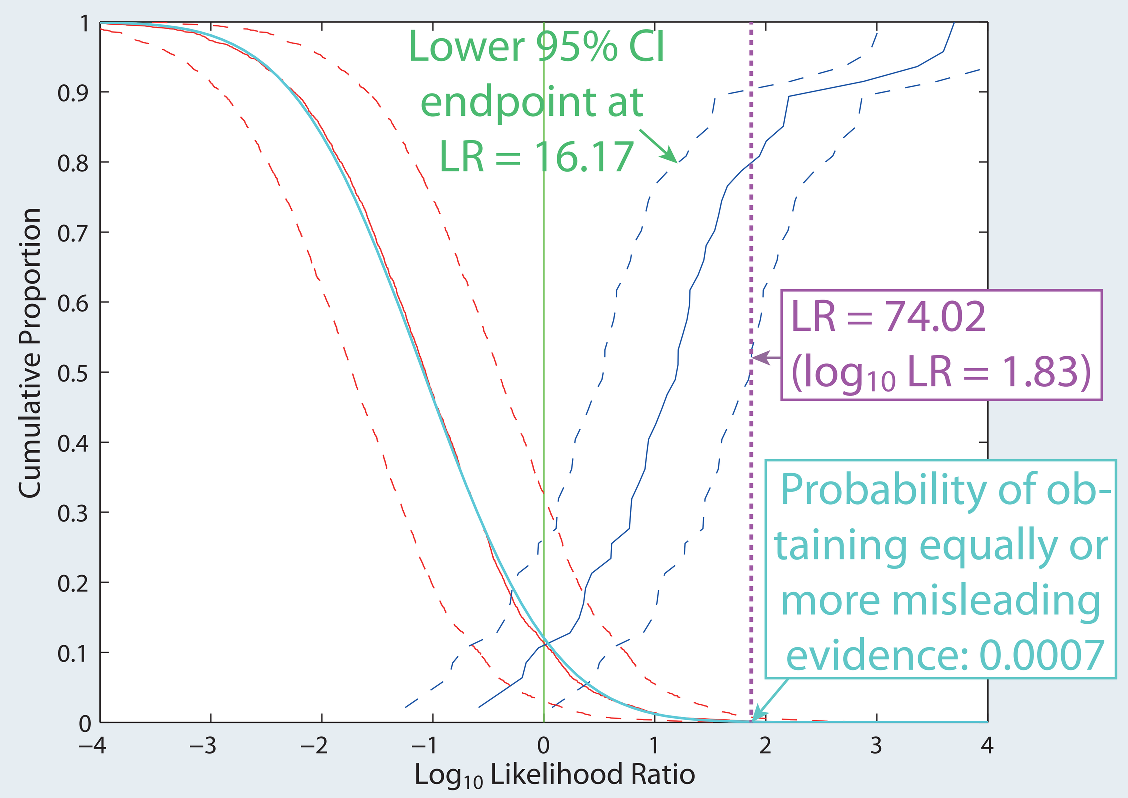


Evaluation of the likelihood ratio

Feature warping + probabilistic feature mapping are used in the final system.

Results from testing of validity and reliability (test set):

- $C_{llr} = 0.347$
- 95% credible interval = 0.66



Summary of evaluation results

- Based on the results of our analysis our best estimate of the strength of the evidence is that the probability of getting the acoustic properties of the offender sample is approximately 74 times greater had it been spoken by the suspect then had it been spoken by another speaker from the relevant population.
- We are 95% certain that the probability of getting the acoustic properties of the offender sample is at least 16 times greater had it been spoken by the suspect versus had it been spoken by another speaker from the relevant population.
- What is the probability of getting this strength of evidence (a value of 74) or greater if the speaker was actually not the suspect but another speaker from the relevant population? Based on our calculations, we estimate that this probability is 0.07%.

Conclusions

- Mismatch compensation greatly improves validity, while reliability moderately deteriorates
- Best performance: Feature warping + probabilistic feature mapping

References

- D. A. van Leeuwen and N. Brümmer, "An introduction to application-independent evaluation of speaker recognition systems," in *Speaker Classification I. Fundamentals, Features, and Methods* (C. Müller, ed.), pp. 330–353, Springer, 2007.
- G. S. Morrison, "Tutorial on logistic-regression calibration and fusion: Converting a score to a likelihood ratio," *Australian Journal of Forensic Sciences*, vol. 45, pp. 173–197, 2013.
- J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Proceedings of Odyssey 2001: The Speaker Recognition Workshop*, pp. 213–218, 2001.
- M.-W. Mak, K.-K. Yiu, and S.-Y. Kung, "Probabilistic feature-based transformation for speaker verification over telephone networks," *Neurocomputing*, vol. 71, no. 1-3, pp. 137–146, 2007.
- W. Campbell, D. Sturim, P. Torres-Carrasquillo, and D. Reynolds, "A comparison of subspace feature-domain methods for language recognition," in *Proc. Interspeech 2008*, pp. 309–312, 2008.
- B. Vesnicer and F. Mihelič, "The likelihood ratio decision criterion for nuisance attribute projection in GMM speaker verification," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 158, 2008.