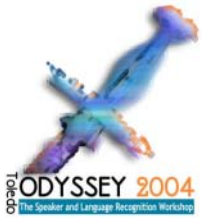


Handling Mismatch in Corpus-Based Forensic Speaker Recognition

Anil Alexander †, Filippo Botti ‡ and Andrzej Drygajlo †

† **Swiss Federal Institute of Technology Lausanne**
Signal Processing Institute
‡ **Université de Lausanne**
École de sciences criminelles

Outline

- Bayesian Interpretation in forensic automatic speaker recognition
- Mismatched recording conditions and its influence on the estimation of the strength of evidence
- Measuring discrimination in matched conditions
- Statistical estimation of mismatch
- Compensation of mismatched conditions of potential population databases
- Conclusion

Bayesian Interpretation in Forensic Automatic Speaker Recognition

- Evidence (E): score obtained comparing statistical model of suspect's voice and a questioned recording (trace)

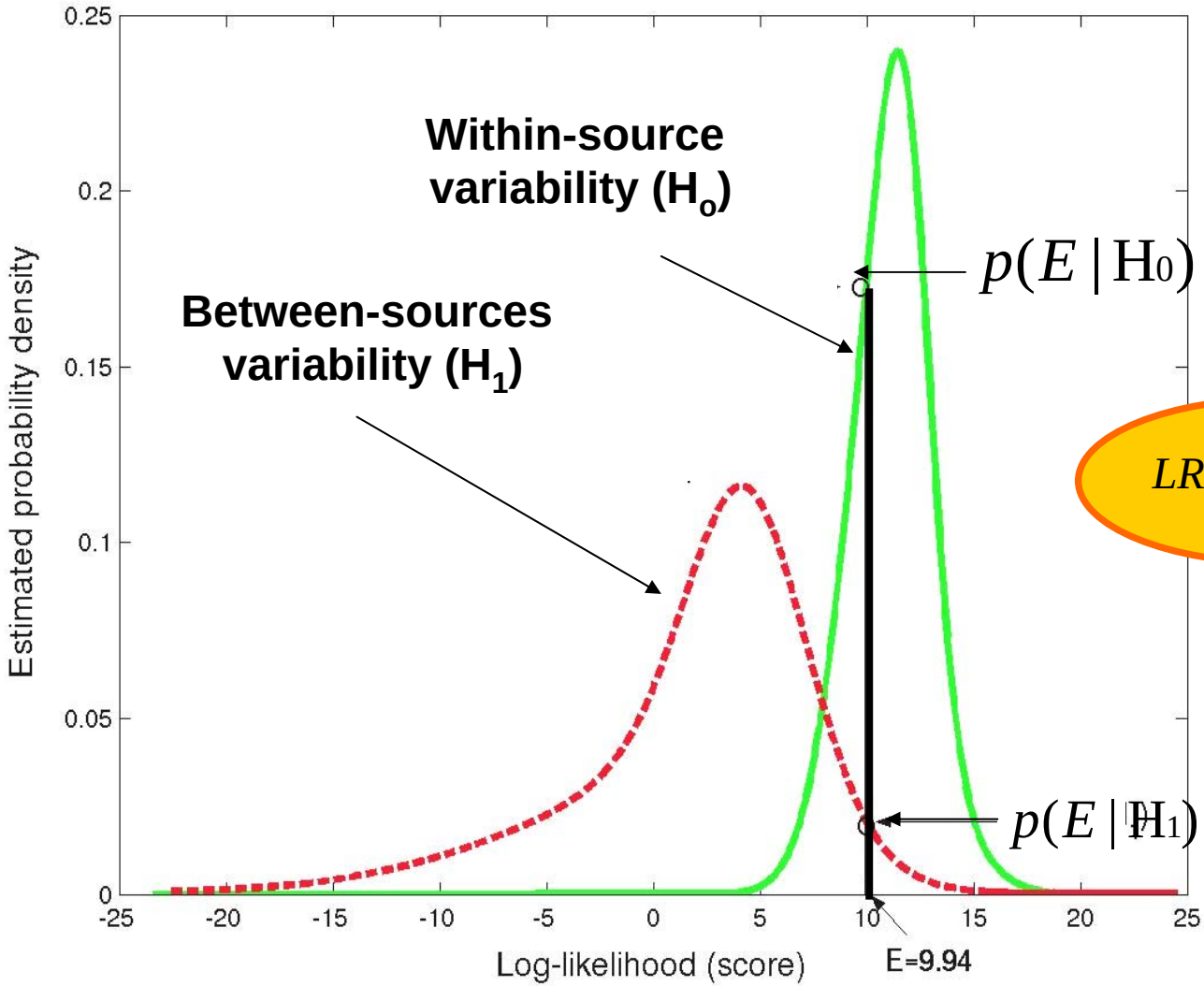
H_0 - Suspect is the source of the trace

H_1 - Another speaker is the source of the trace

$$LR = \frac{p(E \mid H_0)}{p(E \mid H_1)}$$

Likelihood Ratio (LR) is the relative probability of observing a particular score of E, with respect to two competing hypotheses

Likelihood Ratio



$$LR = \frac{p(E \mid H_0)}{p(E \mid H_1)}$$

Choice of Algorithms and Databases

Databases

- Suspect reference (R) and Suspect control (C) databases

  Often not possible to record the suspect in the conditions similar to the case

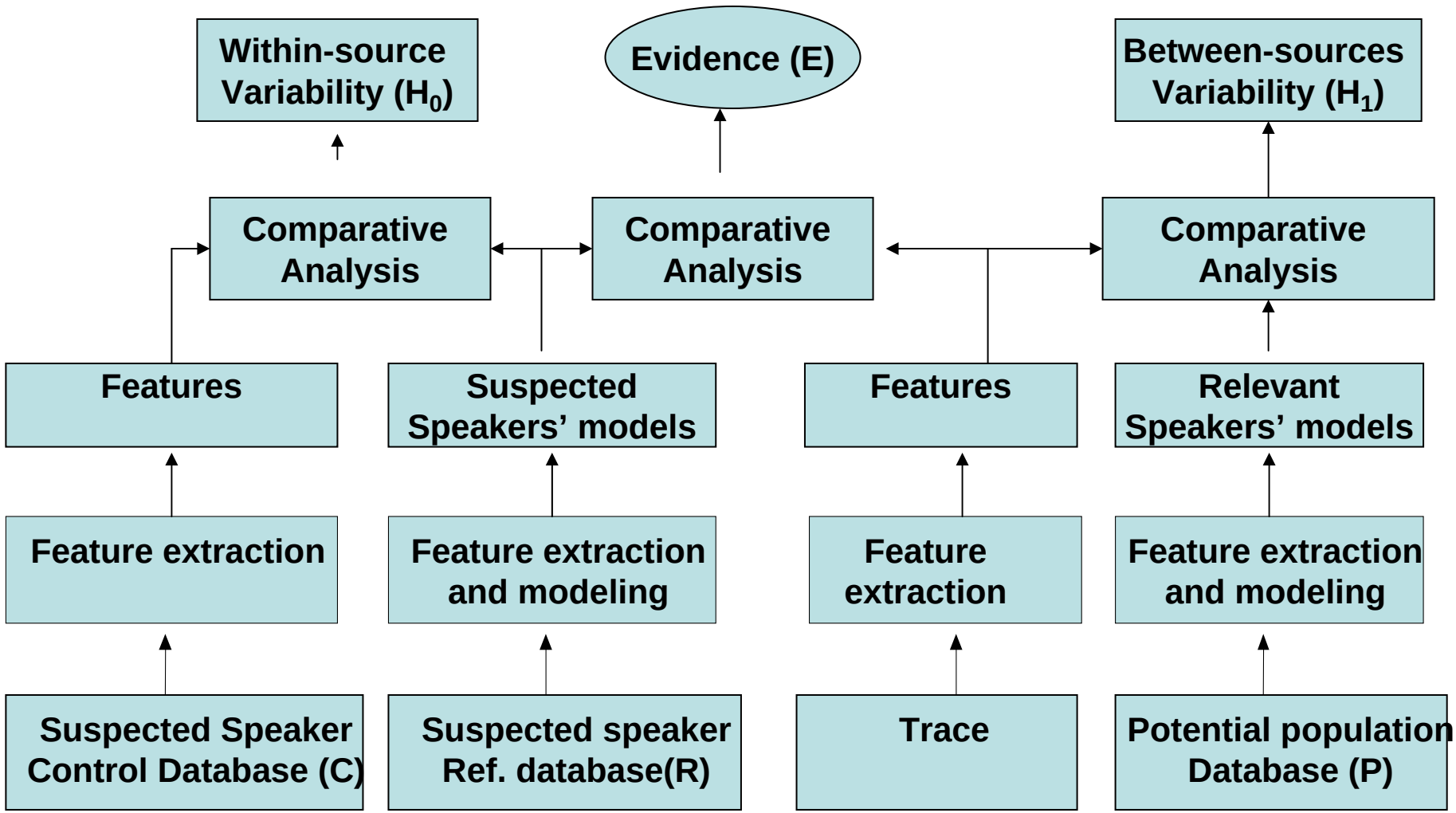
- A relevant potential population (P) database
  - Difficulty : Size, availability and cost

  P database similar in recording conditions and linguistic contents to the R database

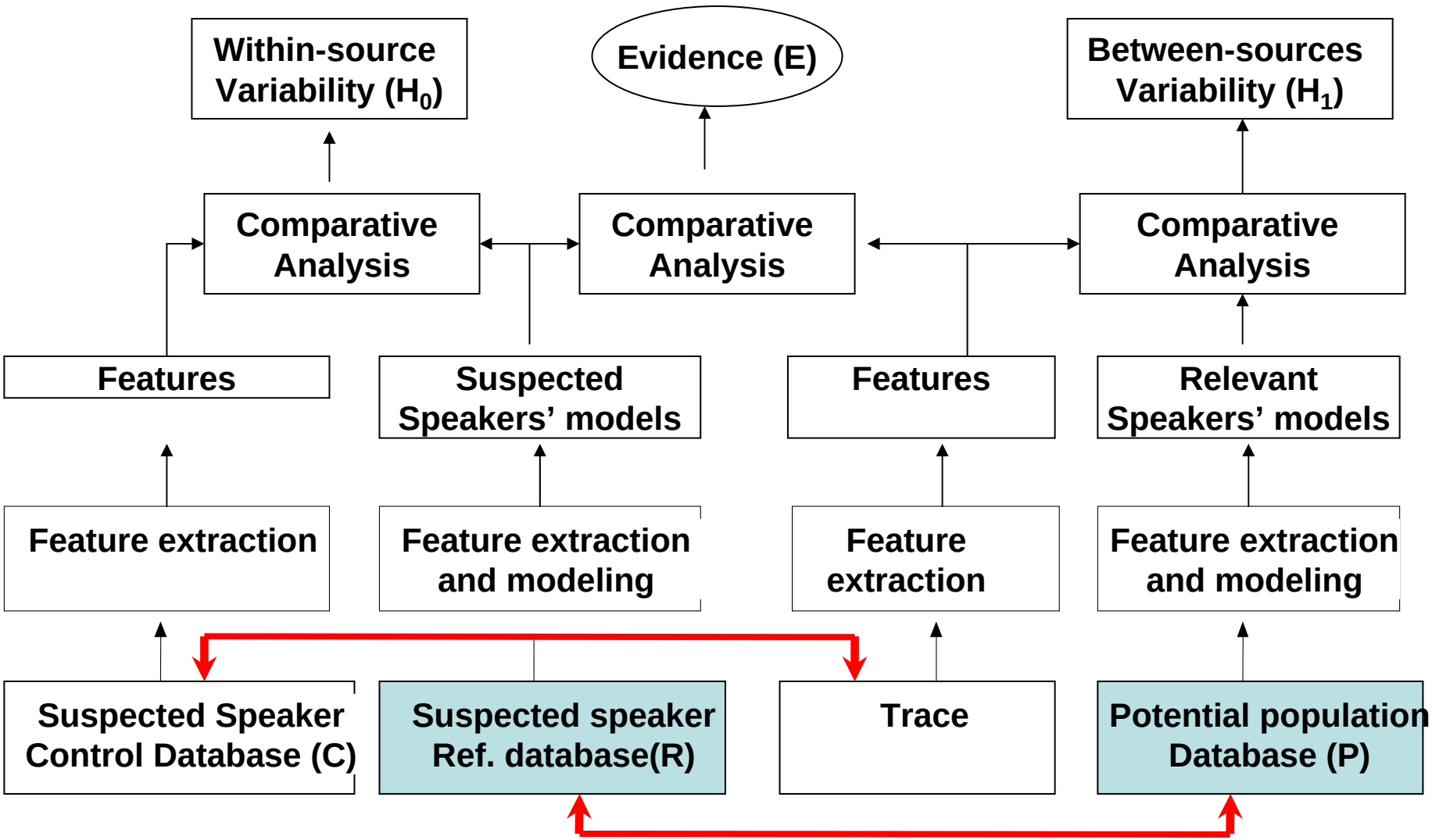
Algorithms

- Appropriate pre-processing, feature extraction and statistical modeling techniques best suited to the case

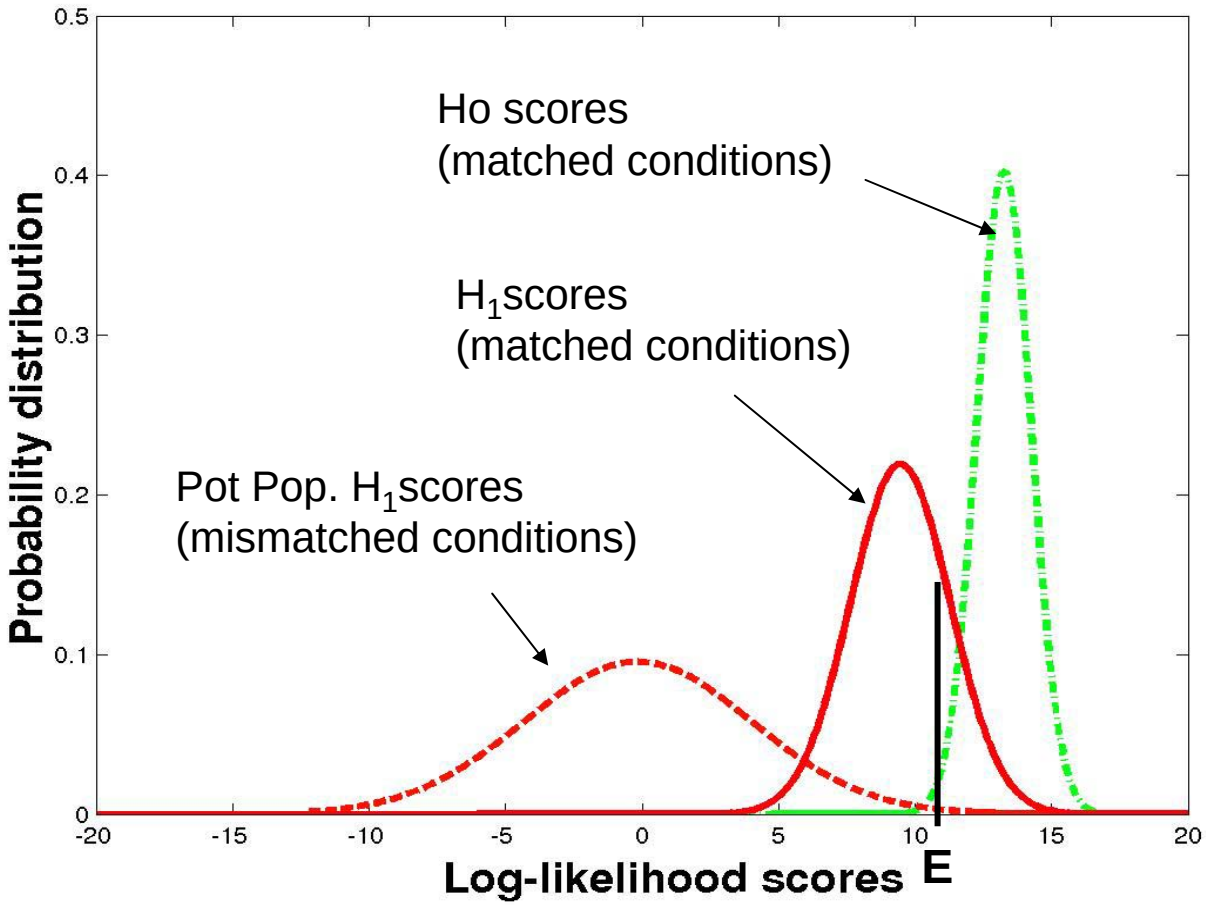
Bayesian Interpretation in Forensic Automatic Speaker Recognition



Bayesian Interpretation in Forensic Automatic Speaker Recognition

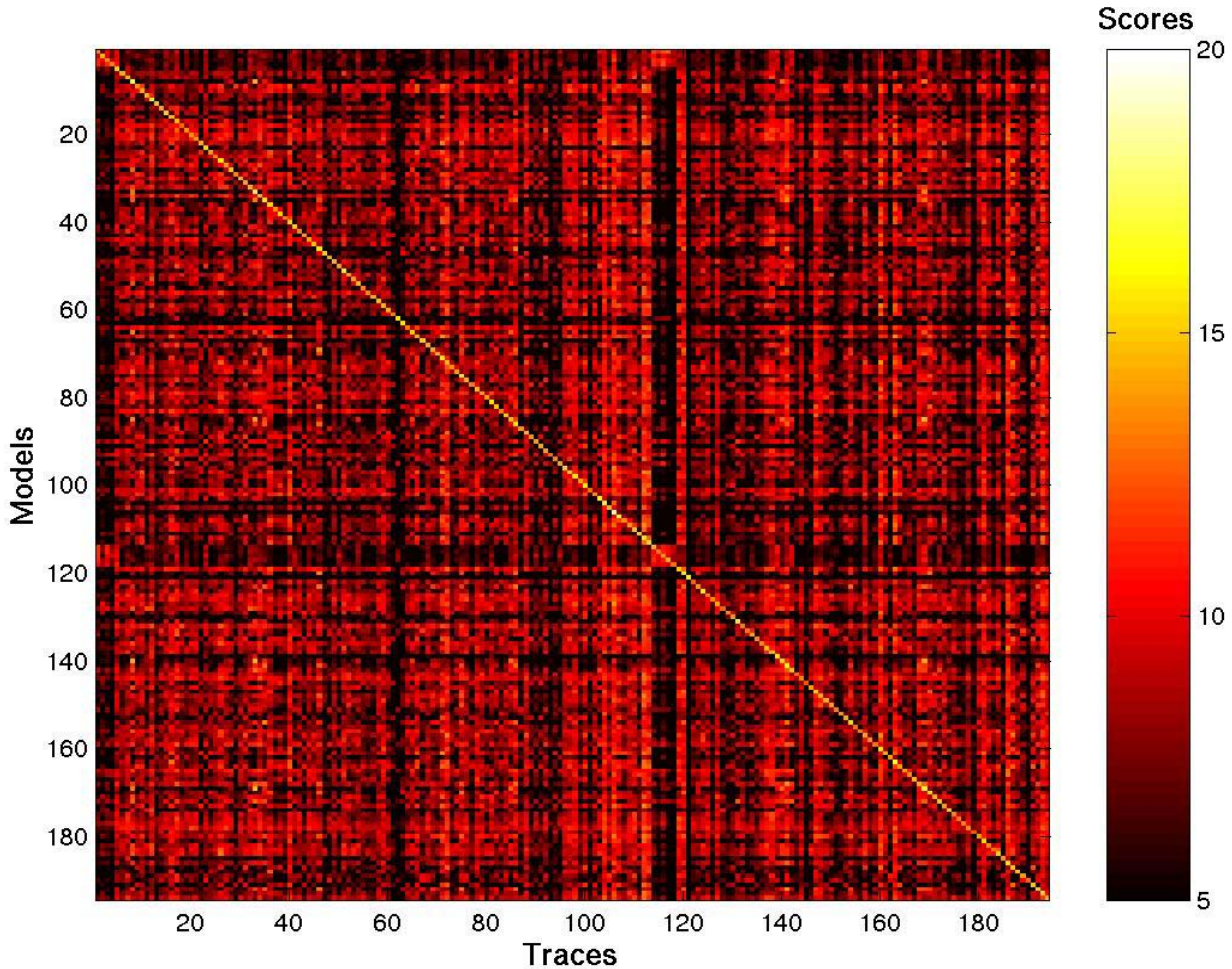


Database Mismatch and its influence LR



Mismatch results in flawed estimation of LR

Discrimination in Matched Conditions



Potential Population Database (Swisscom Polyphone + GMM Based Classifier)

Does the algorithm work well with the database ?

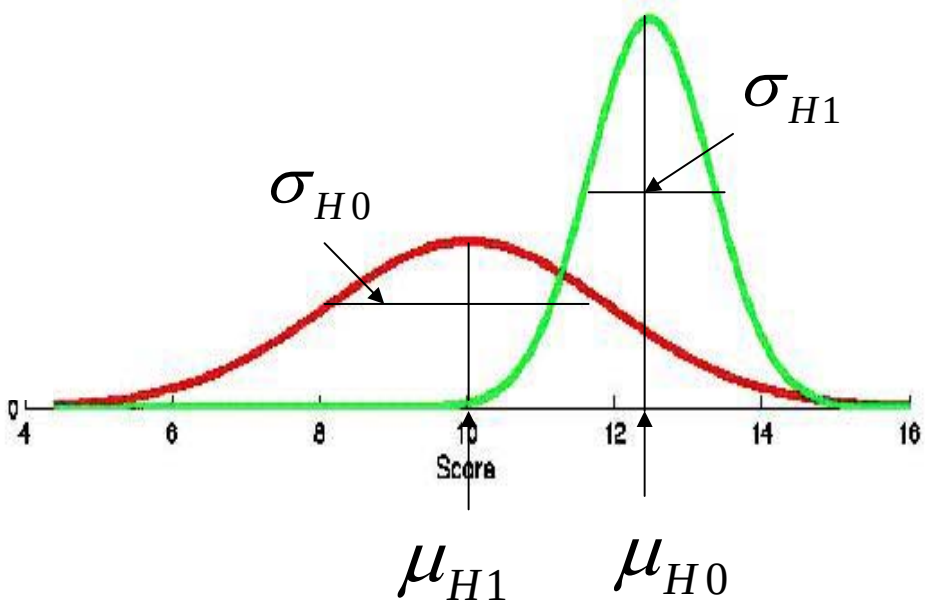
Discrimination in Matched Conditions

Does the algorithm work well with the database ?

- Discrimination coefficient

$$DC = \frac{\mu_{H_0} - \mu_{H_1}}{\sigma_{H_0} + \sigma_{H_1}}$$

Measures the normalized separation between the H0 and H1 distributions

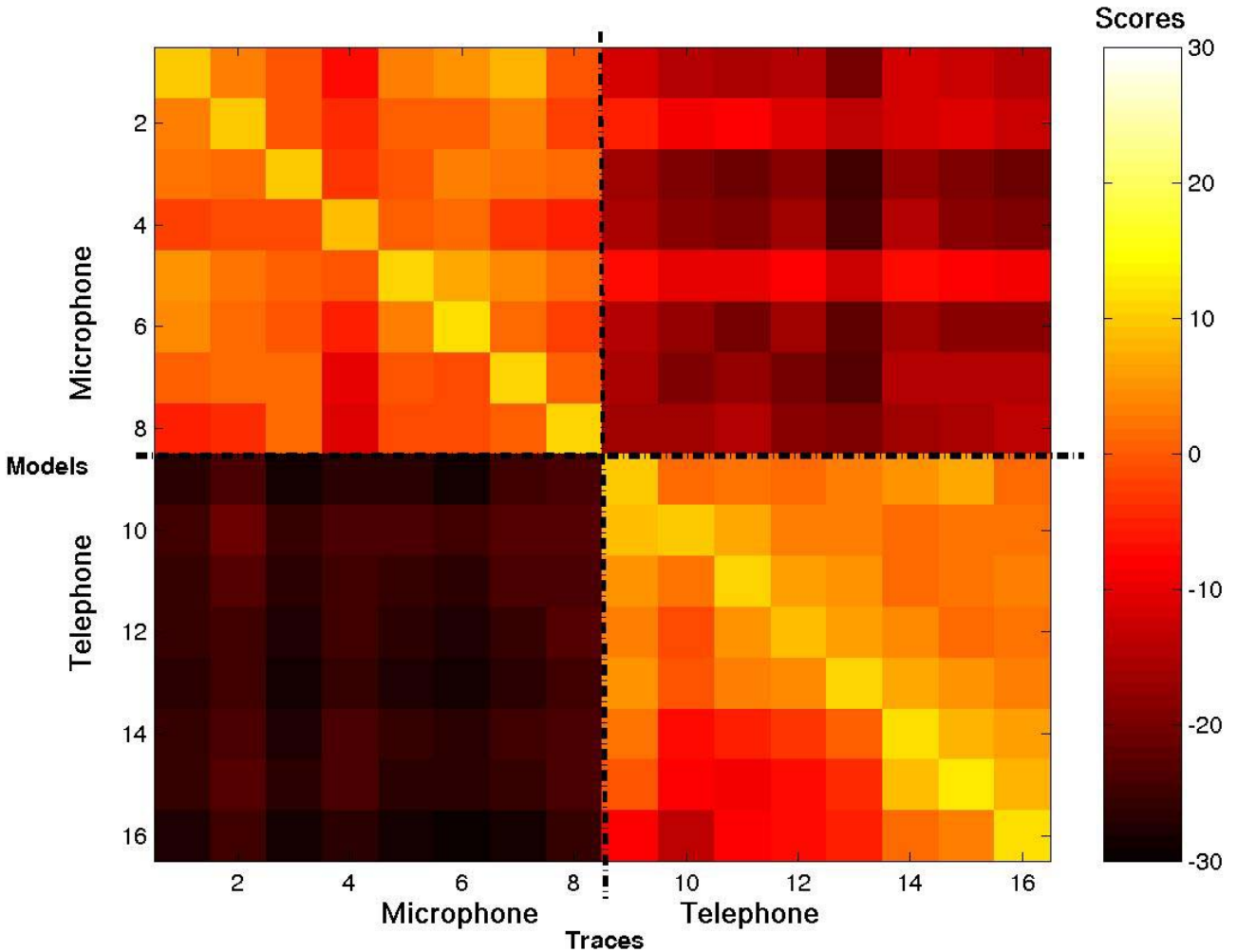


Detection of Mismatch

Are the databases used in mismatched conditions ?

- **A set of speakers from each database is selected**
- **Corresponding traces for each of these speakers is selected**
- **Each of the sets of speakers and traces are compared**
  - **within set**
  - **across sets**
- **Scores are obtained for**
  - **H_1 and H_0 within database and across databases**

Detection of Mismatch

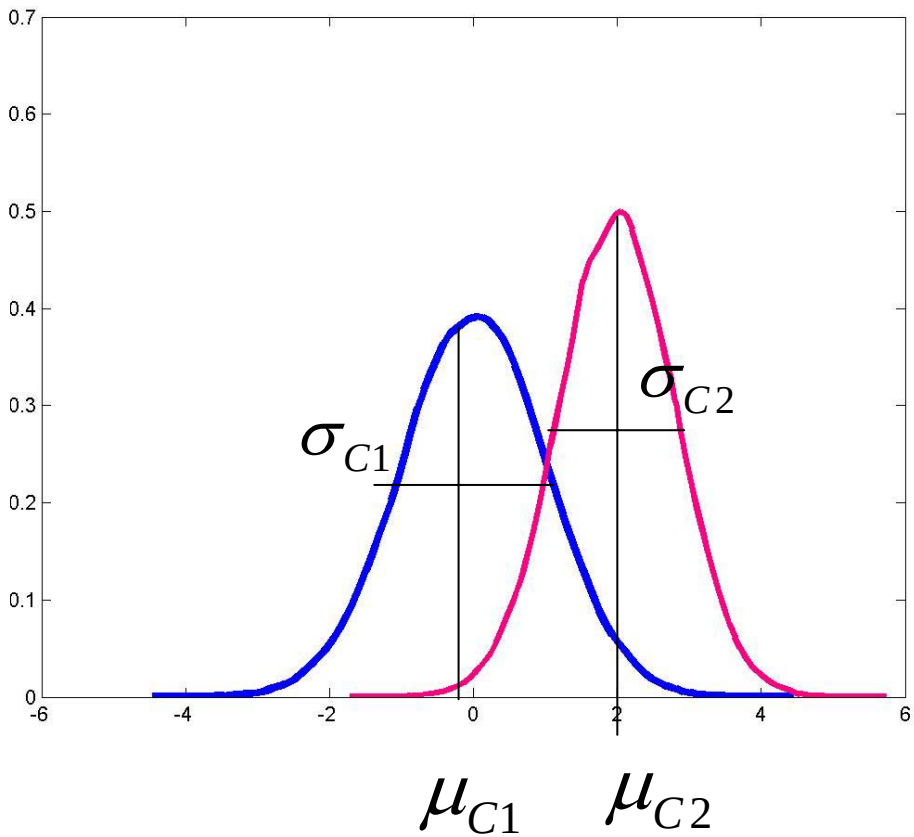


FBI NIST 2002 Database : 2 conditions (Microphone-Telephone)

Detection of Mismatch

- Statistical significance of difference of score distributions within and across databases measured using the z-test

$$Z = \frac{\mu_{C1} - \mu_{C1}}{\sqrt{\frac{\sigma_{C1}^{\ 2}}{n_{C1}} + \frac{\sigma_{C2}^{\ 2}}{n_{C2}}}}$$



Experimental Framework

Databases

- Reference (R), Control (C), and Traces (T) database
  - IPSC Polyphone 02
    - 12 speakers (Fixed phone, Cellular & Analogue tape recordings- French and German)
- Potential Population Database
  - Swisscom Polyphone (Swiss French)
- M-Database (Mismatch Evaluation database)
    - **Subset of the NIST 2002 Speaker recognition database**

Automatic Recognition system

- Feature extraction: RASTA -PLP
- Statistical modeling: Gaussian Mixture Modeling (GMM)
- Likelihood ratio (Kernel Density Estimation)

Compensation for Mismatch

- If significant mismatch is detected
  - Select another compatible potential population database
  - Record the suspect in conditions compatible with the databases available
- If neither of the above options are possible,
  - Decide not to analyze the case using B.I framework or
  - Apply statistical compensation of the mismatched conditions

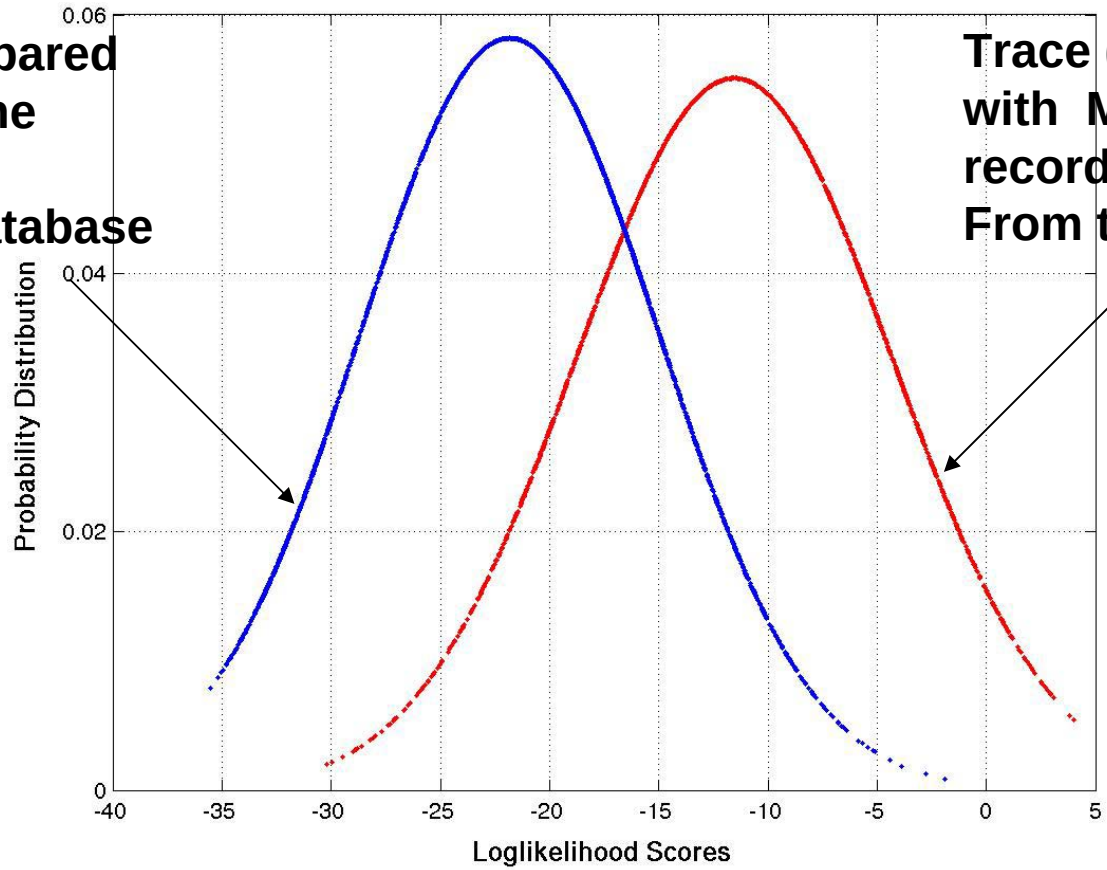
Compensation for Mismatch

- Statistical compensation of mismatch
  - Use of the M database
  - same speakers in two different conditions
    - Condition of the P database.
    - Condition of the R recording
  - Compare the questioned recording with the mismatch database
  - Estimate the speaker independent ‘shift’ of scores due to the mismatched conditions

Compensation for Mismatch

Trace (T) compared with Telephone recordings From the M database

Trace (T) compared with Microphone recordings From the M database



FBI NIST 2002 Database : 2 conditions (Microphone-Telephone)

Compensating for Mismatch

We propose to use the following normalization:

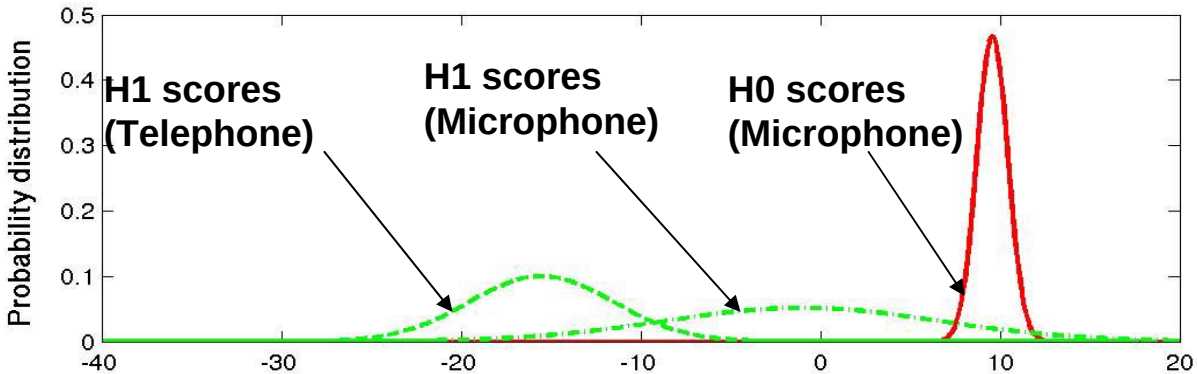
$$f(X) = \left(X + \mu_{H_1C_2} - \mu_{H_1C_1}\right) \cdot \frac{\sigma_{H_1C_2}}{\sigma_{H_1C_1}}$$

C1 and C2 correspond to 2 conditions of the M database

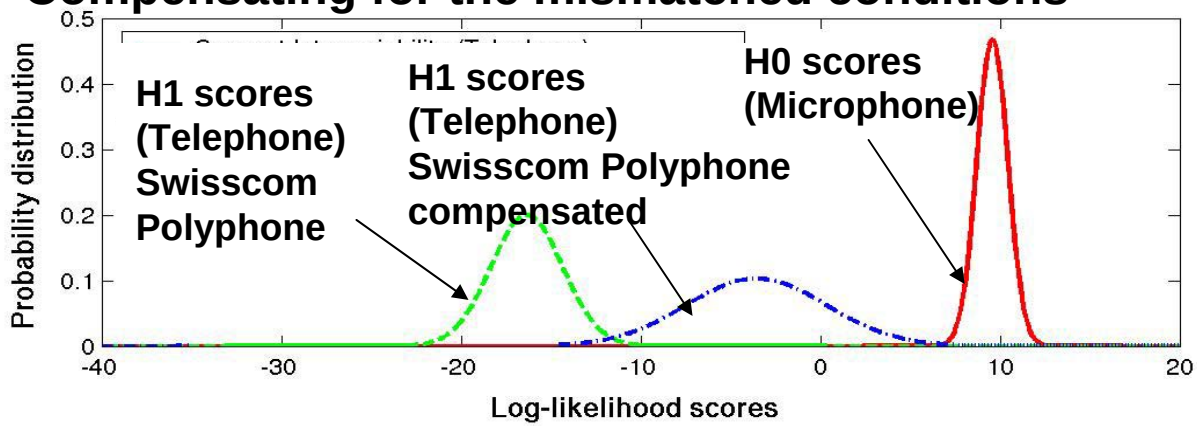
This normalization shifts the mean of the mismatched potential population scores to the conditions similar to that of the suspect recording

Compensating for Mismatch

Estimating the bias due to the two mismatched conditions

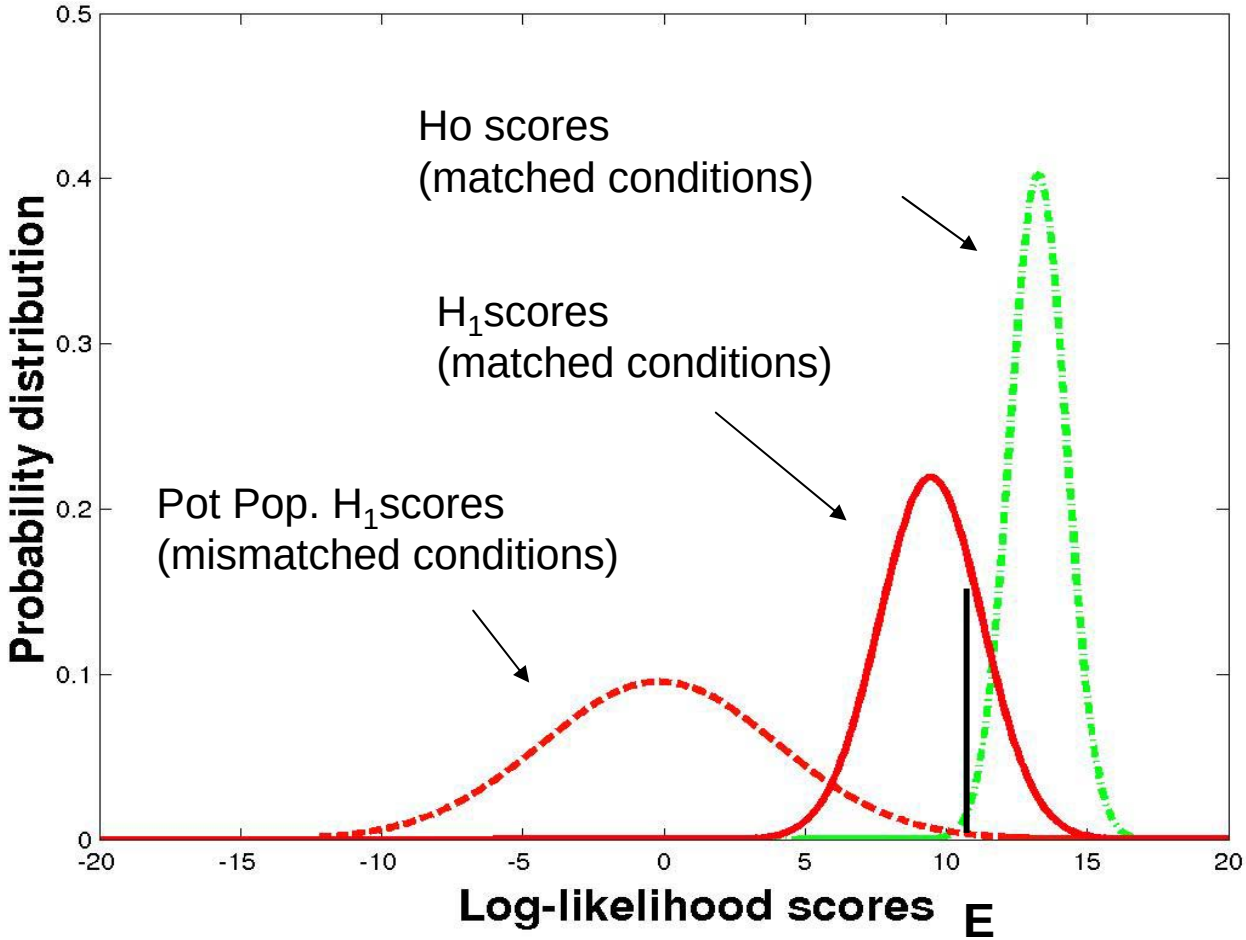


Compensating for the mismatched conditions



The LR estimated using statistical compensation is more representative of the strength of evidence if the databases were compatible.

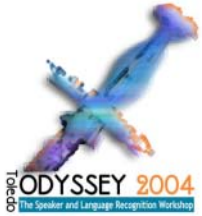
Compensating for Mismatch



Not compensating for mismatch can be the difference Between an LR < 1 and an LR > 1

Conclusion

- Mismatched recording conditions can result in the under over, or even wrong estimation of the strength of evidence
- Care has to be taken to use databases in conditions compatible with the case.
- An evaluation of the mismatch within and across databases must be performed for every case.
- Statistical compensation of scores can be performed in order to minimize the effects of mismatch if it is not possible to choose compatible databases



Questions?

Thank you for your attention