A methodology for the creation of a forensic speaker recognition database to handle mismatched conditions

Anil Alexander and Andrzej Drygajlo

Swiss Federal Institute of Technology Lausanne
Signal Processing Institute

Thursday 4th August 2005
Marrakesh, Morocco

Outline

- Evaluation of the strength of evidence in forensic automatic speaker recognition using a corpus-based Bayesian Interpretation (B.I.) method
- Mismatched recording conditions of the databases and why they influence the estimation of the strength of evidence
- Methodology for the creation of a database to handle the mismatch in recording conditions
- Estimation and compensation of the mismatch using this database
- Conclusion

Bayesian Interpretation in Forensic Automatic Speaker Recognition

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

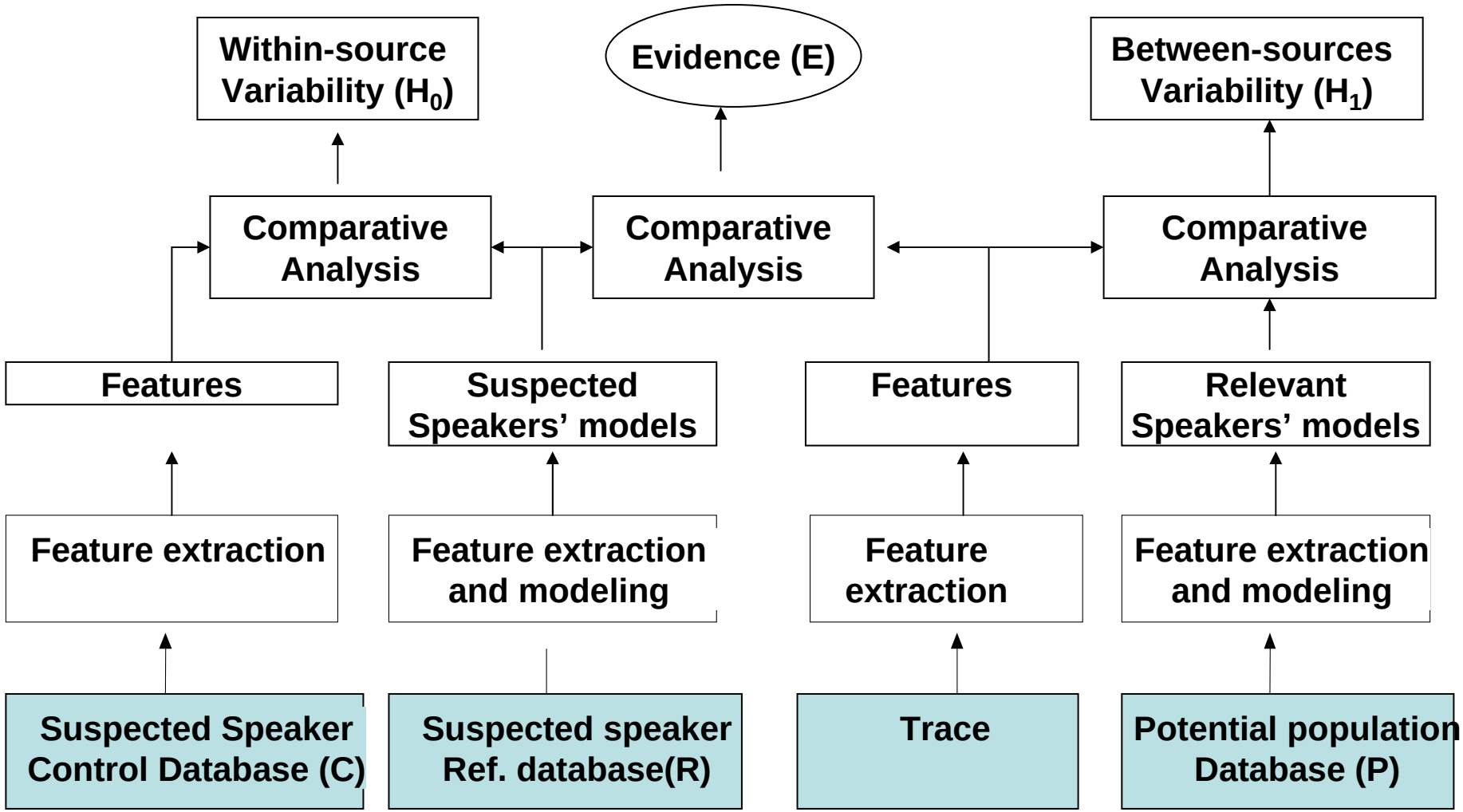
H_0 – The suspected speaker is the source of the trace

H_1 - Another speaker is the source of the trace

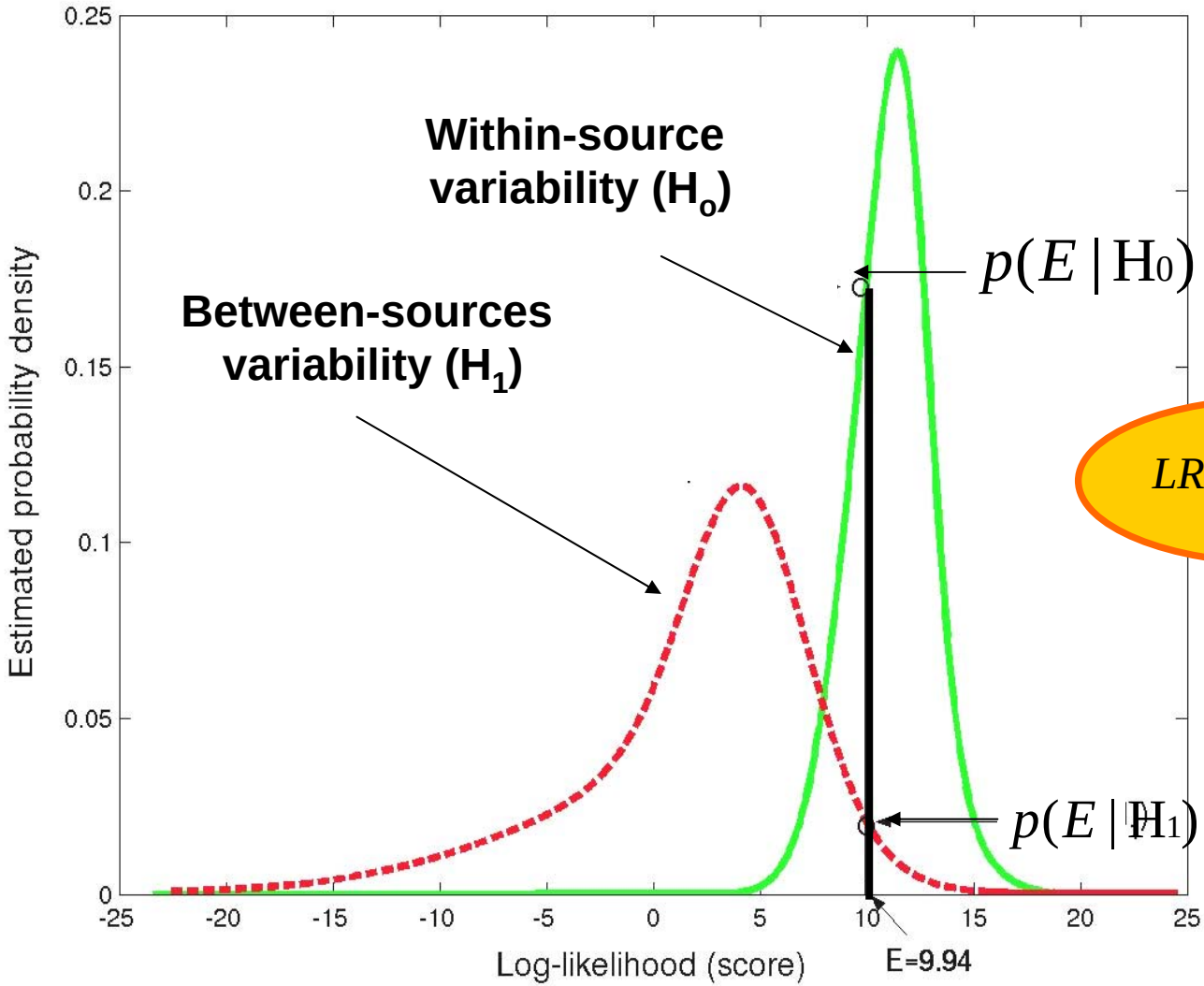
$$LR = \frac{p(E \mid H_0)}{p(E \mid H_1)}$$

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

Corpus-based Bayesian Interpretation in FASR



Likelihood Ratio



Databases required in the B.I. methodology

Databases Required

- Suspect reference database (R)
- Suspect control database (C)
- A relevant potential population (P) database
- For performance evaluation purposes, a trace (T) database is used

– Assumptions:
  – P database similar in recording conditions and linguistic contents to the R database
  – C database similar in recording conditions and linguistic contents to the questioned recording

– Difficulty : Size, availability, and mismatch of databases

Often not possible to record the suspect in conditions of the case or to obtain a representative potential population in matched conditions.

Requirements for forensic speaker database

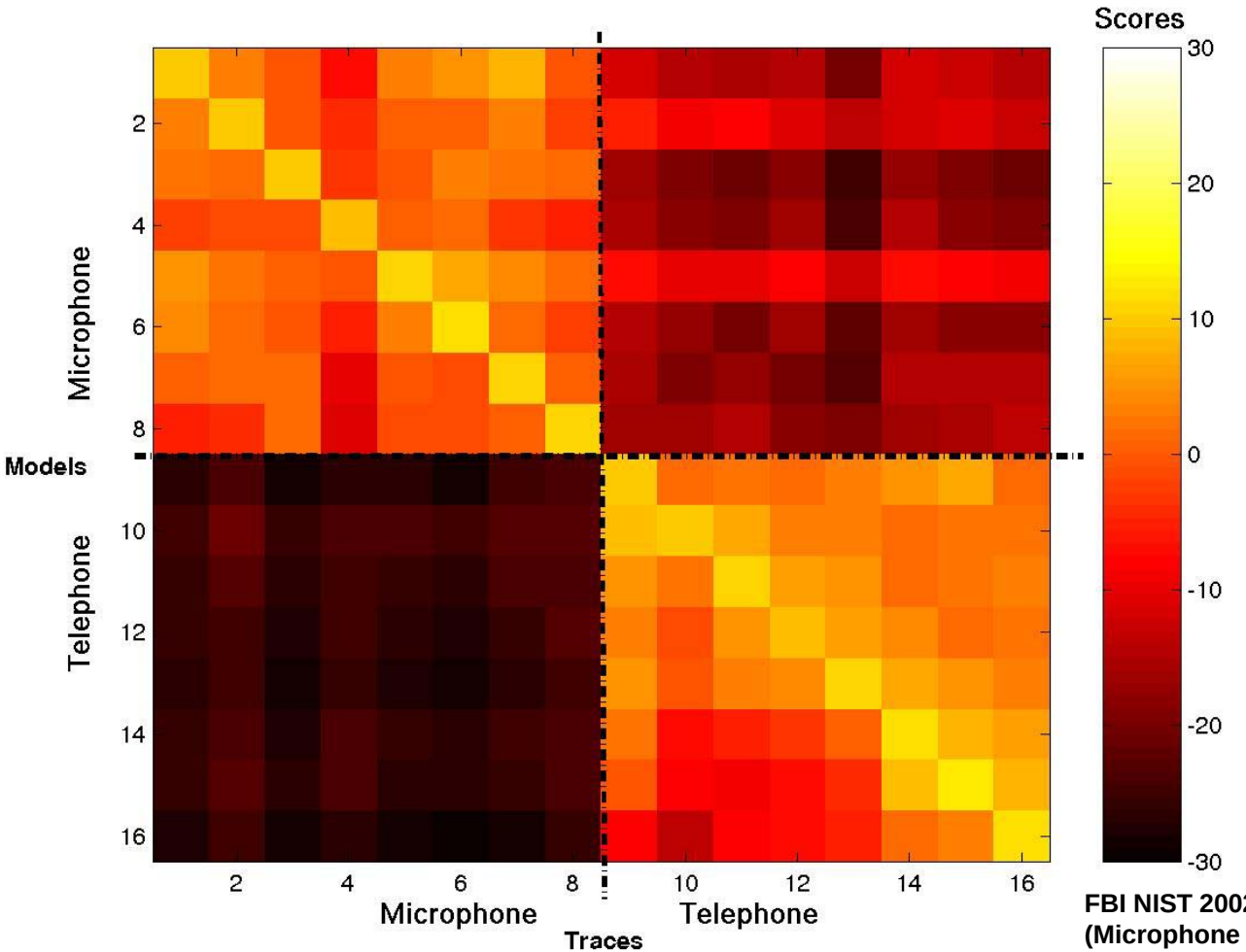
Some practical considerations:

- Attempt to simulate real world conditions closely
- Contain similar-sounding subjects as far as possible. i.e., same language and accent, preferably the same sex.
- Technical conditions (recording conditions, transmission channels)
- Statistically significant number of speakers
- Sufficient duration of recordings to create statistical models of speakers

Mismatched recording conditions of the database

- What if it is not possible to obtain all the files in the same recording conditions ?
  - What if the trace and the three databases (P,R,C) are obtained in different recording conditions ?
- This kind of situation often appears in forensic speaker recognition casework, as
  - The conditions in which the questioned recording (trace) and the suspected speaker's recordings (R, C) were obtained are beyond the control of the forensic expert.
  - A potential population database (P) in the recording conditions of the trace is not available.
- In this presentation the assumptions made are:
  - that the conditions of the trace and P are known
  - the conditions of R and C can be chosen correspondingly

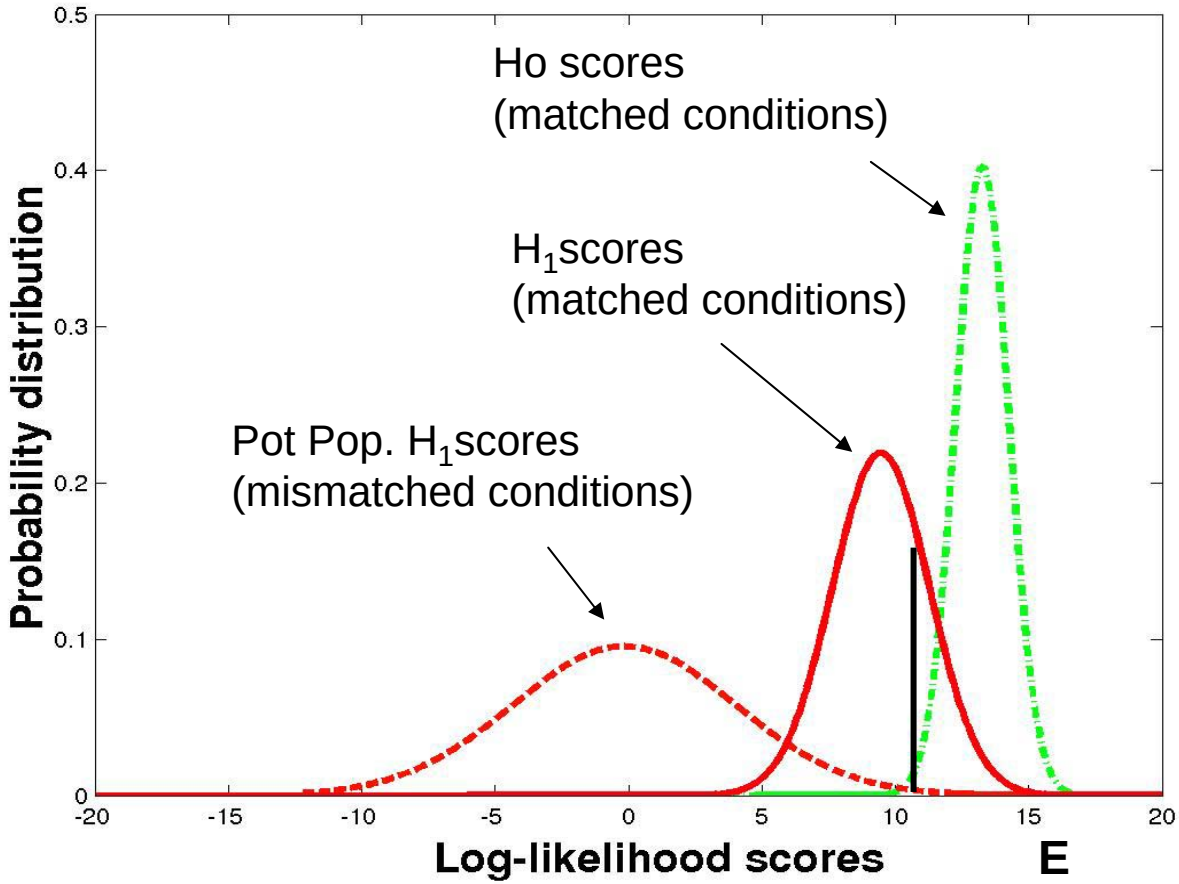
Using databases with mismatched recording conditions



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

The extent of mismatch can be measured using statistical testing [Alexander et al '04]

The influence of mismatched potential population database



P,R,C,T in conditions C1,C2,C2,C2

The influence of mismatch can be the difference between an LR < 1 and an LR > 1

Mismatched databases

Corpus-based B.I. and mismatch in different situations

Sit. No.	P	C	R	T
1.	C_2	C_2	C_2	C_2
2.	C_2	C_2	C_2	C_1
3.	C_2	C_2	C_1	C_2
4.	C_2	C_1	C_2	C_2
5.	C_1	C_2	C_2	C_2
6.	C_2	C_1	C_2	C_1
7.	C_2	C_1	C_1	C_2
8.	C_1	C_1	C_1	C_1

Where C_1 represents recording condition 1 (e.g. GSM)
C_2 represents recording condition 2 (e.g. PSTN)

Handling mismatch

- In order to estimate and compensate for mismatched conditions, we require:
  - Several databases containing the same speakers in the different recording conditions.

    Recording several such databases for each case is very expensive.
  - A methodology to estimate and compensate for the ‘shift’ in scores resulting from a mismatched recording condition.

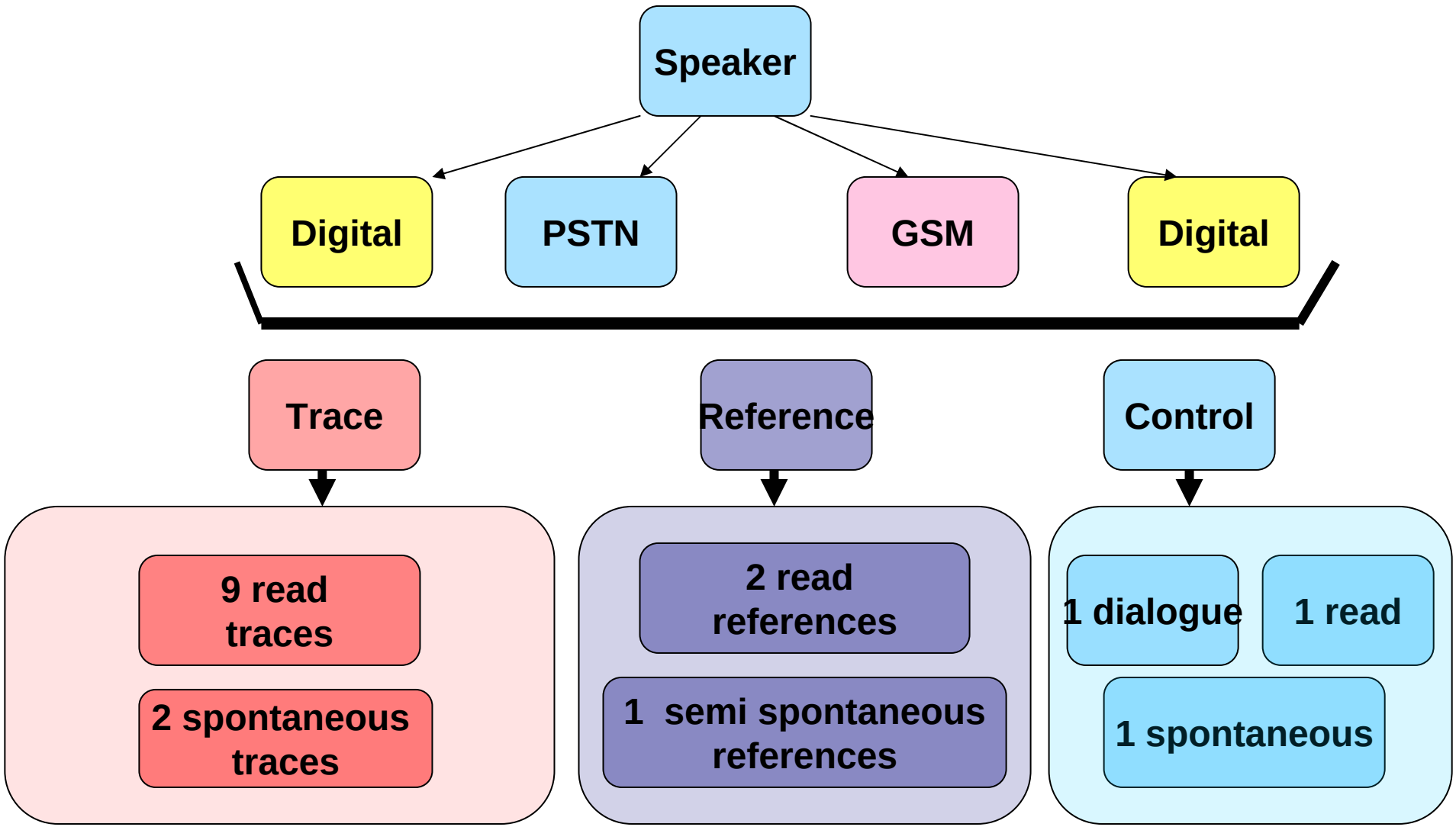
A prototype forensic speaker recognition database

- IPSC03
  - Institut de Police Scientifique, UNIL and the Signal Processing Institute, EPFL
- 70 male speakers in three different recording conditions:
  - Public switched telephone network (PSTN)
  - global system for mobile communications (GSM)
  - calling room acoustic recording using digital recorder
- controlled and uncontrolled speaking modes
- controlled conditions in a quiet room
- over 4800 audio files totalling over 40 hours

Features of the IPSC03 database

- Forensically realistic questioned recordings
- Statistically significant number of speakers
- Same telephones for all speakers to minimize the effects of individual handsets
- Prompted and free questioned recordings in the form of threatening telephone calls
- Digital recordings at the source and after having passed through a transmission channel.

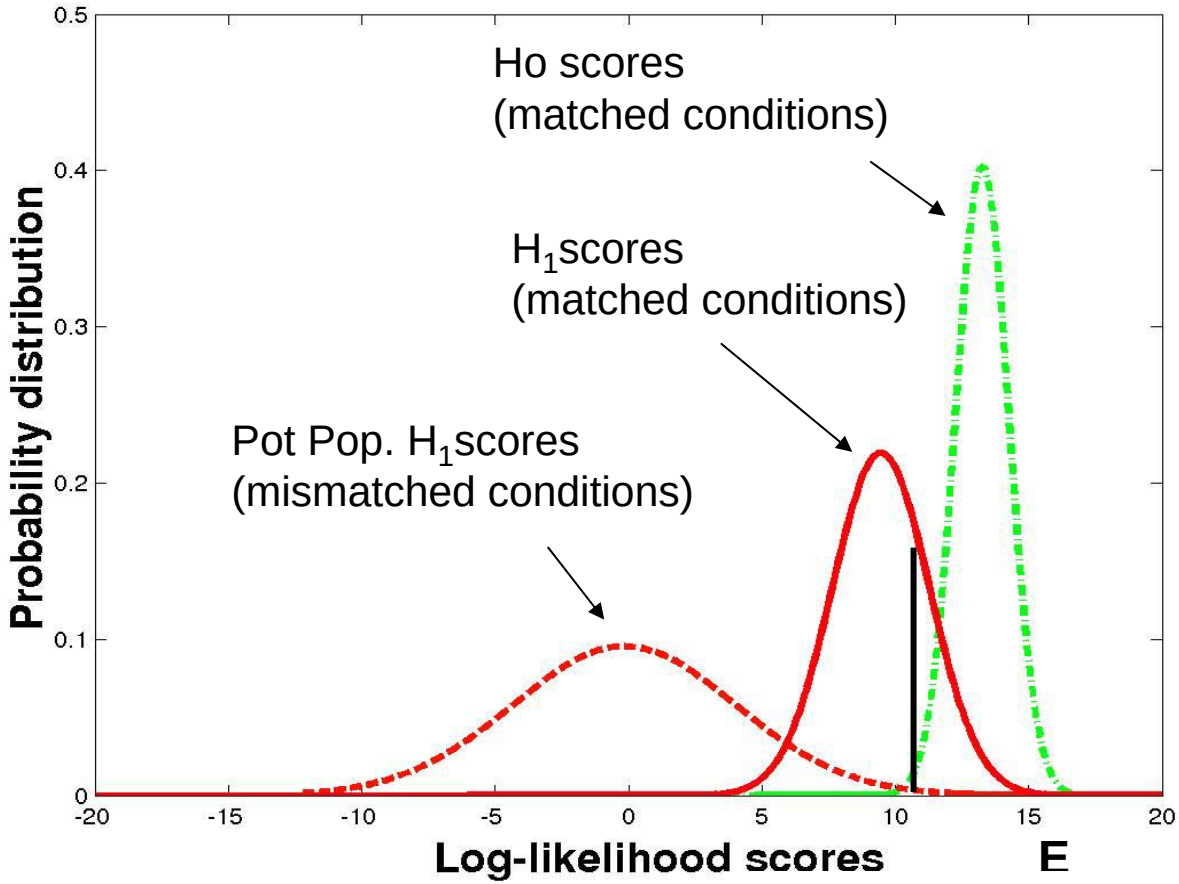
Database Layout



Method for compensation of mismatch

- Statistical compensation of mismatch
  - Use a same set of speakers in two different conditions (C1 and C2)
  - Compare the test recording with models of the set of speakers in two conditions
  - Estimate the speaker independent ‘shift’ of scores due to the mismatched conditions
    - Ideally using the whole set of the database speakers
    - Practically, using a subset of this database

The influence of mismatched potential population database



P,R,C,T in conditions C1,C2,C2,C2

Score distribution compensation using distribution scaling:

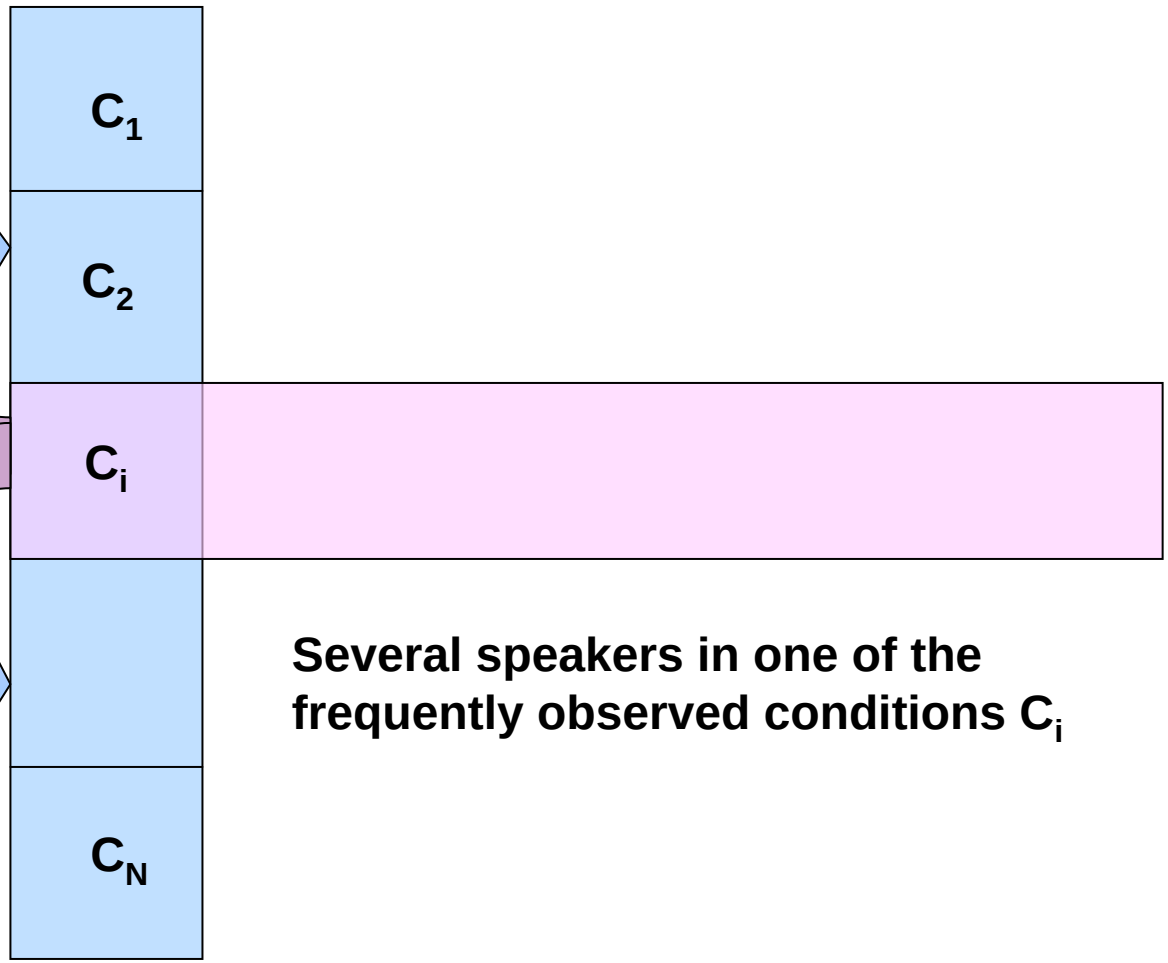
$$X_{C_2} = (X_{C_1} - \mu_{C_1})\frac{\sigma_{C_2}}{\sigma_{C_1}} + \mu_{C_2}$$

Experimental Framework

- The databases chosen are the subsets of the IPSC03 database

Simulated case: the suspect is the source of the questioned recording

Schema of a forensic database to handle mismatch



Shift due to difference in conditions

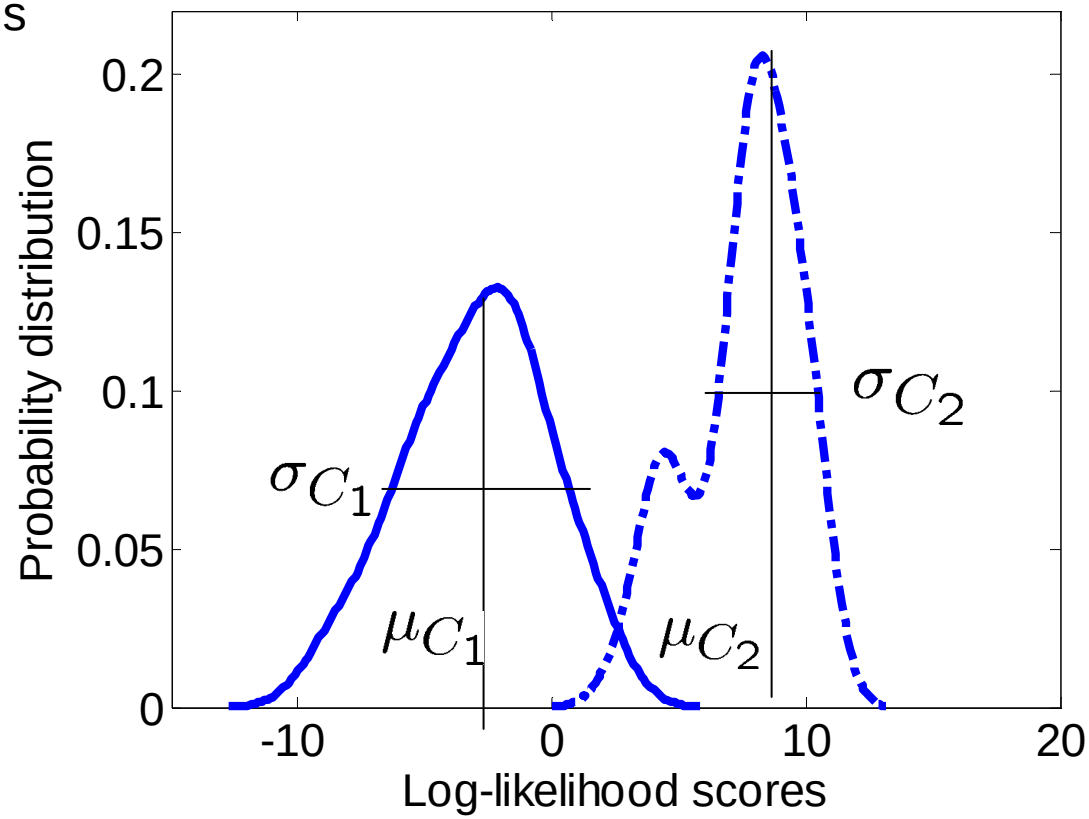
Several speakers in one of the frequently observed conditions C_i

A smaller database of speakers in several recording conditions

Compensating for Mismatch

We choose a sub-database of 22 speakers in both PSTN and GSM conditions, and use these scores to estimate the shift between conditions

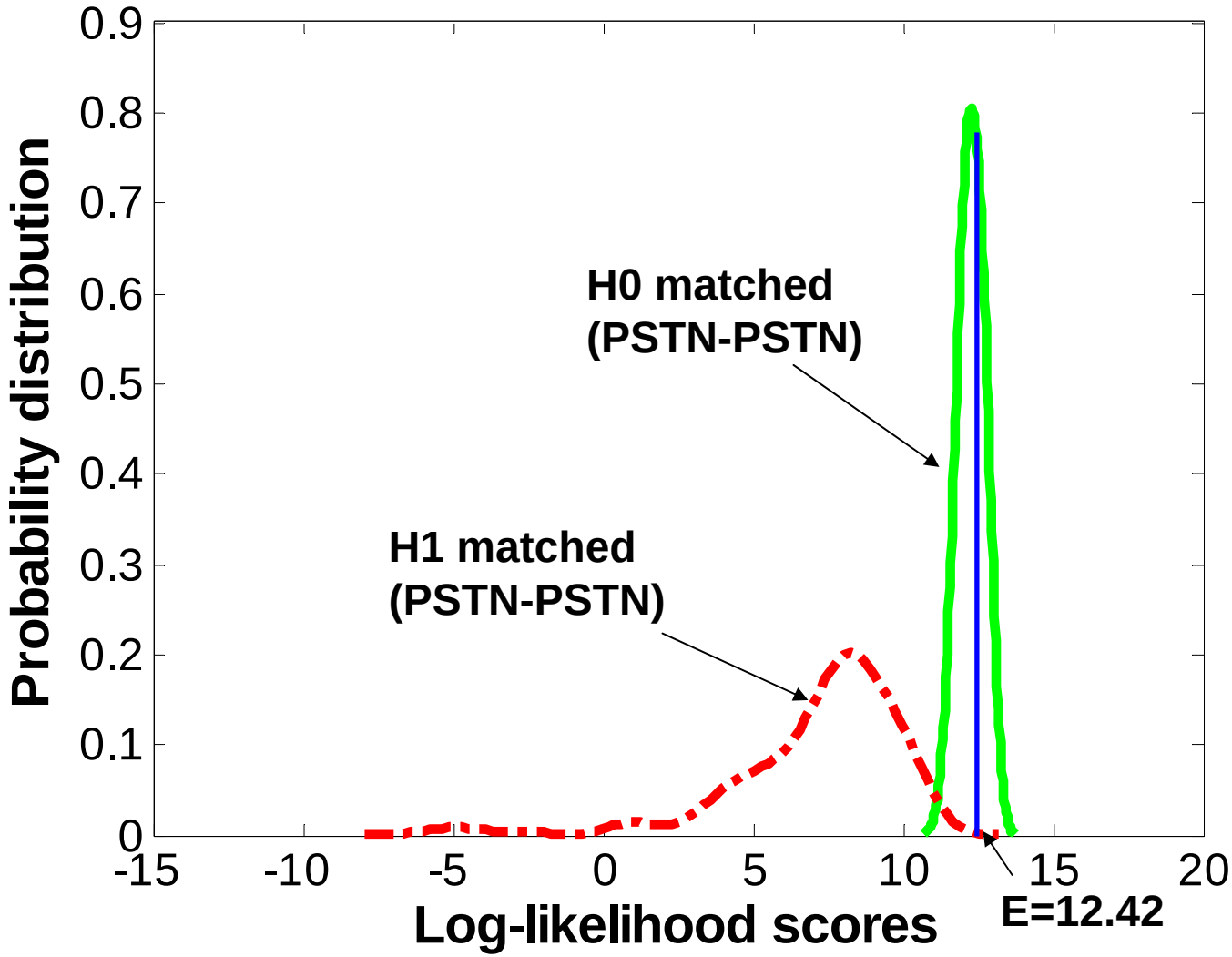
[Similar to feature mapping technique Reynolds, ICASSP03]



We propose to use the following normalization:

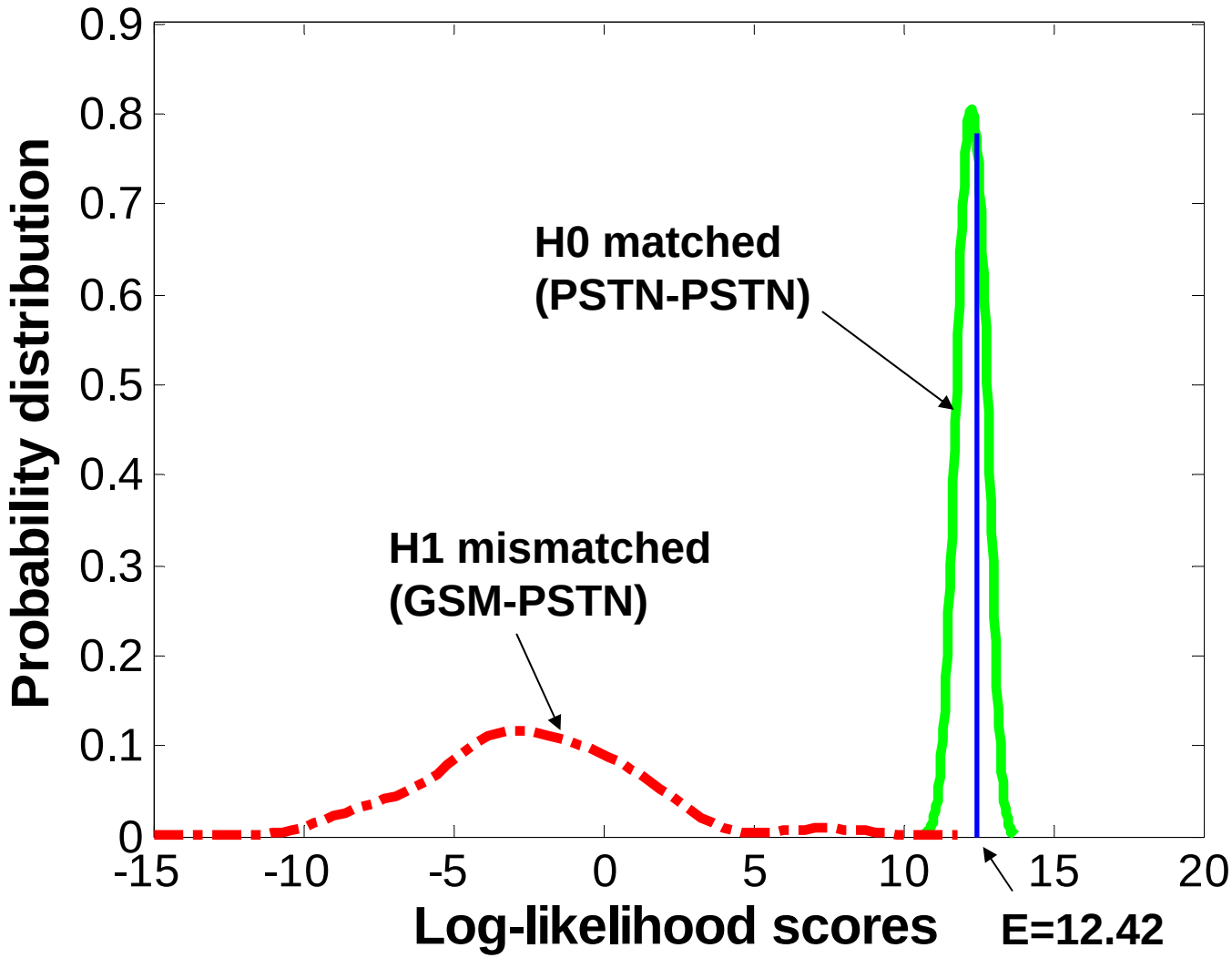
$$X_{C_2} = (X_{C_1} - \mu_{C_1})\frac{\sigma_{C_2}}{\sigma_{C_1}} + \mu_{C_2}$$

Matched conditions



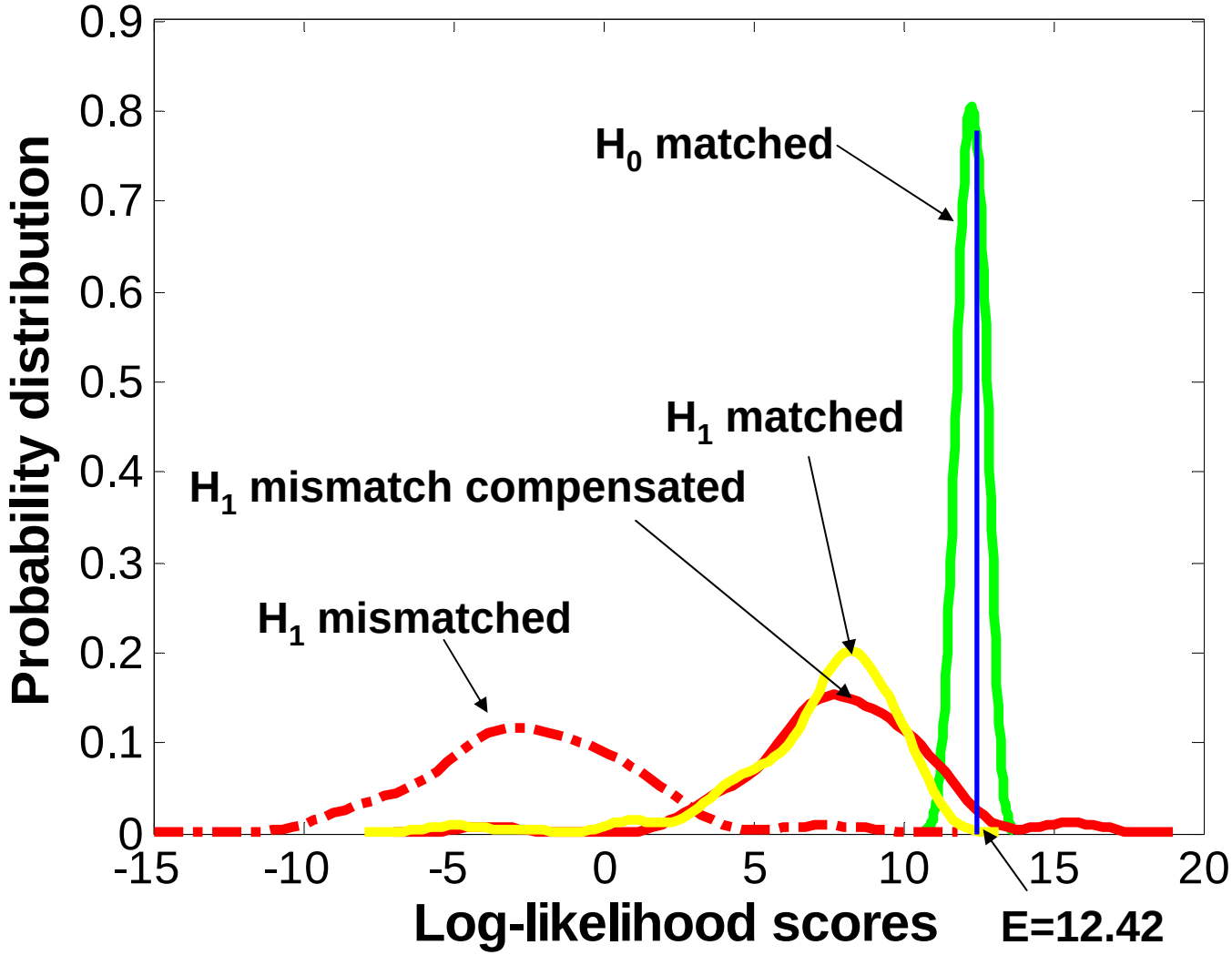
The LR estimated in matched conditions is 354

Mismatched conditions



The LR estimated in mismatched conditions is 73,678 because of the shift in potential population scores

Compensation for mismatch



The LR estimated using statistical compensation (29.34) is more representative of the strength of evidence obtained in matched conditions.

Conclusion

- The strength of evidence (likelihood ratio) depends on the recording conditions of the databases used.
- In certain cases, it may not be possible to obtain all the databases in the same conditions.
- A forensic potential population database should include several smaller databases in different conditions containing the same speakers.
- These databases can be used to estimate and reduce the effects of mismatch due to different recording conditions.

Questions?

Thank you for your attention