Value and Interpretation of Biometric Evidence in Forensic Automatic Speaker Recognition

Dr. Andrzej Drygajlo
Speech Processing and Biometrics Group
Swiss Federal Institute of Technology Lausanne (EPFL)

Forensic Speech and Audio Analysis Working Group

Outline

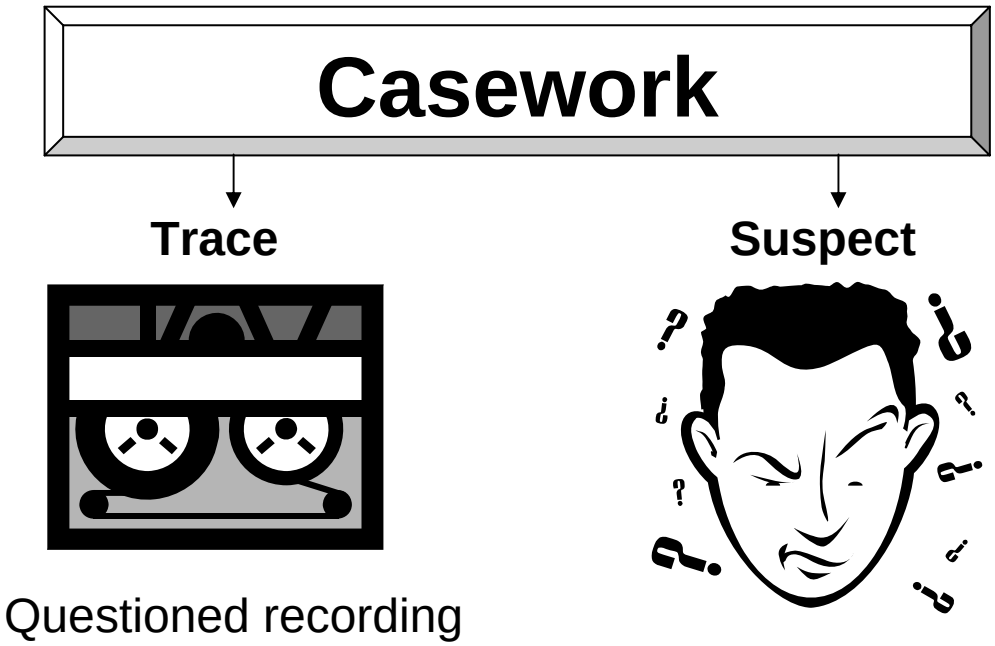
- **Forensics and Biometrics**
- Forensic Speaker Recognition (FSR)
- Bayesian Interpretation of Forensic Evidence
- Forensic Automatic Speaker Recognition (FASR)
- Automatic Speaker Recognition (ASR)
- Deterministic and Statistical Methods
- Voice as Biometric Evidence
- FASR - Univariate (Scoring) and Multivariate (direct) Methods
- Conclusions

Forensics

- **Forensic science (Forensics)** refers to the applications of scientific principles and technical methods to the investigation of criminal activities, in order to demonstrate the existence of a crime, and to determine the identity of its author(s) and their modus operandi.
  - **Forensic** (adj.) means the use of science or technology in the investigation and establishment of facts or evidence in the court of law.

- **Biometrics** is the science of establishing identity of individuals based on their biological and behavioral characteristics

Forensic Speaker Recognition



Forensic speaker recognition (FSR) is the process of determining if a specific individual (suspected speaker) is the source of a questioned voice recording (trace).

Forensic Speaker Recognition

- **Aural-perceptual methods**
  - earwitnesses, line-ups
- Visual methods and « voiceprint? »
  - visual comparison of spectrograms of linguistically identical utterances (utterly misleading!)
- **Aural-instrumental methods**
  - analytical acoustic approach combined with an auditory phonetic analysis
- **Automatic methods**
  - Speaker verification – not adequate
  - Speaker identification – not adequate
  - Bayesian framework for the evaluation of voice as biometric evidence

Despite recent advances in Bayesian Statistics, it is critical not to loose sight of the fact that these methods are merely tools.

Automatic Speaker Recognition

- **Speaker recognition** is the general pattern recognition term used to include all of the many different tasks of discriminating people based on the sound of their voices.

- **Speaker identification** is the task of deciding, given a sample of speech, who among many candidate speakers said it. This is an *N*-class decision task, where *N* is the number of candidate speakers.

- **Speaker verification** is the task of deciding, given a sample of speech, whether a specified candidate speaker said it. This is a 2-class decision task and is sometimes referred to as a speaker detection task.

Forensic Automatic Speaker Recognition

- Forensic automatic speaker recognition – data-driven methodology for quantitative interpretation of recorded speech as evidence
- The interpretation of recorded voice as evidence in the forensic context presents particular challenges, including within-speaker (within-source) variability, between-speakers (between-sources) variability, and differences in recording sessions conditions
- Consequently, FASR methods should provide a probabilistic evaluation which gives the court an indication of the strength of the evidence given the estimated within-source, between-sources and between-session variabilities, and this evaluation should be compatible with other interpretations in other forensic disciplines
- The Bayesian interpretation framework, using a likelihood ratio concept, offers such interoperability

Bayesian probability statements are about states of mind over states of the world, and not about states of the world *per se*).

Forensic specificity

- Short utterances
- Questioned recording - uncontrolled environment
- Investigations in controlled conditions (longer utterances)
- Telephone quality (95%)
- Clear understanding of the inferential process
- Respective duties of the actors involved in the judicial process: jurists, forensic experts, judges, etc.

The forensic expert's role is to testify to the worth of the evidence by using, if possible a quantitative measure of this worth.

It is up to the judge and/or the jury to use this information as an aid to their deliberations and decision.

Inference and Reasoning

- The **role of forensic science** is the provision of information (factual or opinion) to help answer questions of importance to investigators and to courts of law.

- In developing an opinion, the forensic expert has to utilise some form of inference process (from observations to the source).

- **Reasoning**
  - Deductive reasoning occurs in those situations where a logical rule can be applied to a particular set of observations
  - Induction is the process of reasoning from a set of observations within a framework of incomplete knowledge.

- Hypothetical-deductive method combined with statistical inference and inductive reasoning for forensic automatic speaker recognition – **Bayesian interpretation of evidence**

Evaluative forensic science opinion

- **Evaluative opinion** – an opinion of evidential weight, based upon case specific propositions and clear conditioning information (framework of circumstances) that is provided for use as evidence in court.

- An evaluative opinion is an opinion based upon the estimation of a **likelihood ratio**.
  - UK Association of Forensic Science Providers, "Standards for the formulation of evaluative forensic science expert opinion", Science and Justice 49 (2009), 161-164.

Adversary System

The suspected speaker is the source of the questioned recording

The speaker at the origin of the questioned recording is not the suspected speaker

Expert opinion testimony has to be **carefully documented**, and **expressed with precision**, in as neutral and objective a way as the **adversary system** permits.

Bayesian Interpretation of Forensic Evidence

Principle

- The Bayesian model, proposed for forensic speaker recognition by Lewis in 1984, allows for revision based on new information of a measure of uncertainty (likelihood ratio of the evidence (province of the forensic expert)) which is applied to the pair of competing hypotheses.

- The Bayesian model shows how new data (questioned recording) can be combined with prior background knowledge (prior odds (province of the court)) to give posterior odds (province of the court) for judicial outcomes or issues.

prior odds x **?** = **posterior odds**

Bayes' Theorem tells us how we should rationally update subjective, probabilistic beliefs in light of evidence.

Bayesian Interpretation of Forensic Evidence

The odds form of Bayes' theorem

prior background knowledge

New Data

posterior knowledge on the issue

$$\frac{P(H_0)}{P(H_1)} \times \frac{P(E \mid H_0)}{P(E \mid H_1)} = \frac{P(H_0 \mid E)}{P(H_1 \mid E)}$$

Prior odds

Likelihood Ratio (LR)

Posterior odds

province of the court

province of the forensic expert

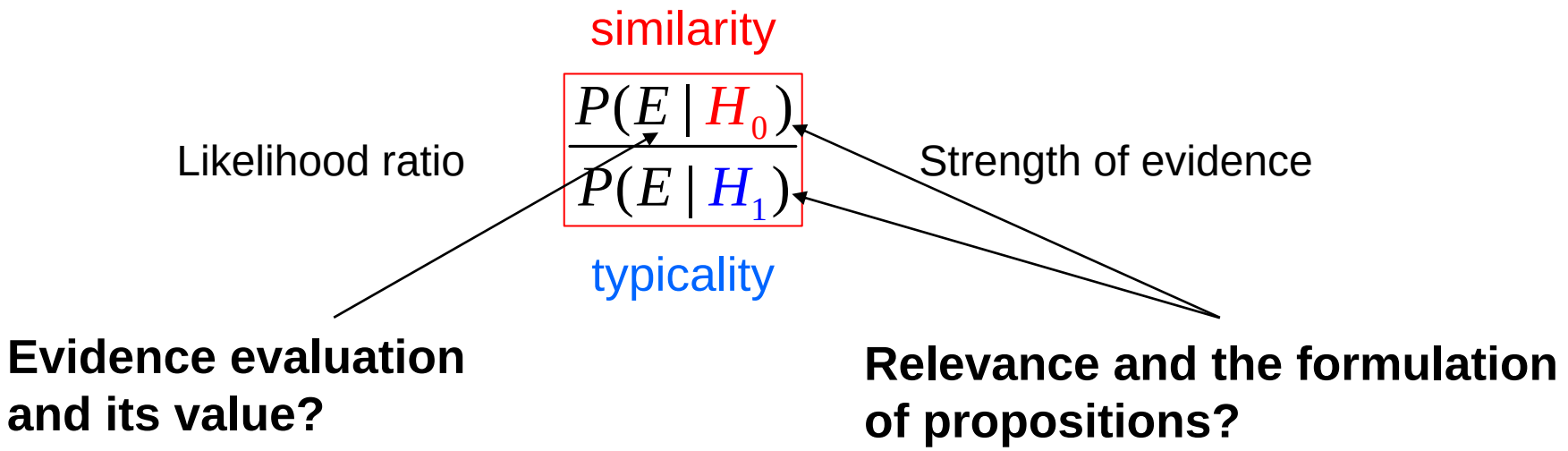
province of the court

Subjective probabilities are whatever a particular person believes, provided they satisfy the axioms of probability.

Bayesian Interpretation of Forensic Evidence

- **H_0** – the suspected speaker is the source of the questioned recording
- **H_1** – the speaker at the origin of the questioned recording is not the suspected speaker

$$\frac{P(H_0)}{P(H_1)} \times \frac{P(E \mid H_0)}{P(E \mid H_1)} = \frac{P(H_0 \mid E)}{P(H_1 \mid E)}$$

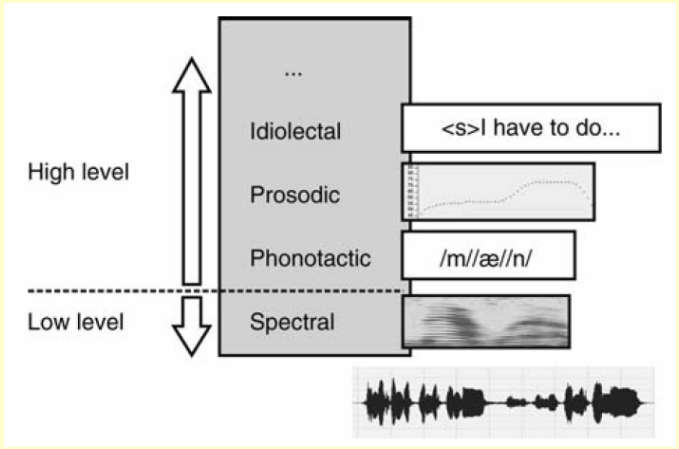


- At a high level of abstraction, Bayesian data analysis is extremely simple:
  - following the same, basic recipe: via Bayes Rule, we use the data to update prior beliefs about unknowns

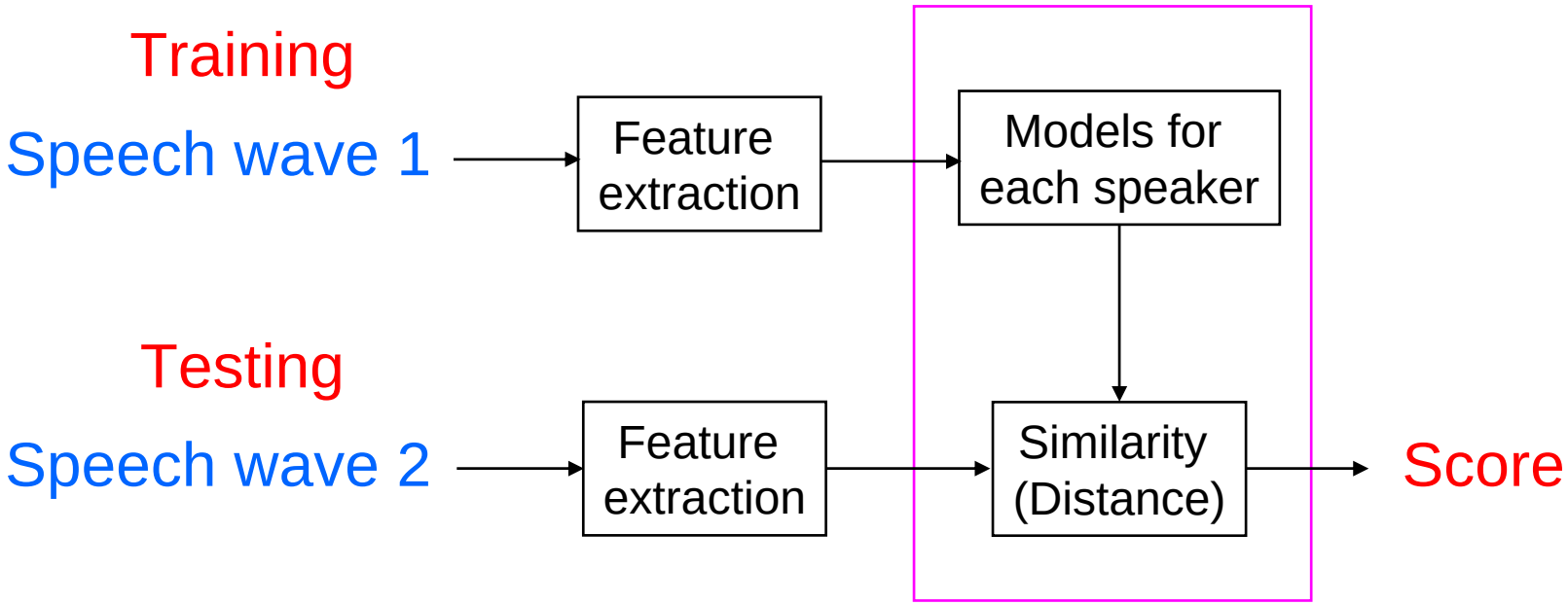
- There is much to be said on the implementation of this procedure in any specific application (e.g. FASR)
  - Freedom of choosing evidence evaluation and its value
  - Freedom of formulating propositions (and corresponding mathematical models) in relevance to the case
  - Freedom of choosing automatic speaker recognition method

Automatic Speaker Recognition

Speaker model is a representation of the identity of a speaker obtained from a speech utterance of known origin



Principal structure of speaker recognition systems



Text-dependent methods:

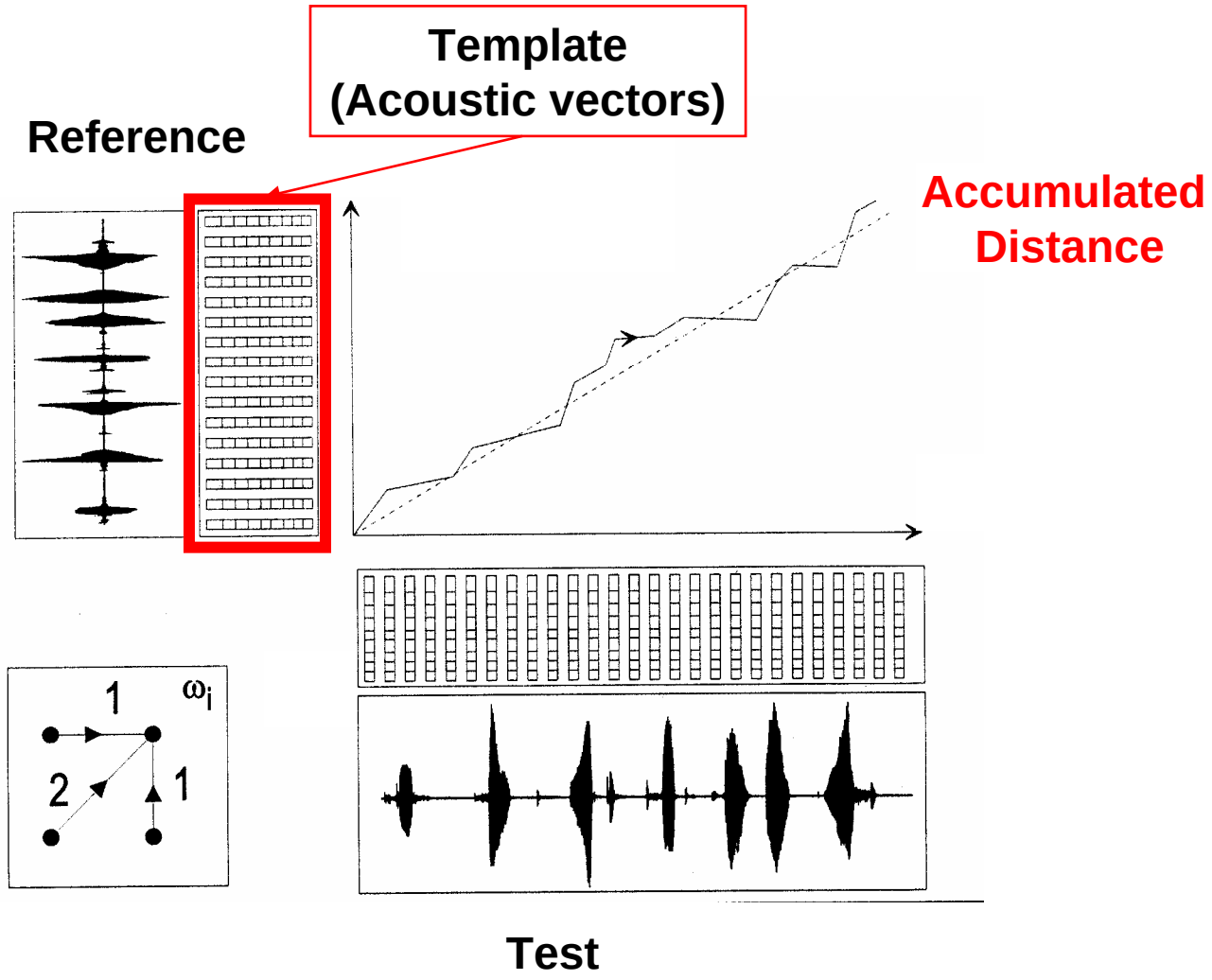
- **Dynamic Time Warping (DTW)**
- **Hidden Markov Models (HMMs)**

Text-independent methods:

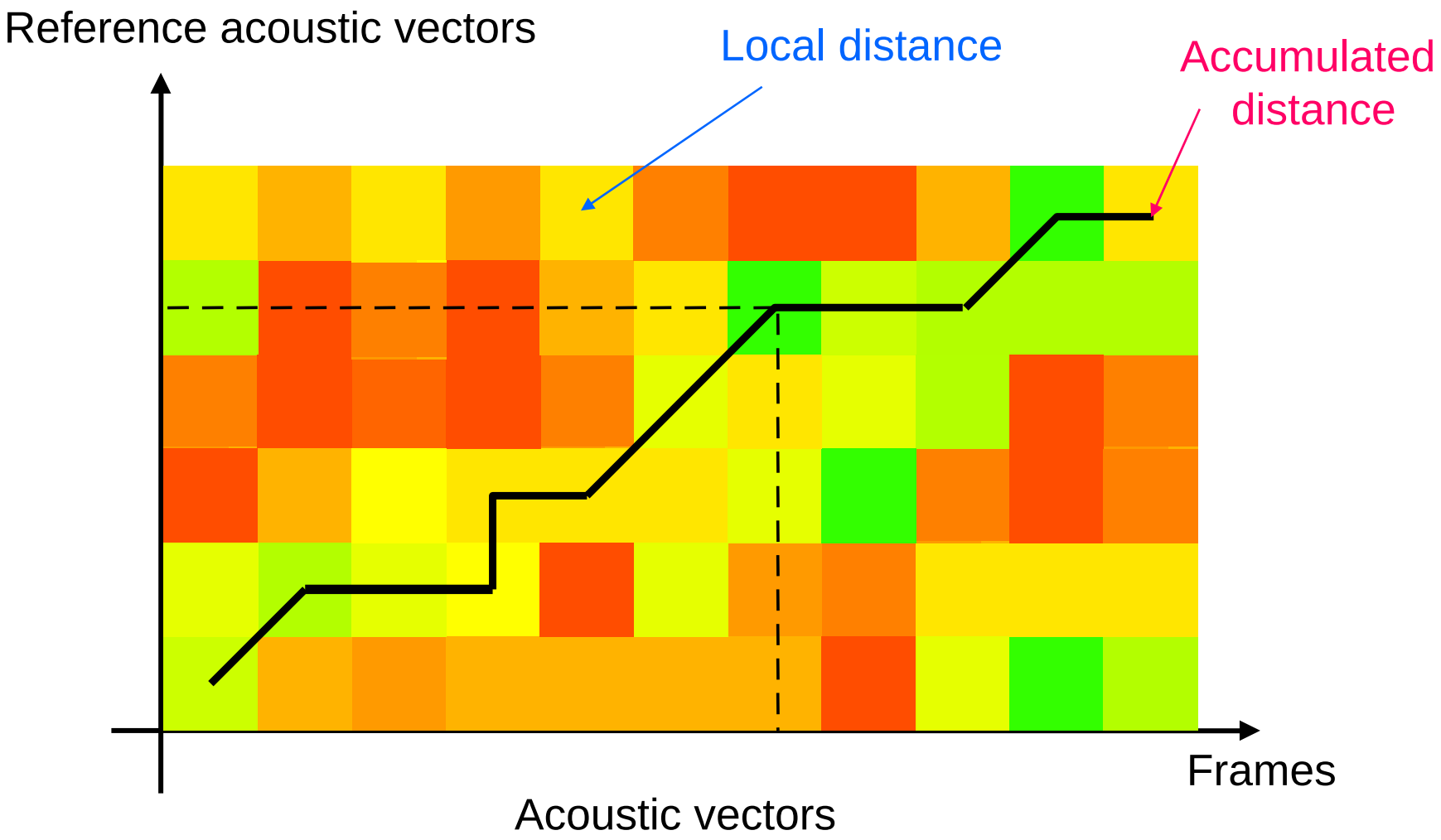
- **Vector Quantization (VQ)**
- **Gaussian Mixture Models (GMMs)**

- **Deterministic Methods**
  - Dynamic Time Warping (DTW)
  - Vector Quantization (VQ)
  - …
- **Statistical Methods**
  - Hidden Markov Model (HMM)
  - Gaussian Mixture Model (GMM)
  - …

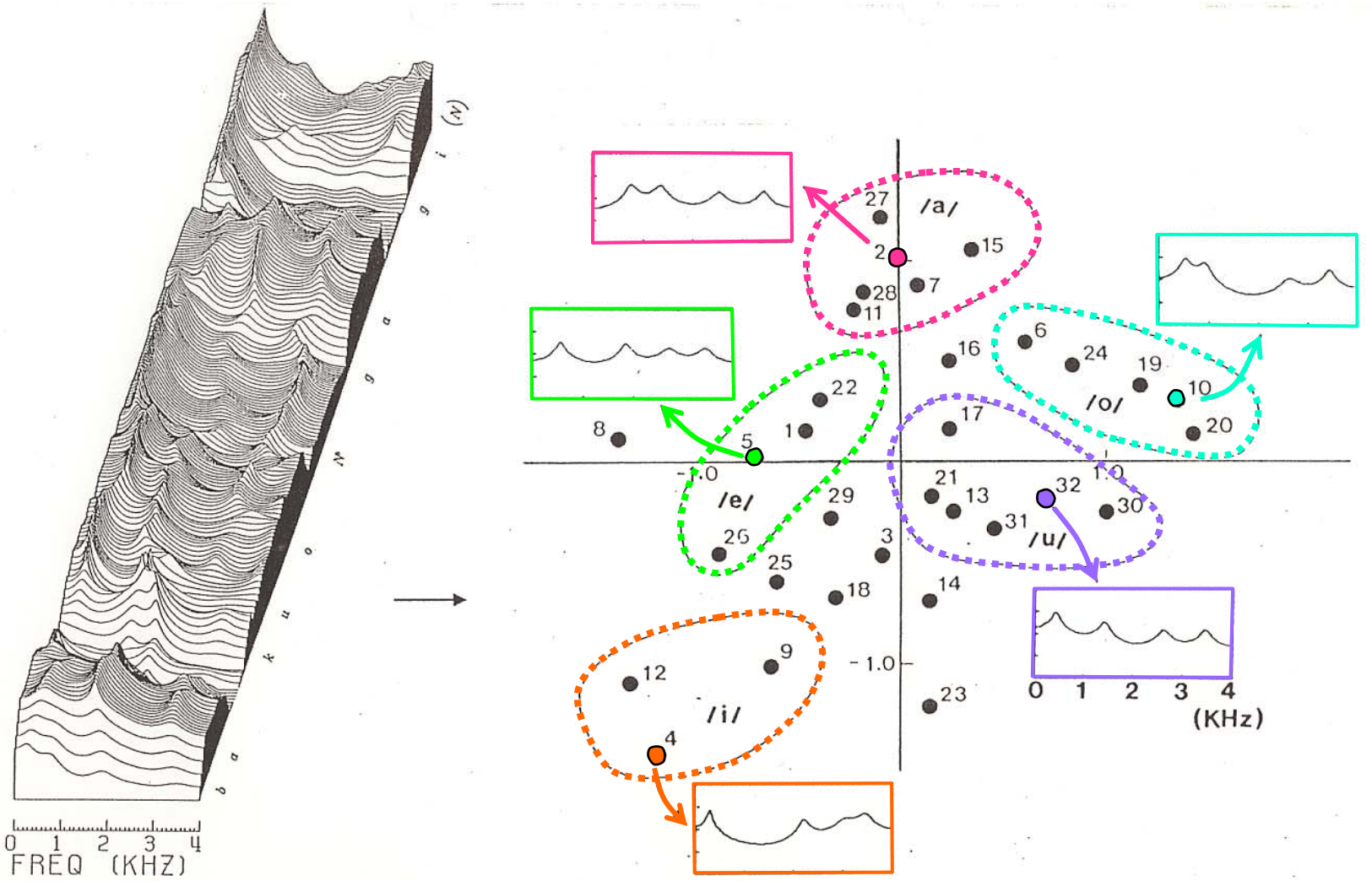
Dynamic Time Warping (DTW)



Dynamic Time Warping (DTW)



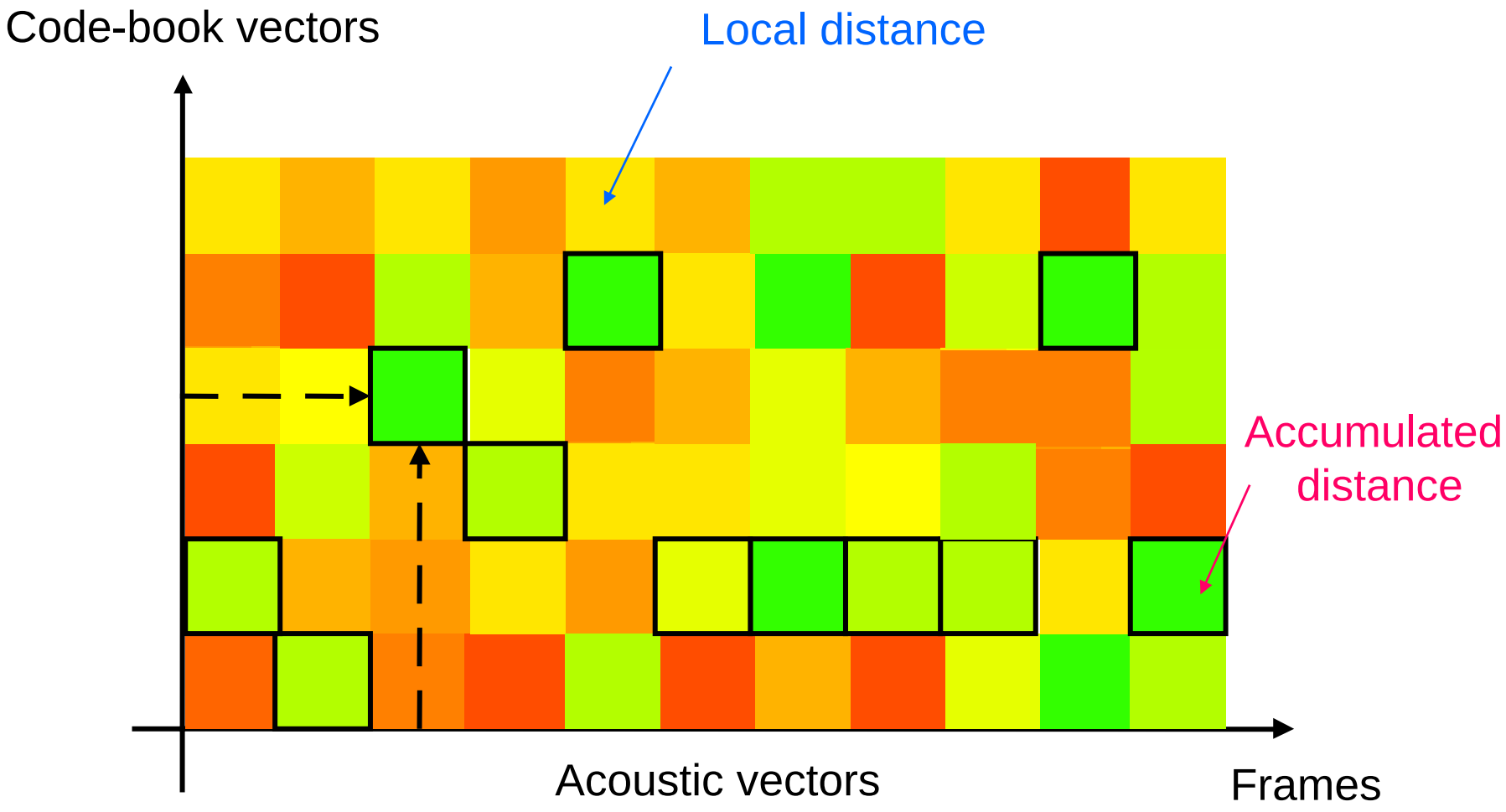
Vector Quantization (VQ)



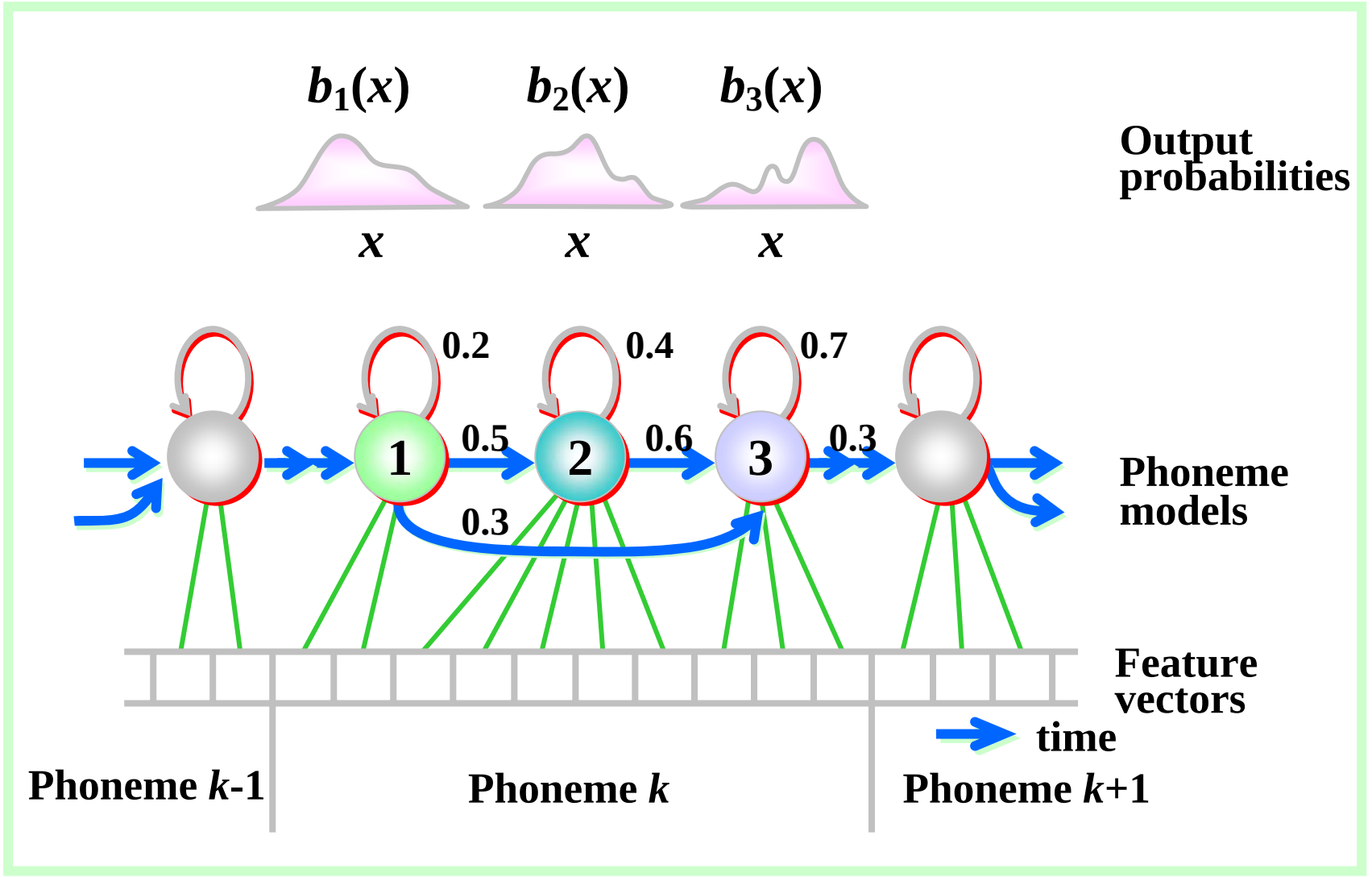
Spectral envelopes

Speaker-specific codebook

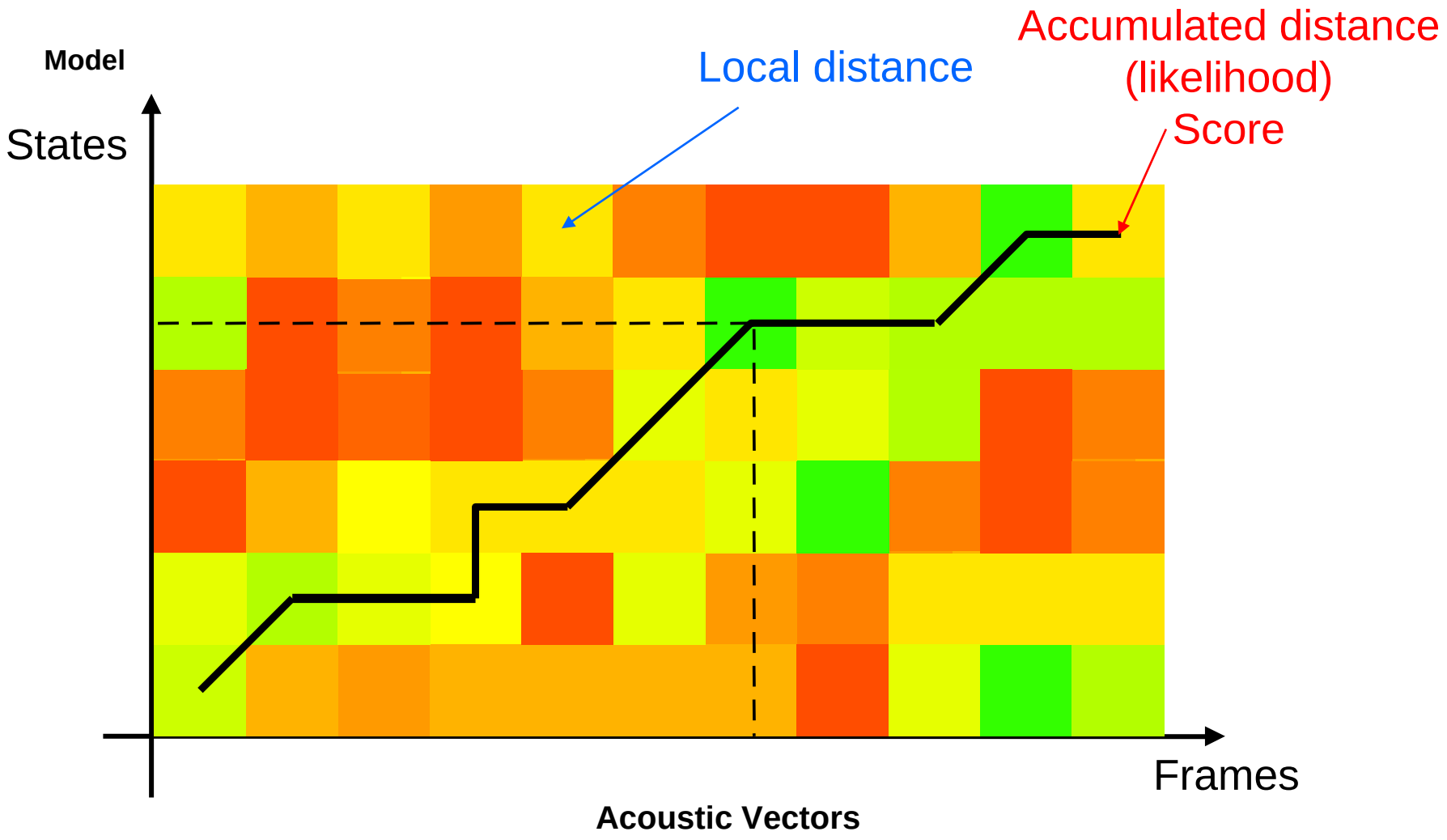
Vector Quantization (VQ)



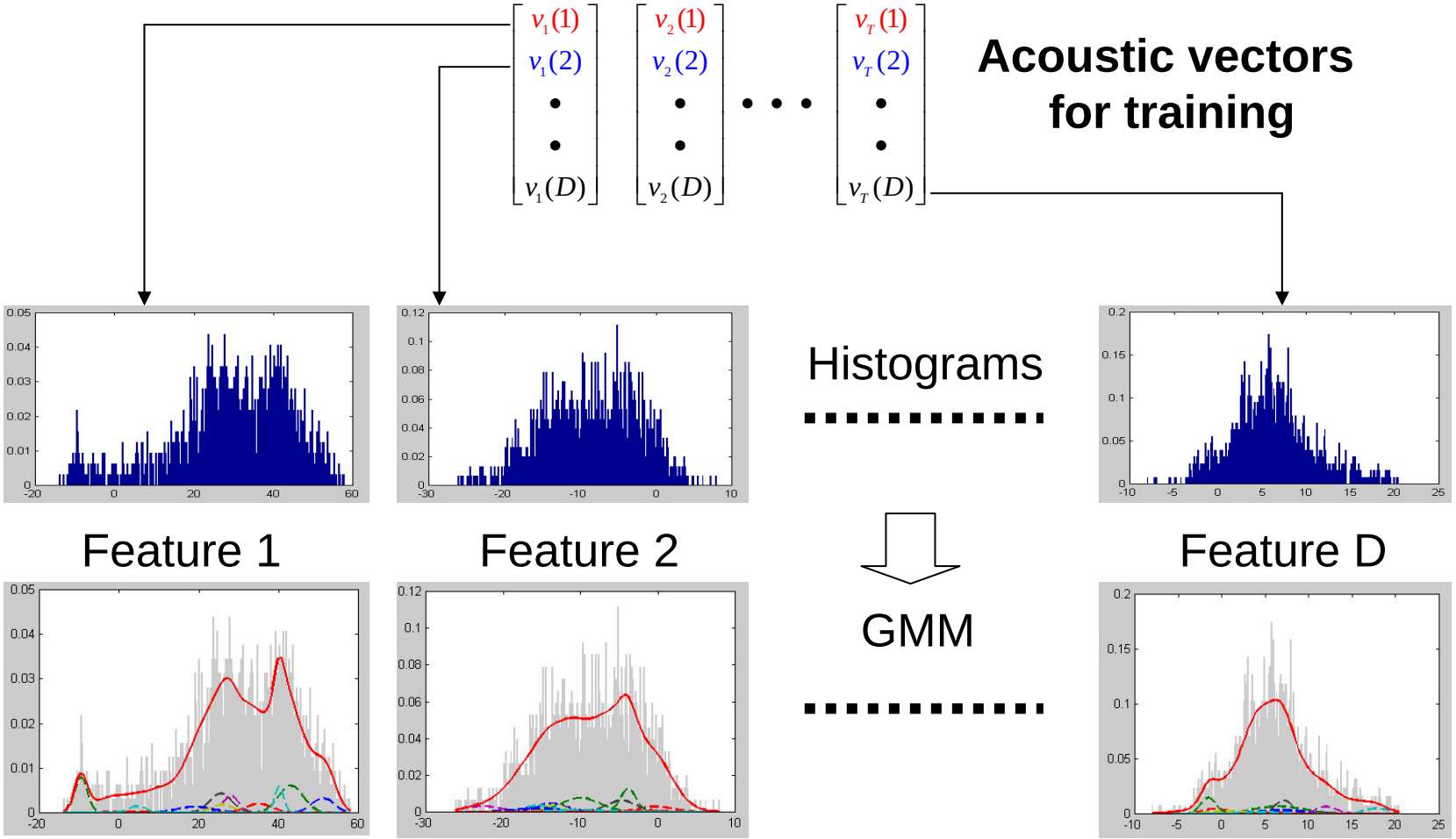
Hidden Markov Model (HMM)



Hidden Markov Model (HMM)

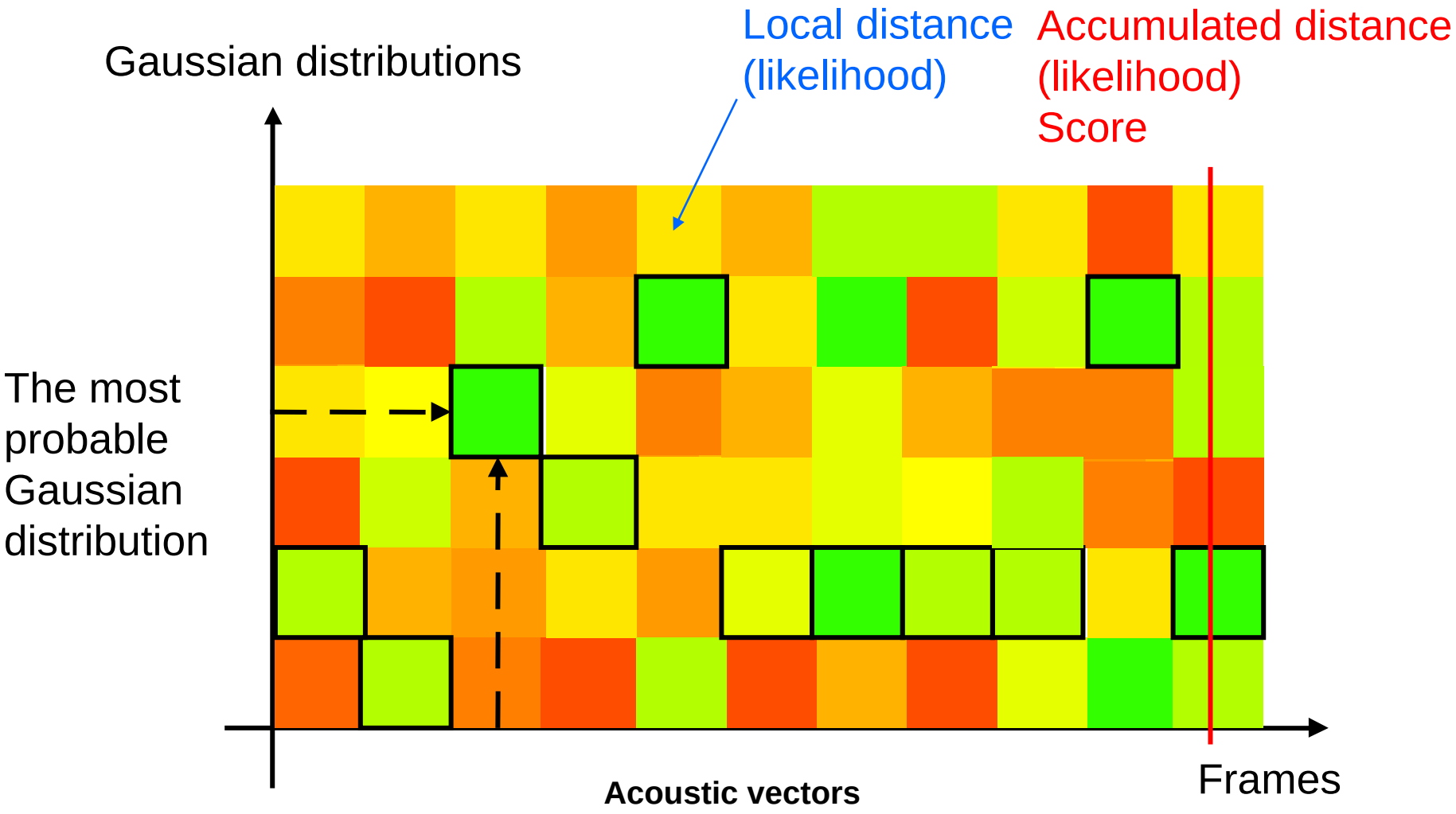


Gaussian Mixture Model (GMM)



score = likelihood (speech | model)

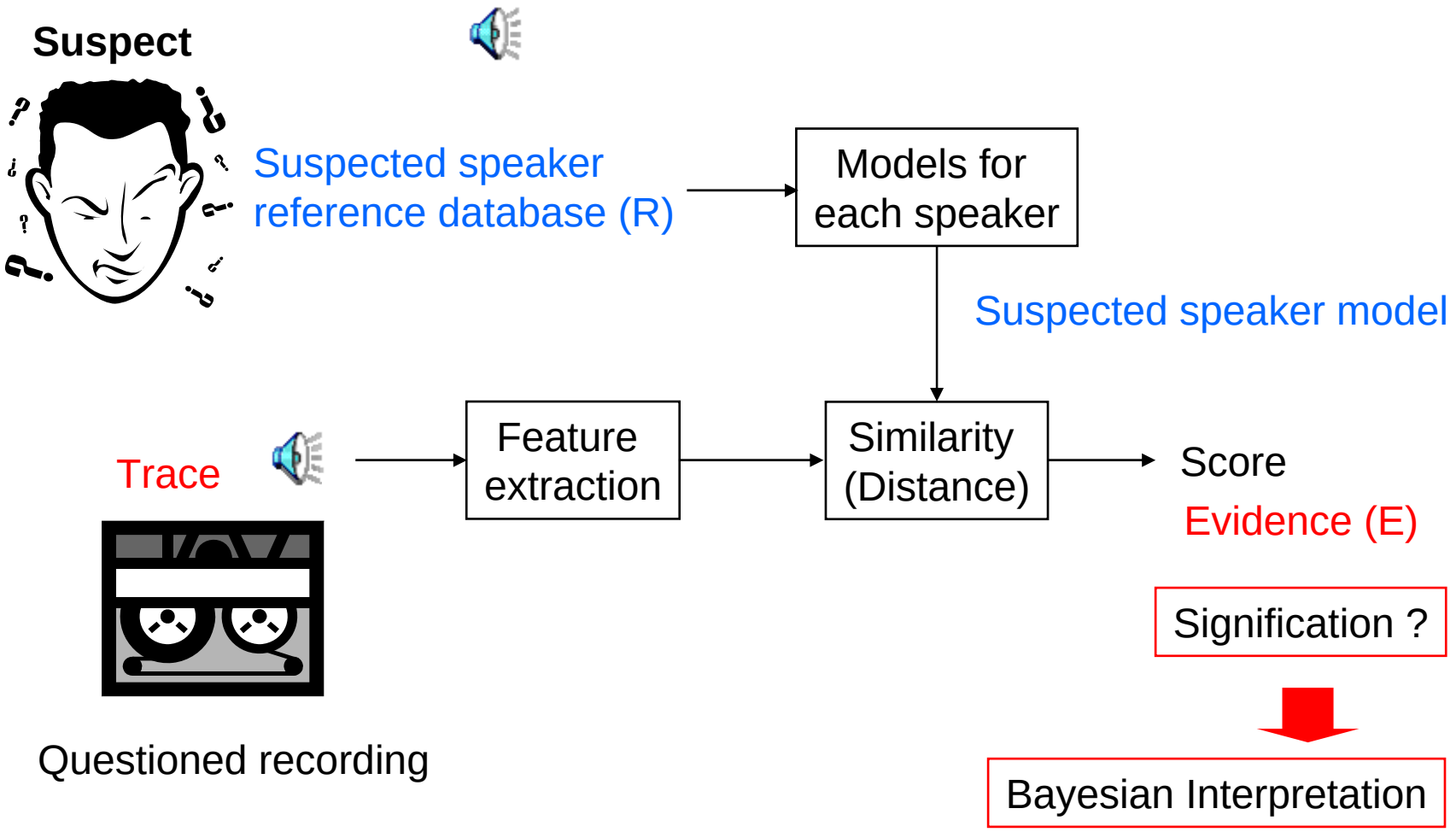
Gaussian Mixture Model



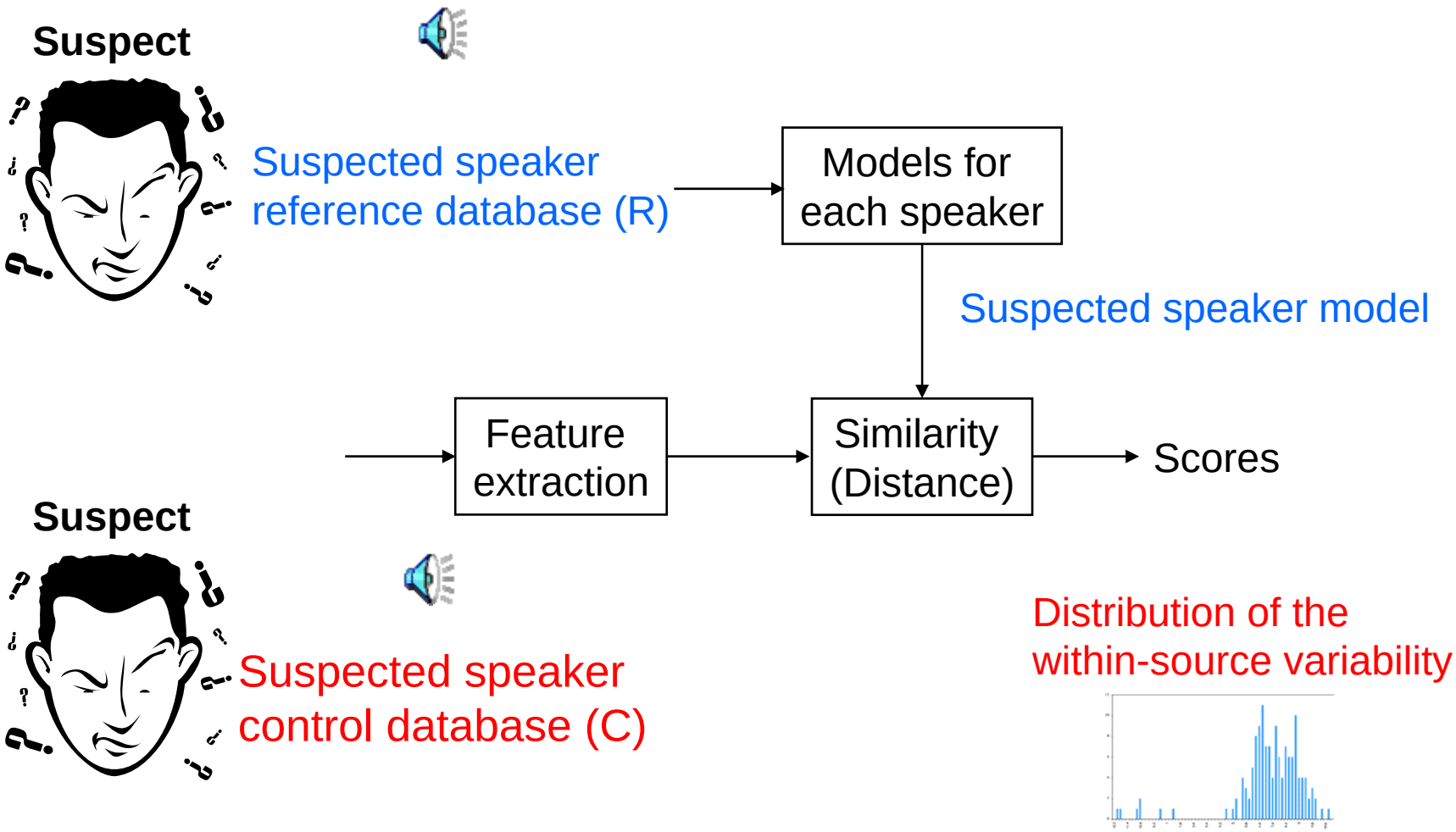
- In the case of questioned recording (trace), the **biometric evidence** does not consist in speech itself, but in the quantified degree of similarity between speaker dependent features extracted from the trace, and speaker dependent features extracted from recorded speech of a suspect, represented by his/her model.

Value of biometric evidence

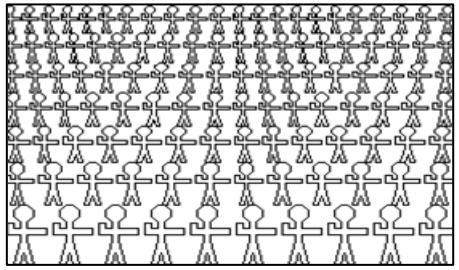
Univariate (Scoring) Method



Within-source variability



Between-sources Variability



Potential population database (P)

Models for each speaker

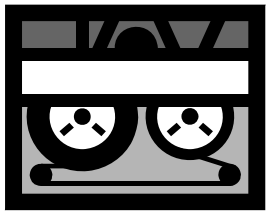
Speaker models of the potential population

Trace

Feature extraction

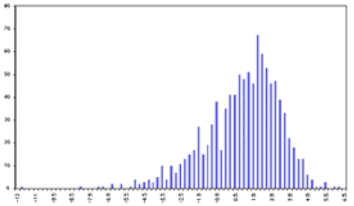
Similarity (Distance)

Scores

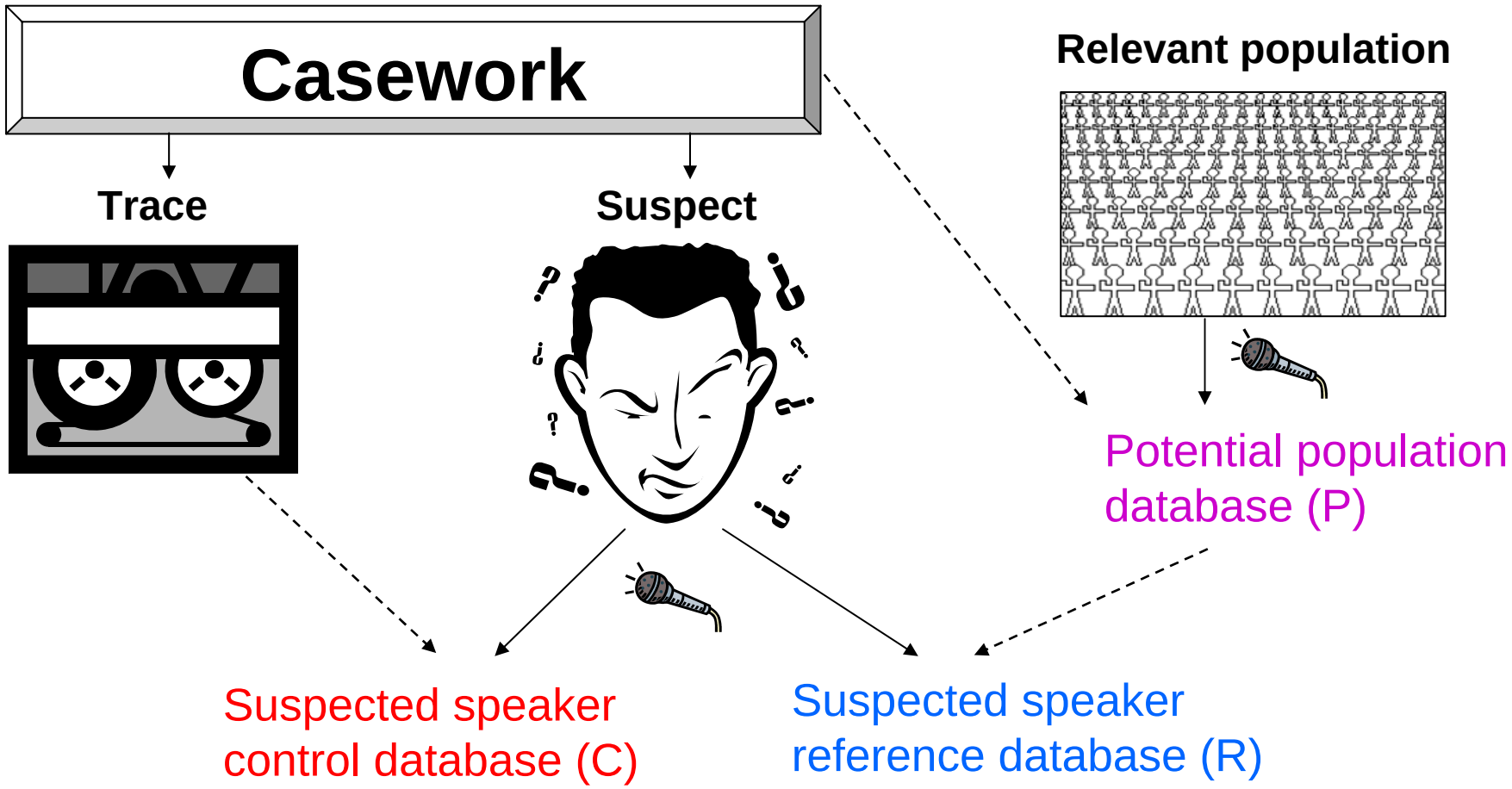


Questioned recording

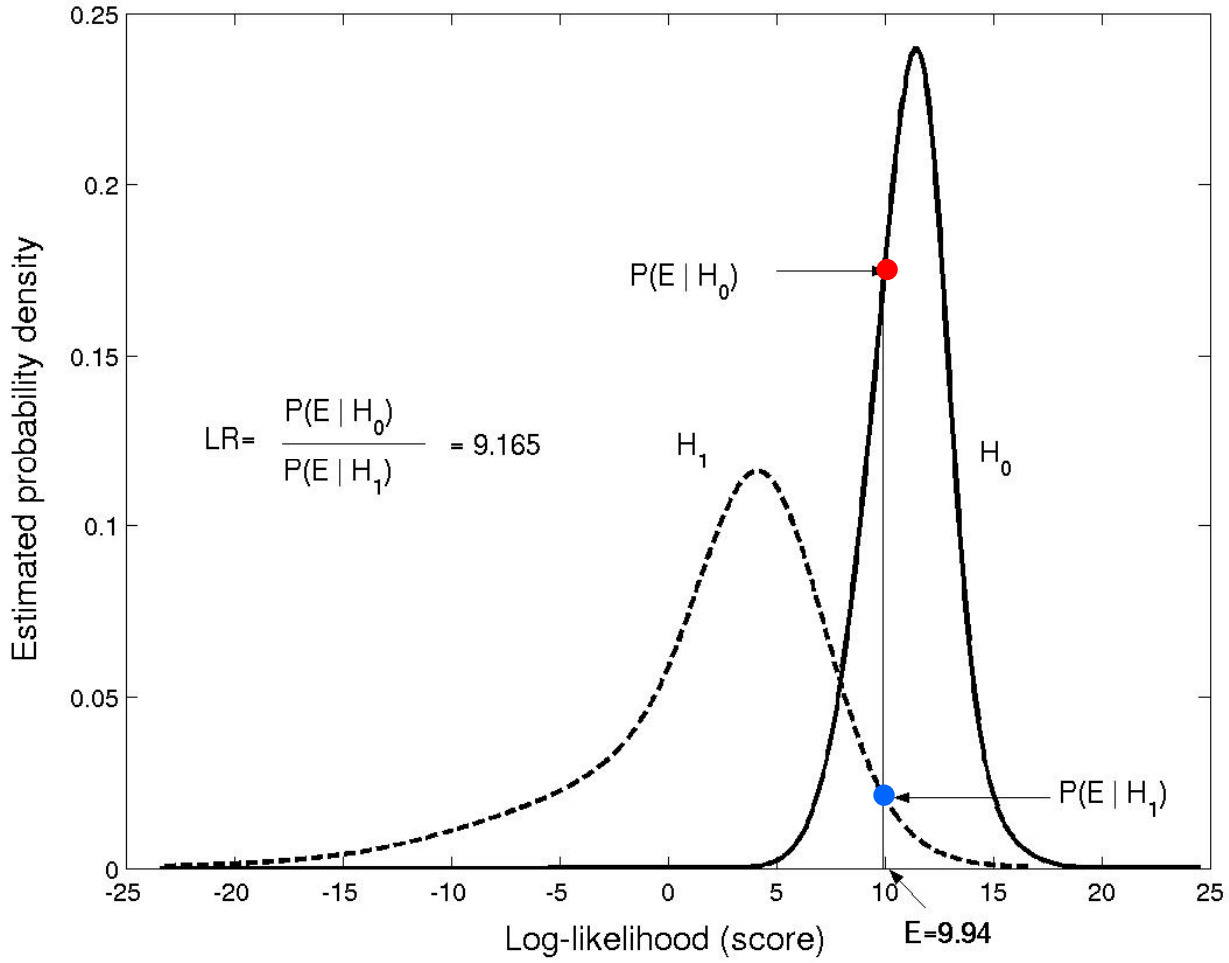
Distribution of the between-sources variability



Univariate (Scoring) Method



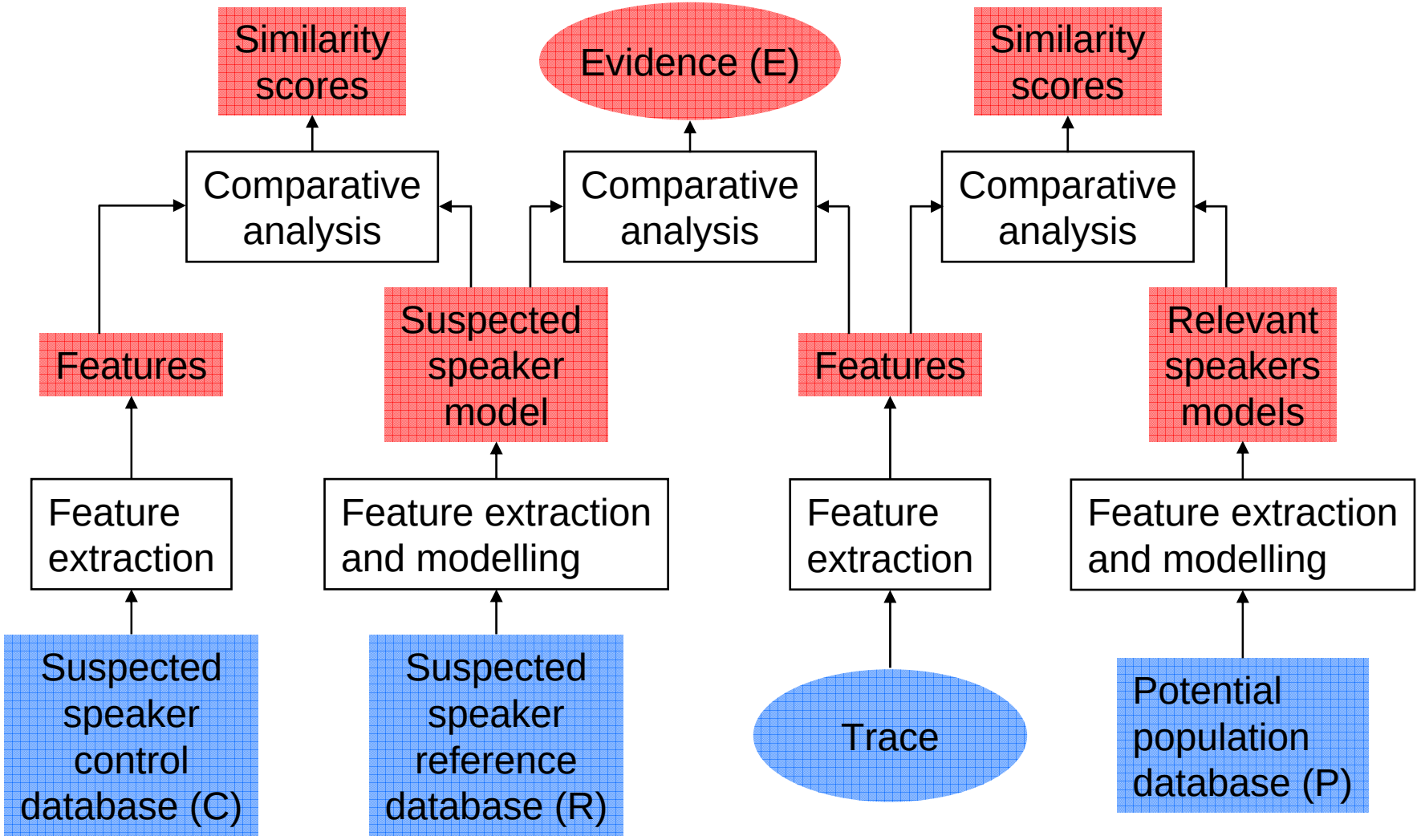
Strength of Evidence - Likelihood Ratio



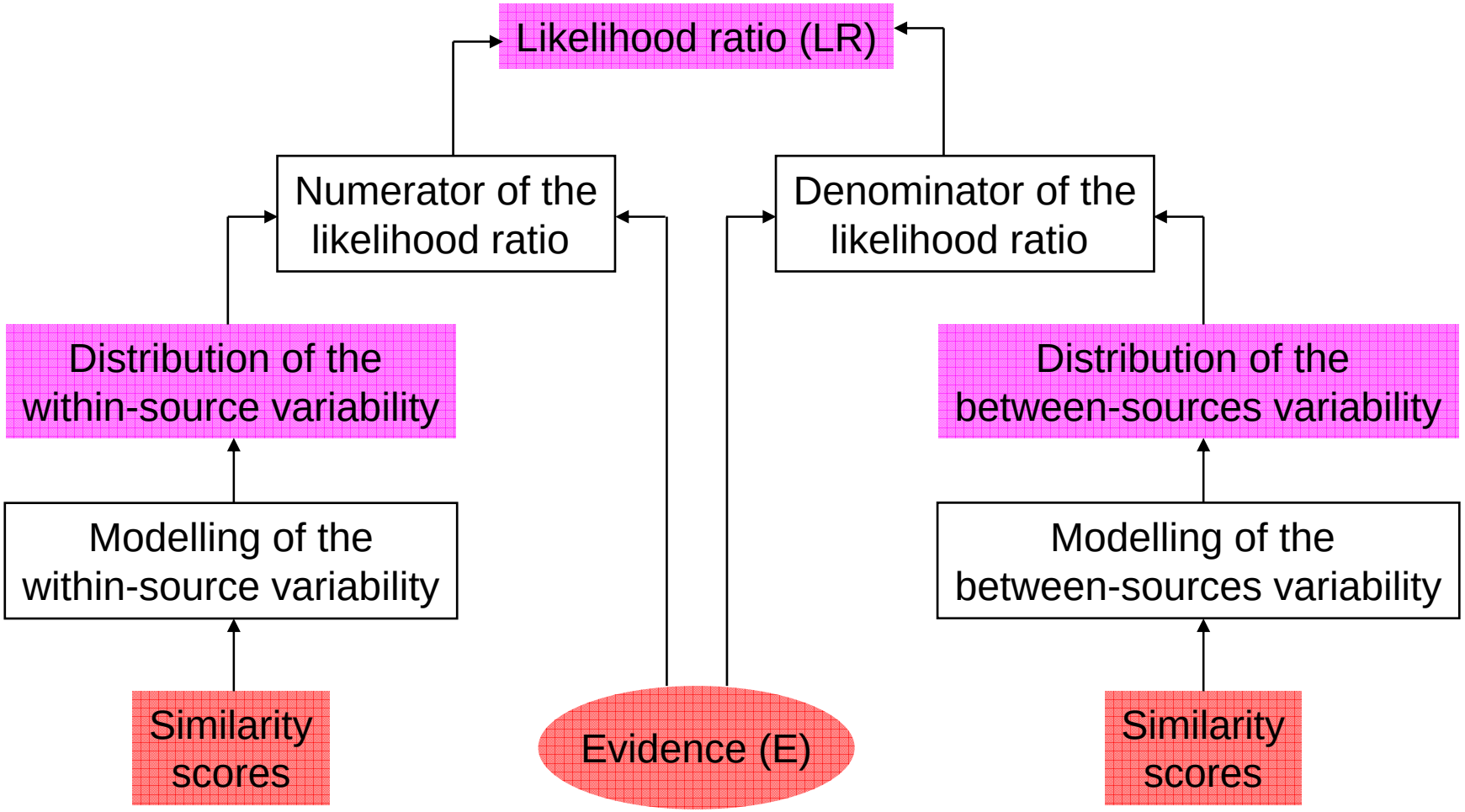
A **likelihood ratio** of 9.16 obtained means that it is **9.16 times more likely** to observe the score (E) given the hypothesis H_0 (the suspect is the source of the questioned recording) than given the hypothesis H_1 (that another speaker from the relevant population is the source of the questioned recording).

Interpretation of Biometric Evidence

Univariate (Scoring) Method

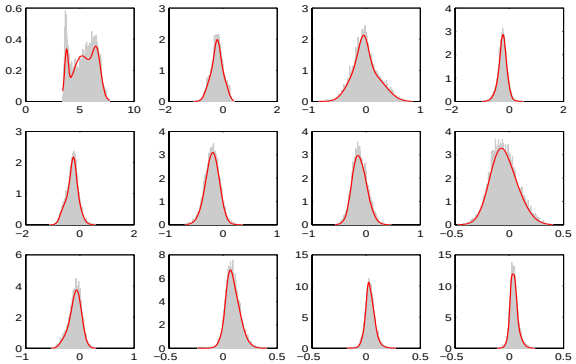


Interpretation of the evidence

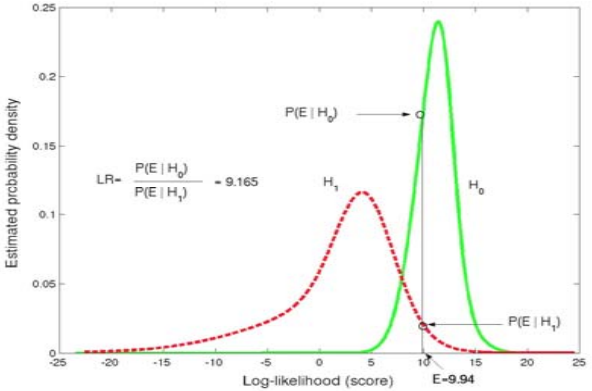


Evaluation of the Strength of Evidence

Double Statistical Model (Scoring method)

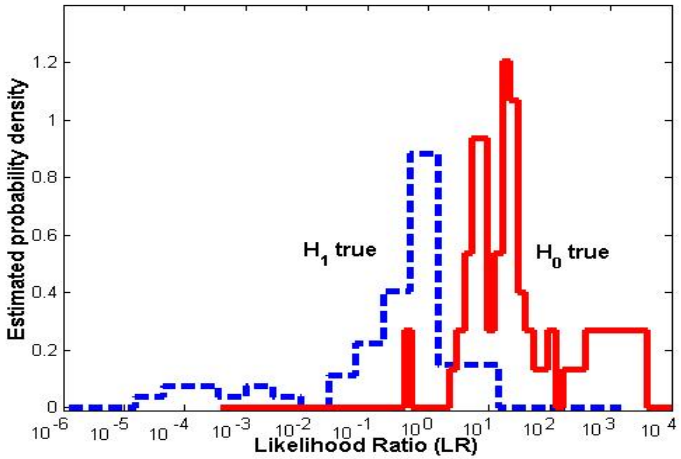


Individual case

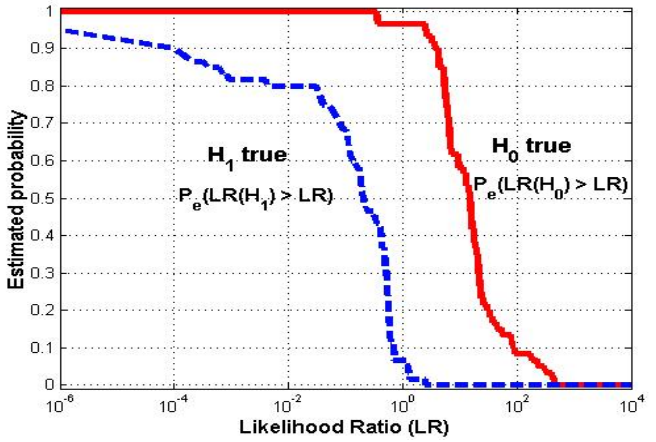


First level: GMM of the features

Second level: Models of within-source and between-sources variability



Performance across several cases



Distributions of LRs

Tippett Plots: Cumulative distributions of LRs

Multivariate (Direct) Method

- **The odds form of Bayes' theorem**
  - **H_0** – the speaker's model (λ_0) and the questioned recording (*T*) have the same source
  - **H_1** – the speaker's model (λ_1) and the questioned recording (*T*) have different sources

$$\frac{P(H_0)}{P(H_1)} \times \frac{P(T \mid H_0)}{P(T \mid H_1)} = \frac{P(H_0 \mid T)}{P(H_1 \mid T)}$$

Likelihood ratio $\boxed{\frac{P(T \mid \lambda_0)}{P(T \mid \lambda_1)}}$ Strength of evidence ?

Multivariate (Direct) Method

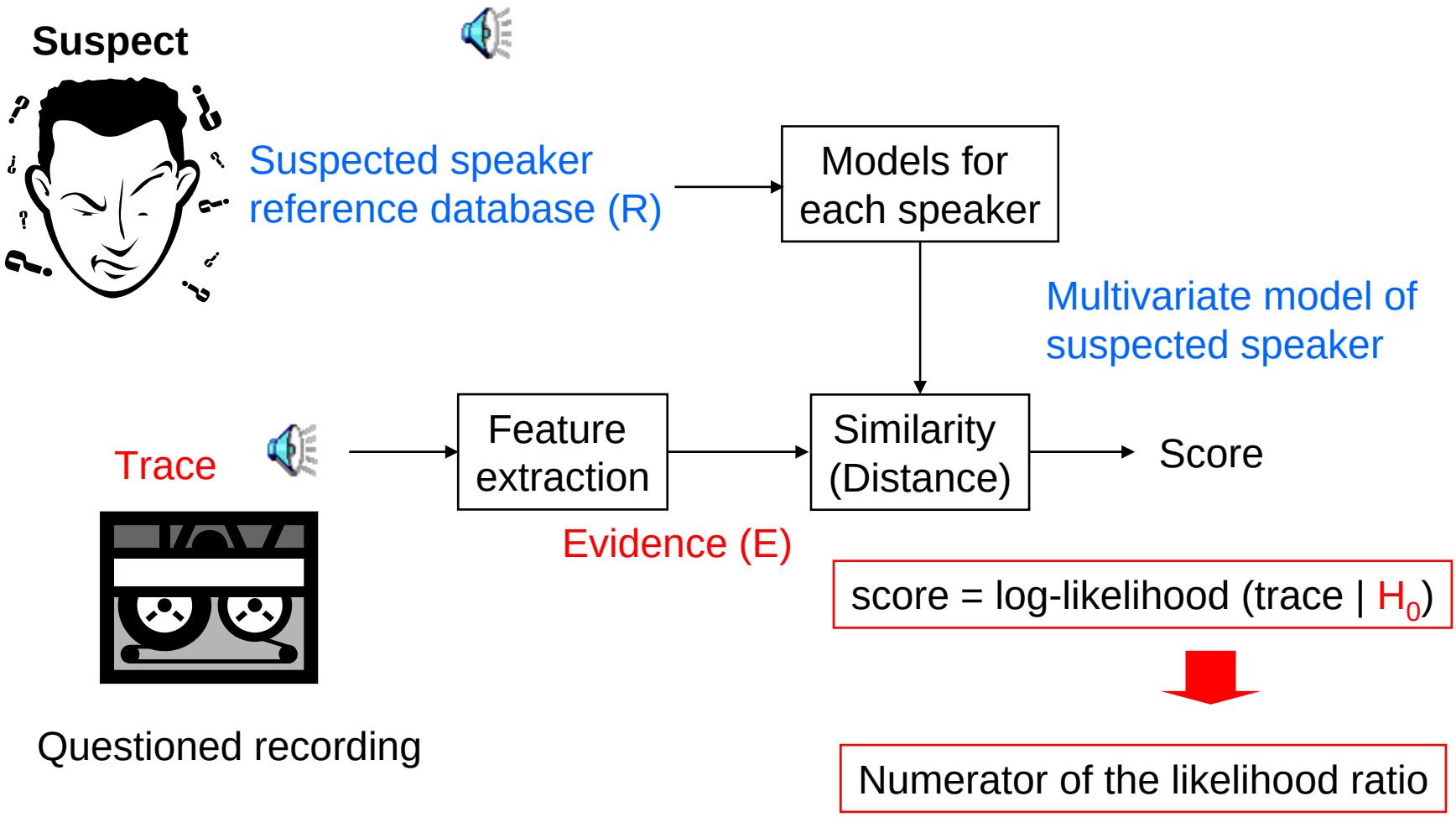
- **The odds form of Bayes' theorem**
  - **H_0** – the speaker's model (λ_0) and multivariate representation of trace (*T*) have the same source
  - **H_1** – the speaker's model (λ_1) and multivariate representation of trace (*T*) have different sources

$$\frac{P(H_0)}{P(H_1)} \times \frac{P(E \mid H_0)}{P(E \mid H_1)} = \frac{P(H_0 \mid E)}{P(H_1 \mid E)}$$

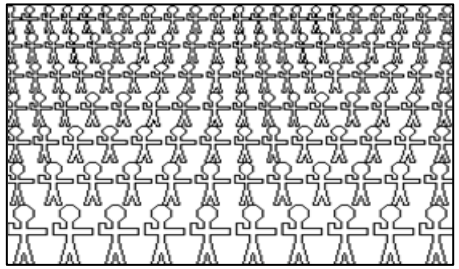
Likelihood ratio $\boxed{\frac{P(E \mid H_0)}{P(E \mid H_1)}}$ Strength of trace evidence with respect to new hypotheses

E – multivariate feature representation of trace evidence

Multivariate (Direct) Method – LR Numerator



Multivariate (Direct) Method – LR Denominator



Potential population database (P)

Model of all speakers

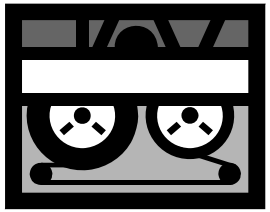
Multivariate model of potential population

Trace

Feature extraction

Similarity (Distance)

Score



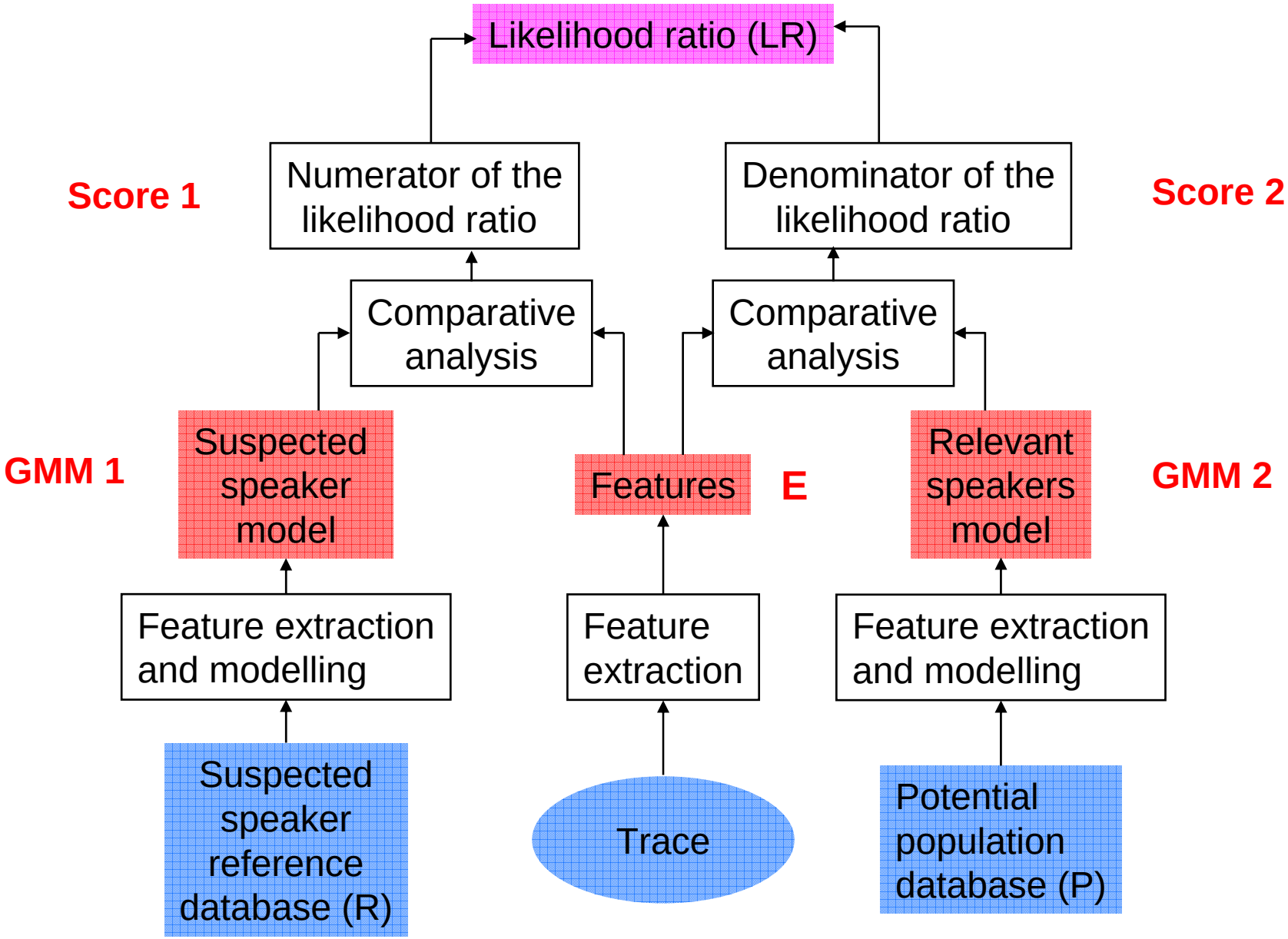
Questioned recording

Evidence (E)

score = log-likelihood (trace | H_1)

Denominator of the likelihood ratio

Multivariate (Direct) Method



Conclusions

- Statistical evaluation, and particularly **Bayesian methods** such as calculation of likelihood ratios based on automatic (deterministic and statistical) pattern recognition methods, have been criticized, but they are the only demonstrably rational means of quantifying and evaluating the value of biometric evidence available at the moment.
- The **data-driven based methodology** provides a coherent way of assessing and presenting the biometric evidence of questioned recording.
- The **future methods** to be developed for interpretation of voice as **forensic evidence** should combine the advantages of automatic signal processing and pattern recognition objectivity with the methodological transparency solicited in forensic investigations.

References (1998-2009)

- D. Meuwly, M. El-Maliki, A. Drygajlo, "Forensic Speaker Recognition Using Gaussian Mixture Models and a Bayesian Framework", COST 250 Workshop on Speaker Recognition by Man and by Machine: Directions for Forensic Applications, Ankara, Turkey, April **1998**, pp. 52-55.

- D. Meuwly, A. Drygajlo, "Forensic Speaker Recognition Based on a Bayesian Framework and Gaussian Mixture Modelling (GMM)", The Workshop on Speaker Recognition “2001: A Speaker Odyssey”, Crete, Greece, June, 2001, pp. 145-150.

- D. Meuwly, "Reconnaissance de locuteurs en sciences forensiques: l’apport d’une approche automatique“, PhD thesis, IPSC, University of Lausanne, 2001.

- A. Drygajlo, D. Meuwly, A. Alexander, "Statistical Methods and Bayesian Interpretation of Evidence in Forensic Automatic Speaker Recognition", EUROSPEECH'2003, Geneva, Switzerland, Sept. 2003, pp. 689-692.

- A. Alexander, A. Drygajlo, "Scoring and Direct Methods for the Interpretation of Evidence in Forensic Speaker Recognition“, ICSLP 2004, Jeju, Korea, 2004.

- A. Alexander, F. Botti, D. Dessimoz, A. Drygajlo, "The Effect of Mismatched Recording Conditions on Human and Automatic Speaker Recognition in Forensic Applications", Forensic Science International, 146S (2004), pp. S95-S99.

References (1998-2009)

- A. Alexander, D. Dessimoz, F. Botti, and A. Drygajlo, "Aural and Automatic Forensic Speaker Recognition in Mismatched Conditions", The International Journal of Speech, Language and the Law, vol. 12, Dec. 2005, pp. 214-234.

- M. Arcienega, A. Alexander, P. Zimmermann, A. Drygajlo, "A Bayesian Network Approach Combining Pitch and Spectral Envelope Features to Reduce Channel Mismatch in Speaker Verification and Forensic Speaker Recognition", InterSpeech 2005, Lisbon, Portugal, Sept. 4-8, 2005.

- J. Gonzalez-Rodriguez, A. Drygajlo, D. Ramos-Castro, M. Garcia-Gomar, J. Ortega-Garcia, "Robust Estimation, Interpretation and Assessment of Likelihood Ratios in Forensic Speaker Recognition", invited paper, Computer Speech and Language, vol. 20, 2006, pp. 331-355.

- A. Drygajlo, "Forensic Automatic Speaker Recognition", IEEE Signal Processing Magazine, 24 (2): 132-135 (2007).

- A. Drygajlo, “Statistical Evaluation of Biometric Evidence in Forensic Automatic Speaker Recognition", 3rd International Workshop on Computational Forensics (IWCF 2009), The Hague, The Netherlands: 1-12 (2009).

- A. Drygajlo, "Forensic evidence of voice", Chapter in S.Z. Li (Ed.), "Encyclopedia of Biometrics", Springer, New York, August **2009,** pp. 1388-1396