Forensic Speaker Recognition

Based on

a Bayesian Framework and

Gaussian Mixture Modelling

(GMM)
Outline

• Inference of identity
  – Bayesian interpretation vs speaker identification / verification

• Corpus based methodology
  – 3 databases
    • Potential population (P) – Control (C) – Reference (R)

• Automatic speaker recognition method
  – Gaussian Mixture Modelling (GMM)

• Experiment
  – Calculation of a likelihood ratio

• Evaluation of the system performance
  – Tippet plots
Our contribution

– IAFP Conference, Edinburgh 1997
  • D. Meuwly, A. Drygajlo "Likelihood Ratios for Automatic Speaker Recognition in Forensic Applications"

  • D. Meuwly, M. El-Maliki, A. Drygajlo "Forensic Speaker Recognition Using Gaussian Mixture Models and a Bayesian Framework"

– RL2C, Avignon 1998
  • C. Champod, D. Meuwly "The Inference of Identity in Forensic Speaker Recognition"

– ENFSI Meeting, Krakow 2000
  • D. Meuwly, A. Drygajlo "The Influence of the Telephone Network on Automatic Forensic Speaker Recognition"
• Speaker verification
  – Discrimination task based on a threshold
    • Discrimination: rejection
    • No discrimination: acceptation

  – Legal interpretation of the notion of threshold
    • Acceptable level of reasonable doubt

  – Speaker verification for forensic individualisation
    • Not adequate: If the random match probability is not null, the conclusion of identification is inadequate and misleading
Inference of identity (2)

• Speaker identification
  – Closed-set of speakers
    • Not adequate: The assessment of the credibility of the exhaustiveness of the number of suspects is outside the duties of the expert
  – Open-set of speakers
    • Not adequate: this framework implies a final discrimination decision based on a threshold and suffers from the same drawbacks as the verification task
Inference of identity (3)

• Bayesian interpretation (BI)
  – Principle
    • The Bayesian model, proposed for forensic speaker recognition by Lewis in 1984, allows for revision of a measure of uncertainty about the truth or falsity of an issue.
    • This approach shows how data can be combined with prior background knowledge (prior odds) and new data (questioned recording) to give posterior odds for judicial outcomes or issues.

Prior odds * ? = posterior odds
• Bayesian interpretation (BI)
  – Calculation of the evidence (E)
    \( x \) – speaker dependent features extracted from the questioned recording (X)
    \( y \) – speaker dependent features extracted from utterances of the suspected speaker (Y)

  \( E \) – result of the comparative analysis of \( x \) and \( y \) using an automatic speaker recognition method
• Bayesian interpretation (BI)
  – Concept of likelihood ratio
    - $H_1$ – the suspected speaker is the source of the questioned recording
    - $H_2$ – the source of the questioned recording is another speaker of the potential population

Evaluation of $E$ according to the hypotheses $H_1$ and $H_2$

$P(E \mid H_1)$ – Probability of $E$ when $H_1$ is verified
$P(E \mid H_2)$ – Probability of $E$ when $H_2$ is verified

\[
\frac{P(E \mid H_1)}{P(E \mid H_2)} \quad \text{– Likelihood ratio of } E \text{ according to } H_1 \text{ and } H_2
\]
Inference of identity (6)

- Bayesian interpretation (BI)

\[
\frac{P(H_1)}{P(H_2)} \times \frac{P(E|H_1)}{P(E|H_2)} = \frac{P(H_1|E)}{P(H_2|E)}
\]

- Prior background knowledge
- New Data
- Posterior knowledge on the issue

- A priori probability ratio
- Likelihood Ratio (LR)
- A posteriori probability ratio

- Province of the court
- Province of the forensic scientist
- Province of the court
• 3 databases (DBs)
  – Potential population database (P)
    • Large-scale database used to model the potential population of speakers to evaluate the between-source variability
  – Suspected speaker reference database (R)
    • Database recorded with the suspected speaker to model its speech
  – Suspected speaker control database (C)
    • Database recorded with the suspected speaker to evaluate the within-source variability
Automatic speaker recognition

• Principle
  – Text-independent automatic speaker recognition method

• Feature vectors
  – 12 perceptual linear prediction coefficients (PLP)

• Modelling and recognition method
  – Gaussian Mixture Models (64 density functions)
Experiment (1)

• Simulated questioned recording
  – Anonymous call: 10 seconds, PSTN, SNR = 40dB
  – Aural analysis
    • Spoken language and accent: Swiss French
    • Assumption on the sex and the voice disguise: male speaker, no perceptible voice disguise

• Selection of the potential population database (P)
  – Population: $10^6$ Swiss-French male speakers
  – Database: 1000 male speakers of the Swiss-French Polyphone DB
    • 1 session of 100–140 seconds of speech per speaker
    • Used to calculate 1000 Gaussian mixture speaker models
• Selection of a Swiss-French suspected male speaker
  – In the experiment, the suspected speaker is the source of the questioned recording

• Recording of the suspected speaker reference DB (R)
  – Six speech sessions of 100–140 seconds recorded during 2 months
  – The sessions are recorded in the same way as the Swiss-French Polyphone DB
  – Used to calculate 6 speaker models ($\lambda_1$ – $\lambda_6$)
Experiment (3)

• Calculation of the evidence
  – Comparison of the questioned recording with the GMM of the suspected speaker ($\lambda_1$)
  – Result: $E = \log P(x | \lambda_1) = 6$

• Interpretation of the evidence
  – Recording of the suspected speaker control DB (C)
    • 35 utterances: simulation of 5 phone discussions and description of 30 pictures recorded through the PSTN
    • Used to evaluate the within-source variability of the suspected speaker
Experiment (4)

- Evaluation of the within-source variability
  - The 6 suspected speaker models ($\lambda_1 - \lambda_6$) are compared to the C database.
Experiment (5)

- Approximation of the within-source variability with kernel density estimation (KDE)
Experiment (6)

- Numerator of the likelihood ratio
  - Calculation of $P(E | H_1)$ for $E = 6$
Experiment (7)

• Evaluation of the between source variability
  – The 1000 suspected speaker models of the P database are compared to the questioned recording
Experiment (8)

- Approximation of the between-source variability with kernel density estimation (KDE)
Experiment (9)

- Denominator of the likelihood ratio
  - Calculation of $P(E \mid H_2)$ for $E = 6$
• Calculation of the likelihood ratio
  \[ \frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{0.15}{0.002} = 75 \]
Experiment (11)

• A likelihood ratio of 75 means that the forensic analysis of the questioned recording allows to revise the prior odds defined by the judge by multiplying them by 75:

• Supposing an *a priori* probability of $p(H_1) = 0.5$ (prior odds : 1/1):

\[
\begin{align*}
\text{prior odds} & \quad \times \quad 75 & = & \quad \text{posterior odds} \\
\frac{1}{1} & \quad \times \quad 75 & = & \quad \frac{75}{1}
\end{align*}
\]
Experiment (11)

\[ p(H_1|E) = \frac{\text{posterior odds}}{1 + \text{posterior odds}} \]

\[ \frac{75}{76} = 0.986 \]
• Principle
  – Estimation and comparison of likelihood ratios obtained for the evidence E:

  – When $H_1$ is verified
    • The suspected speaker truly is the source of the questioned recording

  – When $H_2$ is verified
    • The suspected speaker truly is not the source of the questioned recording
Evaluation of the system performance (2)

- **Tippet plots**
  - Representation of the results proposed in the field of interpretation of forensic DNA analysis.

  **Vertical axis:** Indication of the probability that the result of the experiment exceeds a given value of likelihood ratio.

  **Horizontal axis:** Graduation with increasing values of likelihood ratios.

- **$H_2$**
- **$H_1$**
• **Example of evaluation with a Tippet plot**
  – Simulated questioned recordings
    • 8 utterances of 10 seconds recorded by 8 Swiss-French male speakers
  – **Tests when \( H_1 \) is verified**
    • *Comparison of the simulated questioned recordings, each with the corresponding speaker model and calculation of the likelihood ratios*
  – **Tests when \( H_2 \) is verified**
    • *Comparison of the simulated questioned recordings, each with the 1000 male speaker models of the Polyphone DB and calculation of the likelihood ratios*
Results presented with a Tippet plot

- This way of presentation illustrates simultaneously the performance of the forensic speaker recognition system when the hypotheses $H_1$ or $H_2$ are verified.

Evolution of the LR when the hypothesis $H_1$ is verified (N = 48)

Evolution of the LR when the hypothesis $H_2$ is verified (N = 8000)
Conclusion

- The Bayes model, current interpretation framework used in forensic science, is adapted for forensic automatic speaker recognition.

- The corpus based methodology, introduced in this paper, provides a coherent way of assessing and interpreting the evidence of questioned recording.

- The Tippet plot, is a graphical representation appropriate for the evaluation of the performance of the forensic recognition system.