A Bayesian Network Approach Combining Pitch and Spectral Envelope Features to Reduce Channel Mismatch in Speaker Verification and Forensic Speaker Recognition

Mijail Arcienega, Anil Alexander, Philipp Zimmermann*, Andrzej Drygajlo  
Interspeech 2005, Lisboa, Portugal

**Context:**
- Channel mismatch
- Speaker verification
- Forensic speaker recognition

**Key Points:**
- Bayesian networks
- Combining pitch and spectral envelope features

**The Bayesian Network**

![Bayesian Network Diagram]

- Pitch \( \hat{\varphi} \) and spectral envelope features \( \hat{x} \) are conditionally independent, given the voicing status \( s \).

- \( p(\hat{x}|\hat{\varphi}, s) = p(\hat{x}|s) \)

- \( p(\hat{\varphi}|s = i) = w_i \)
- \( p(\hat{\varphi}|s = 0) \) defined by two Gaussian mixtures \( \lambda_1^\varphi \)
- \( p(\hat{\varphi}|s = 1) \) defined by one Gaussian mixture \( \lambda_0 \)
- \( p(\varphi = 0|s = 2) = 1 \); \( p(\varphi \neq 0|s = 2) = 0 \).

**Likelihood Estimation**

**Definitions:**
- \( O = \{ \eta_1, ..., \eta_T \} \) Set of testing data
  - \( \eta_t = \{ \hat{\varphi}_t, \hat{x}_t \} \)
- \( S = \{ s_1, ..., s_T \} \) Set of voicing status values

**Likelihood Expression:**
- \( p(O|S, \lambda) = p(X|S, \lambda) \cdot p(P|S, \lambda) \)
  - \( p(O|S, \lambda) = p(X_V|\lambda_1^\varphi) \cdot p(X_U|\lambda_2^\varphi) \cdot p(P_V|\lambda^0) \)

**The Database: EPFL-IPSC03**

Forensic speaker recognition database (EPFL-IPSC03)

- Six speech segments (15 to 180 seconds) for 60 Swiss French speakers which include recordings through:
  - switched public telephone network (PSTN).
  - global system for mobile communications (GSM).
  - direct recording in the calling room, via a digital recorder (room).

**Speaker Verification Results**

**Forensic Speaker Recognition Evaluation**

**CONCLUSIONS:**
- Convolutional modifications introduced by the PSTN or GSM channel severely affect the spectral envelope features but have almost no influence on the pitch values.
- The Bayesian network efficiently combines both features and improves both speaker verification and forensic speaker recognition systems.

---

Signal Processing Institute  
Swiss Federal Institute of Technology, Lausanne

*Institut de Police Scientifique  
University of Lausanne