Testing the validity and reliability of forensic voice comparison based on reassigned time-frequency representations of Chinese /iau/

Ewald Enzinger
School of Electrical Engineering & Telecommunications, The University of New South Wales, Sydney, Australia
Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria
Paradigm for evaluation of FVC evidence

• Likelihood-ratio framework:
  – Statement of strength of the evidence as an answer to a specific question
    \[ \text{LR} = \frac{p(E \mid H_p)}{p(E \mid H_d)} \]

• Quantitative measurements, statistical models, databases representative of the relevant population

• Testing of validity and reliability under conditions reflecting those of the case
• **Fulop & Disner (2007, 2009):**
  – Pruned T-F-reassigned spectrograms of short vowel segments ([æ], [a] etc.)
  – visual comparison of spectrograms by human experts (“voiceprint”)

• **Fulop & Kim (2013):**
  • Quantitative approach
  • Automatic SVM-based closed-set identification
  • 24 enrolled speakers, 6 test segments
TF reassigned spectrograms

- Short-time Fourier transform of /iau/
- Channelized Instantaneous Frequency (CIF)
- Local Group Delay (LGD)

\[
\text{CIF}(\omega, T) = \frac{\delta}{\delta T} \arg(X_h(\omega, T))
\]

\[
\text{LGD}(\omega, T) = \frac{\delta}{\delta \omega} \arg(X_h(\omega, T))
\]

- “Reassign” T-F magnitudes to locations corresponding to local center of gravity
• Pruning (threshold) to reduce noise/artefacts
  - Based on second-order mixed partial derivative (Nelson, 2001)
  - Set to retain line components and impulses
• Fulop & Kim (2013): Feature representation based on discretization using a coarse grid
  – 50 time bins
  – 85 frequency bins

• Dimensionality reduction via PCA
  – 10 time features
  – 20 frequency features
• Chinese /iau/ triphthong:
  – Significant correlation over time and frequency
  – 2D Discrete cosine transform (DCT)

lower-order 7 x 7 coefficients
Feature representation – MFCC-on-/iau/

- Mel frequency cepstral coefficients (MFCC)
  - Common feature in FVC / speaker recognition
  - Extracted from /iau/ triphthong tokens
  - 16 MFCC + 16 Delta ($\Delta$) coefficients
Likelihood ratio calculation

- Score obtained using Gaussian mixture model-Universal background model (GMM-UBM) approach

\[ \lambda = (p_i, \mu_i, \Sigma_i)_{i=1,\ldots,M} \]

\[ s = \frac{1}{N} \sum_{j=1}^{N} \log \left( \frac{p(x_j | \lambda_{\text{suspect}})}{p(x_j | \lambda_{\text{UBM}})} \right) \]

- Logistic regression calibration and fusion

- Baseline automatic FVC system
  - Entire speech-active portion of recording
  - 16 MFCC + 16 delta ($\Delta$) coefficients
  - 1024 Gaussian mixture components (UBM)
• 60 female Standard Chinese speakers
• Split into 3 groups of 20 speakers
  • background set
  • development set
  • test set
• Manually marked /iau/ triphthongs
• Information-exchange task over telephone
• High quality and mobile-to-landline data
• Two recording sessions separated by 2–3 weeks

http://databases.forensic-voice-comparison.net/
Evaluation

• Validity / Accuracy
  • log-likelihood ratio cost \((C_{llr})\) metric

• Reliability / Precision
  • 95% credible interval (Morrison, 2011)

• Conditions:
  • High-quality v high-quality
  • Mobile-to-landline v mobile-to-landline
  • High-quality v mobile-to-landline
Results – high-quality v high-quality

The image shows a scatter plot with the x-axis labeled as "log10 95% Credible Interval (orders of magnitude)" and the y-axis labeled as "log likelihood ratio cost (C_{lr})". The plot includes data points for different methods such as TFR AVG, TFR DCT, MFCC, and Baseline. The legend indicates these methods with distinct marker shapes and colors.
Results – mobile-to-landline v mobile-to-land
Results – mobile-to-landline v high-quality
Tippett plot – Baseline system
Tippett plot – Fusion Baseline + TFR DCT
Conclusion

• High-quality v high-quality
  – no substantial improvement

• Mobile v mobile, mobile v high-quality
  – Improvement in validity, reliability deteriorates
  – TFR DCT improves upon TFR AVG
  – MFCC-on-/iau/ similar or slightly better

• Caveat:
  – Results give only an indication of performance
    (not tested: background noise, reverberation, ..)
  – Testing on per-case basis
Thank You!!


