A Generalised Normalisation Method for Speaker Verification

Dat Tran and Michael Wagner
University of Canberra, School of Computing
ACT 2601, Australia
{DatT, MichaelW}@ise.canberra.edu.au
Contents

- Introduction
- Current Normalisation Methods
- Proposed Method
  - A solution for the false rejection problem
  - A solution for the false acceptance problem
- Experimental Results
- Conclusion
Introduction

- Input Utterance
- Speaker Model
- Claimed Identity
- Score
- Threshold
- Accept or Reject
Score

Claimed Speaker’s Likelihood

Cohort Speakers’ Likelihood

Equal weight assumption

relativity of ratio-based values

False Acceptance Error

False Rejection Error

Proposed Method
Current Methods

\( \lambda_0 \) : claimed speaker model
\( \lambda \) : impostor speakers model
\( \lambda_i \) : cohort speaker models, \( i = 1, \ldots, B \)
\( X \) : input utterance
\( P(X | .) \) : likelihood function
\( S(X) \) : claimed speaker’s score
\[ S_1(X) = \log P(X \mid \lambda_0) - \log P(X \mid \lambda) \]

Likelihood Ratio

\[ S_2(X) = \log P(X \mid \lambda_0) - \max_{\lambda \neq \lambda_0} \log P(X \mid \lambda) \]

Higgins et al

\[ S_3(X) = \log P(X \mid \lambda_0) - \log \left\{ \frac{1}{B} \sum_{i=1}^{B} P(X \mid \lambda_i) \right\} \]

Reynolds
\[ S_4(X) = \log P(X \mid \lambda_0) - \log \sum_{i=0}^{B} P(X \mid \lambda_i) \]

Matsui & Furui

\[ S_5(X) = \log P(X \mid \lambda_0) - \frac{1}{B} \sum_{i=1}^{B} \log P(X \mid \lambda_i) \]

Liu et al

\[ S_6(X) = \sum_{t=1}^{T} \left[ \log P(x_t \mid \lambda_0) - \log \sum_{i=1}^{B} P(x_t \mid \lambda_i) \right] \]

Markov & Nakagawa
The Proposed Method

- A resolution for some false acceptances:

\[ S(X) = \frac{0.07}{0.03} \quad \text{and} \quad S(Y) = \frac{0.00000007}{0.00000003} \]

\[ \Rightarrow S(X) = S(Y) : \] the relativity of ratio-based values
But \( S_{nc}(X) = \frac{0.07}{0.03 + 0.01} = 1.75 \)

and \( S_{nc}(Y) = \frac{0.0000007}{0.0000003 + 0.01} = 0.000069 \)

\[ S_{nc}(X) > S_{nc}(Y) \]

Claimed Speaker’s Likelihood

Score \( \varepsilon = \frac{\text{Cohort Speakers’ Likelihood}}{\text{Cohort Speakers’ Likelihood} + \varepsilon} \)
A resolution for some **false rejections**:

False rejections can arise because the likelihood values of the cohort speakers are assumed to be of equal weight.

- Different weights based on fuzzy integration were proposed.
- Likelihood transformation.
Likelihood Transformation:

\[
\text{Score} = \frac{\text{Claimed Speaker's Likelihood}}{\text{Cohort Speakers' Likelihood}}
\]

\[
\text{Score}_T = \frac{T(\text{Claimed Speaker's Likelihood})}{T(\text{Cohort Speakers' Likelihood})}
\]

\[T(P): \text{a nonlinear function of } P\]
$P_0$ : the claimed speaker’s likelihood

$P_i$ : the $i$-th cohort speaker’s likelihood

**false rejection**

**true acceptance**

assuming thresholds $\theta = \theta_T = 1$
\[ S_{7T}(X) = \frac{\left( P(X | \lambda_0) \right)^\alpha}{\sum_{i=0}^{B} \left( P(X | \lambda_i) \right)^\alpha} \]

\[ S_{8T}(X) = \frac{-\log P(X | \lambda_0)^{-\beta}}{\sum_{i=0}^{B} -\log P(X | \lambda_i)^{-\beta}} \]

\( \alpha, \beta > 0 \)
Generalised Score:

\[
\text{Score} = \frac{\text{Claimed Speaker's Likelihood}}{\text{Cohort Speakers' Likelihood}}
\]

\[
\text{Score}_{T\varepsilon} = \frac{T(\text{Claimed Speaker’s Likelihood})}{T(\text{Cohort Speakers’ Likelihood} + \varepsilon)}
\]
\[ S_{7T \varepsilon}(X) = \sum_{i=0}^{B} \left[ P(X | \lambda_i) \right]^\alpha \left[ P(X | \lambda_0) \right]^\alpha + \varepsilon^\alpha \]

\[ S_{8T \varepsilon}(X) = \sum_{i=0}^{B} \left[ -\log P(X | \lambda_i) \right]^{-\beta} + \left[ -\log \varepsilon \right]^{-\beta} \]
Experimental Results

- **ANDOSL database:**
  - 108 speakers: 54 males, 54 females
  - Each speaker: 10 training & 190 test utterances
  - **Training:** GMMs using 16 & 32 mixtures
  - **Verification:** 2,093,040 utterances (claimed test + impostors test) for each result
  - Two cohort speakers sets:
    - 5 closest speakers and
    - 5 same-group speakers
EER(%) of S3 vs S3nc on Andosl

EER(%) of S4 vs S4nc on Andosl

EER(%) of S5 vs S5nc on Andosl

EER(%) of S6 vs S6nc on Andosl
**EER(%) of S4 vs S7T vs S7Tnc on Andosl**

- GMM16bst
- GMM32bst
- GMM16grp
- GMM32grp

**EER(%) of S4 vs S8T vs S8Tnc on Andosl**

- GMM16bst
- GMM32bst
- GMM16grp
- GMM32grp
• **YOHO database:**
  - 138 speakers: 108 males, 30 females
  - Each speaker: 48 training & 40 test utterances

  - Training:
    - **GMMs** using 16, 32 & 64 mixtures and
    - **VQ** models using 16, 32 & 64 codevectors
  - **Verification**: 728,640 utterances (claimed test + impostors test) for each result

  - Cohort speakers set: 5 closest speakers
  - Text-independent mode
Conclusion

- Based on considering the false acceptances and the false rejections, a generalised normalisation method has been proposed and experimentally evaluated.
- **Experiments were performed on the ANDOSL and YOHO copora**
- Experiments showed better results for the proposed method.