Noise Robustness in Forensic Speaker Verification

Yosef A. Solewicz

Bar-Ilan University
Israel National Police
Objective:

Evidence → Suspect’s model → score → Identity statement
Problem:

Evidence

Suspect’s model

Noise

Identity statement

Error

Identity statement

Score

Score
Goal: keep suspect’s identity statement constrained, even in noise

Must keep individual errors low and not only general **EER**: maximize clean/noise scores correlation
Decrease error:

- **Signal level** - spectral subtraction
- **Statistical level** - combine multiple (independent) observations

Partition of feature space

Multiple feature sets
Partition of feature space

Time / Frequency

Low correlation
(Efficient combination)

Multiple feature sets

Cep / Lsf / Δ’s

High correlation
(Inefficient combination)
Architecture for segmentation

Freq.

Time

Architecture for segmentation
Time/frequency segmentation

16 ERB

Data partition:

\[ 3 \text{ modulation spectrum bands:} \]
\[ 0.125-2; \ 2-8; \ 0.125-8 \text{ Hz (rasta)} \]
\[ (5:1 \text{ downsample}) \]

Sub-classifiers:

12 cep. coeff.
16 mixture GMM’s

Recombination:

\[ f = \text{mean, median, max} \]
\[ g = \text{mean, median, perceptrons} \]

Baseline:

Wideband/ 12 mel-cep
Experiments:

38 speakers (OGI)

Training: 1:30 min

Testing: 5 (10 sec.) utterances /speaker

Clean (original)

Noise: SNR=10±5 dB
Some results:

mean error $\not\leftrightarrow$ individual errors
(bias/variance dilemma)

Wideband/ overall/ low modulation spectrum:

(lowest individual EER / low correlation)

EER=15.5% (36.6% in noise)
Correlation coefficient (clean/noise)=0.35

Sub-band/ time-segmented:

(highest clean/noise correlation coefficient / high EER)

EER=19.2% (39.2% in noise) (29%)
Correlation coefficient (clean/noise)=0.61 (0.7)
Error management (bias/variance dilemma)

Model Complexity relative to effective training Data (data overfitting)
Segmentation gain over baseline (%):

<table>
<thead>
<tr>
<th></th>
<th>EER (clean data =&gt; noisy data)</th>
<th>corr. coeff. (clean/noisy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>20 - 10</td>
<td>10</td>
</tr>
<tr>
<td>Freq</td>
<td>5 - 10</td>
<td>40</td>
</tr>
<tr>
<td>T+F</td>
<td>10 - 0(25)</td>
<td>50(75)</td>
</tr>
</tbody>
</table>
Real cases:

Effects of using different amounts of spectral subtraction before verification
Improving modeling

- individual classifiers
- combined classifier

Error management (bias/variance dilemma)

Variance (low correlation)

Model Complexity relative to effective training Data
(data overfitting)
Danger of overfitting: (high variance)

Individual classifiers (Time/freq. Information balance)

Low modulation spectrum (low informative temporal data) :

Wideband / overall (highly informative freq. data): best EER=15.5\% (20.3)

Sub-band / segmented (low informative freq. data): worst EER=26.6\% (43.8)

Worst modulation spectrum band for correlation in baseline \sim 0.3
Danger of overfitting: (high variance)

In particular...

Narrow sub-bands (low informative freq. data) : high variance
We would like to prune optimally!

Individual classifiers:

Discarding lower sub-bands > using all sub-bands

Combined classifier :

max > median > mean
(perceptrons)
Current directions:

1- Optimize subclassifiers modeling.

2- Use machine learning methods to optimally recombine subclassifiers:

- Promising learning technique - ‘Winnow’ (multiplicative updating rule, prunes irrelevant subclassifiers)

- Learning to maximize clean/noisy correlations
Training recombination: maximize correlation

Suspect’s training data

Evidence/test

Extract noise sample

model

Anti-model

Recombination parameters

Train recombination: maximize correlation