A Study on Separation between Acoustic Class Models and Its Applications

Yu Tsao
04/28/2005
Outline

- Characterization of model separation
  - Defining target and competing sets
  - Computing target and competing scores (GLLR)
  - Preparing competing score histograms
  - GLLR plot as a tool to estimate model separations

- Applications of model separations
  - Model separation and acoustic discrimination
  - Model separation and acoustic mismatch
  - Model separation and training criteria
  - Model separation and acoustic resolution
  - Model separation and speech parameter selection

- Advantages of GLLR measure
Detention-based speech recognition paradigm

- Information about separations between models helps to:
  1. Develop new detector
  2. Measure confidence score
Separation between models

- Separation between phones “n” and “en” is smaller than that between phones “n” and “ch”
- Deterministic distance measurement can be used to estimate the separation
- Knowledge about the underlying speech data distribution is incomplete
- Probabilistic distance is more applicable => GLLR measure

Separation of the target phone “n” between the imposter phones “en” and “ch”
A Probabilistic distance : LLR

**LLR (Log Likelihood Ratio):**

\[ T(X | \lambda_0, \lambda_1) = \log[\ell(X | \lambda_0)] - \log[\ell(X | \lambda_1)] \]

Null hypothesis: \( H_0 \ (X \in S_0) \)

Alternative hypothesis: \( H_1 \ (X \notin S_0) \)

- A most competitive imposter
  \[ \lambda_1 = \arg \max_{r \in C_q} \log[\ell(X | \lambda_r)] \]
- Universal background model
  \[ \lambda_1 = \lambda_{UBM} \]
- Cohort set
A Probabilistic distance: GLLR

- Examples for Cohort sets to target phone models

<table>
<thead>
<tr>
<th>Target</th>
<th>Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>“w”</td>
<td>“l”, “el”, “ao”, “ow”, “uw”</td>
</tr>
<tr>
<td>“ah”</td>
<td>“aa”, “ax”, “ow”, “eh”, “aw”</td>
</tr>
<tr>
<td>“n”</td>
<td>“en”, “m”, “ng”, “ix”, “l”</td>
</tr>
</tbody>
</table>

(“w-ah+n” is the phoneme sequence for the word “one”)

- Distance measure - GLLR (Generalized Log Likelihood Ratio):

\[
T(X | \lambda_q, \bar{\Lambda}_q) = \log[\ell(X | \lambda_q)] - \log[f(X | \bar{\Lambda}_q)]
\]

\[
f(X | \bar{\Lambda}_q) = \left\{ \left| C_q \right|^{-1} \sum_r \exp[\eta \log \ell(X | \lambda_r)] \right\}^{\frac{1}{\eta}}
\]

(\left| C_q \right| is the size of the cohort set \( C_q \) of the claimed target \( q \) )
GLLR Plot

- A GLLR plot consists of two GLLR histograms:

![GLLR Plot Diagram]

GLLR measure:

\[ T(X | \lambda_q, \Lambda_q) = \log[ \ell(X | \lambda_q)] - \log[ f(X | \Lambda_q)] \]

\[ f(X | \Lambda_q) = \{|C_q|^{-1} \sum_r \exp[\eta \log \ell(X | \lambda_r)]\}^{1/\eta} \]

- GLLR plots is a useful tool to visually analyze:
  Type I: target samples missed
  Type II: false alarm
  Separations between models
Applications of model separations

- Experimental set up characterization

- GLLLR measure for model separation analysis

1. Model separation and acoustic discrimination
2. Model separation and acoustic mismatch
3. Model separation and training criteria
4. Model separation and acoustic resolution
5. Model separation and speech parameter selection
## Experimental environment

### Databases characterization

<table>
<thead>
<tr>
<th>Database</th>
<th>TIMIT DB</th>
<th>NTIMIT DB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data description</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Clean environment</td>
<td>1. Noisy environment</td>
<td></td>
</tr>
<tr>
<td>2. 16K Hz sampling</td>
<td>2. 8K Hz sampling</td>
<td></td>
</tr>
<tr>
<td>3. Microphone speech</td>
<td>3. Telephone speech</td>
<td></td>
</tr>
</tbody>
</table>

### Model sets characterization

<table>
<thead>
<tr>
<th>Model set</th>
<th>Monophone model</th>
<th>Triphone model</th>
<th>Manner class model</th>
<th>Tri-manner model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model description</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total no :44 + sil.</td>
<td>Total no :4331</td>
<td>Total no :5 + sil.</td>
<td>Total no :196</td>
<td></td>
</tr>
<tr>
<td>3 states per model</td>
<td>3 states per model</td>
<td>3 states per model</td>
<td>3 states per model</td>
<td></td>
</tr>
<tr>
<td>8 mixtures per state</td>
<td>8 mixtures per state</td>
<td>8 mixtures per state</td>
<td>8 mixtures per state</td>
<td></td>
</tr>
</tbody>
</table>
Applications of model separations

- Experimental set up characterization
- GLLLR measure for model separation analysis

-1. Model separation and acoustic discrimination
-2. Model separation and acoustic mismatch
-3. Model separation and training criteria
-4. Model separation and acoustic resolution
-5. Model separation and speech parameter selection
Model separation and acoustic discrimination

Evaluate the confusability of word in an ASR vocabulary

Target phone: ix

Target phone: ay
Model separation and acoustic mismatch

- Mismatched training and testing environments

<table>
<thead>
<tr>
<th>Overall accuracy (%)</th>
<th>matched</th>
<th>mismatched</th>
</tr>
</thead>
<tbody>
<tr>
<td>58.29 %</td>
<td>24.56 %</td>
<td></td>
</tr>
</tbody>
</table>

More serious degradation for unvoiced sounds
Model separation and training criteria

<table>
<thead>
<tr>
<th>Overall Accuracy (%)</th>
<th>Without MCE Training (phone)</th>
<th>With MCE Training (phone)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58.29 %</td>
<td>60.88 %</td>
</tr>
<tr>
<td>Overall Accuracy (%)</td>
<td>71.09 %</td>
<td>78.32 %</td>
</tr>
</tbody>
</table>

Target phone: ay
Before MCE training

Target phone: ay
After MCE training

Target class: Vowel
Before MCE training

Target class: Vowel
After MCE training
Model separation and acoustic resolution

CD model has better acoustic resolution

<table>
<thead>
<tr>
<th>Overall Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI Manner</td>
</tr>
<tr>
<td>71.09 %</td>
</tr>
</tbody>
</table>

GLLR : $T(X | λ_q,CI,\overline{λ}_q,CI) = \log[ \ell(X | λ_q,CI) ] - \log[ f(X | \overline{λ}_q,CI) ]$  
GLLR : $T(X | λ_q,CD,\overline{λ}_q,CI) = \log[ \ell(X | λ_q,CD) ] - \log[ f(X | \overline{λ}_q,CI) ]$
Model separation and speech parameter select.

- **Speech model separation and parameter selection**
  - VOT (voice onset time) parameter vs. 39 MFCC parameters


- **Speaker model separation and parameter selection**
  - Pitch parameter vs. 39 MFCC parameters

  Ma, C. and Lee, C.-H., "Speaker Verification Based on Combining Speaker Parameter Selection and Decisions," 
  *submitted to InterSpeech 2005.*
Advantages of GLLR measure

- Measure the distance between models
- Help to visually analyzes the separation between models
- Serve as a good tool to develop:
  1. improved speech models
  2. better training/compensation algorithms
  3. new speech parameters
- Need not to conduct large scale experiments

Thank you