Hiromitsu Nishizaki, Yuto Furuya, Satoshi Natori and Yoshihiro Sekiguchi,
University of Yamanashi, Japan

# Summary

Future works

10 speech recognizers

5 types of Language Models
- Word based trigram : WBC
- Hiragana based trigram : WBH
- Syllable based trigram : CB
- Bi-syllable based trigram : BM
- Nothing : Non

2 types of Acoustic Models
- Syllable based HMM : Syl
- Tri-phone HMM : Tri

LVCSR decoder
- Julius rev.4.1.3

Outputs of 10 recognition systems (all outputs are converted into phoneme sequence)

<table>
<thead>
<tr>
<th>System</th>
<th>Phoneme Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WBC/Tri</td>
<td>k o s @ a @ @ i @</td>
</tr>
<tr>
<td>WBH/Tri</td>
<td>q o s u a @ @ a N</td>
</tr>
<tr>
<td>CB/Tri</td>
<td>k o s @ a m a i @</td>
</tr>
<tr>
<td>CSB/Tri</td>
<td>k o s @ a @ @ N</td>
</tr>
<tr>
<td>BM/Tri</td>
<td>b o s @ @ @ a @</td>
</tr>
<tr>
<td>BM/Syl</td>
<td>b o s @ @ @ a @</td>
</tr>
<tr>
<td>CSB/Syl</td>
<td>@ @ s @ a @ @ N</td>
</tr>
<tr>
<td>Non/Syl</td>
<td>@ @ s @ a @ @ N</td>
</tr>
</tbody>
</table>

Arc
Node
Terminal Node

Distance: 0.3

Baseline STD is performed by the DP without NULL transition on the transcription of “CB/Tri.”
Baseline STD is performed by the simple DP on the transcription of “CB/Tri.”

The maximum F-measure of “PTN” is 71.4%
The maximum F-measure of “Baseline” is 55.6%

4. False detection control in DP framework

- Introduction of the false detection control parameters
  - “Voting”: the number of recognizers outputting the same phoneme on the same arc
  - “ArcWidth”: the number of arcs between successive two nodes

- The parameters are installed to the calculation of DP cost

5. Conclusion

Summary

Future works

- Using multiple speech recognizers for STD
- Multiple recognizers make STD performance better
- Integrating multiple recognizers’ output in to PTN was very powerful to improve the performance