Japanese Dictation Toolkit
– Free Software Repository for Speech Recognition –

Tatsuya Kawahara (Kyoto Univ.), Tetsunori Kobayashi (Waseda Univ.), Kazuya Takeda (Nagoya Univ.), Nobuaki Minematsu (Toyoohashi Univ. of Tech.), Katsunobu Itou (ETL), Mikio Yamamoto (Tsukuba Univ.), Atsushi Yamada (ASTEM), Takehito Utsuro, Kiyohiro Shikano (Nara Inst. of Sci. & Tech.)

http://winnie.kuis.kyoto-u.ac.jp/pub/julius/doc/00readme.html

1. Overview

- A sharable software repository
- Project sponsored by a governmental organization IPA (Information-technology Promotion Agency), Japan
- Collaboration of researchers of different academic institutes in Japan
- Available to public freely

2. Specification of Models and Programs

2.1 Acoustic Model

Continuous density HMM (HTK format) trained with ASJ (Acoustical Society of Japan) databases

<table>
<thead>
<tr>
<th>model</th>
<th>#states</th>
<th>#mixtures</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>monophone</td>
<td>129</td>
<td>4, 8, 16</td>
<td>GD, GI</td>
</tr>
<tr>
<td>triphone 1000</td>
<td>1000</td>
<td>4, 8, 16</td>
<td>GD</td>
</tr>
<tr>
<td>triphone 2000</td>
<td>2000</td>
<td>4, 8, 16</td>
<td>GD, GI</td>
</tr>
<tr>
<td>triphone 3000</td>
<td>3000</td>
<td>4, 8, 16</td>
<td>GD</td>
</tr>
</tbody>
</table>

2.2 Japanese Text Processor

- Morphological Analyzer: ChaSen
- Pronunciation Post-Processor: ChaWan to handle undelimited Kanji-Kana texts

2.3 Lexicon and Language Model

<table>
<thead>
<tr>
<th>vocabulary size</th>
<th>coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>88.2%</td>
</tr>
<tr>
<td>20000</td>
<td>96.5%</td>
</tr>
</tbody>
</table>

Word 2-gram, 3-gram (ARPA format) trained with Mainichi newspaper articles of 7 years

2.4 Decoder: Julius

<table>
<thead>
<tr>
<th></th>
<th>acoustic model</th>
<th>language model</th>
<th>search approx</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st pass</td>
<td>intra-word CD</td>
<td>2-gram</td>
<td>1-best</td>
</tr>
<tr>
<td>2nd pass</td>
<td>inter-word CD</td>
<td>3-gram</td>
<td>N-best</td>
</tr>
</tbody>
</table>

3. Key-Property of Toolkit

- open format & interface
- easy to revise and replace modules
- reasonable performance as the baseline word accuracy 85~93% / 1~8 real time
- current 1998 version
- 20K model & system
- Unix platform
- condition
  Copyright reserved
  NO WARRANTY

REFERENCES


For further information:
http://www.lang.astem.or.jp/dictation-tk/
mailto: dictation-tk-request@astem.or.jp

Copyright reserved
NO WARRANTY

ISCA Archive
http://www.isca-speech.org/archive

ESCA Workshop on Interactive Dialogue in Multi-Modal Systems (IDS-99)
Kloster Irsee, Germany
June 22-25, 1999