THISL progress report - 1999sep

Dan Ellis
International Computer Science Institute, Berkeley CA
<dpwe@icsi.berkeley.edu>

Outline

1 Spoken queries recognized
2 TREC modspec processed
3 ThisIGUI updated
4 FYI: General audio retrieval
Spoken query recognition

- 23 TREC query sentences (351 words)
- Pronounced by 9 speakers (from BBC) + 1
- Recognized by RNN and RNN+MSG

<table>
<thead>
<tr>
<th>Speaker</th>
<th>RNN WER</th>
<th>RNN+MSG WER</th>
</tr>
</thead>
<tbody>
<tr>
<td>nch</td>
<td>16.0%</td>
<td>15.1%</td>
</tr>
<tr>
<td>mrm</td>
<td>21.4%</td>
<td>21.9%</td>
</tr>
<tr>
<td>cdm</td>
<td>24.5%</td>
<td>21.1%</td>
</tr>
<tr>
<td>jp</td>
<td>24.5%</td>
<td>24.8%</td>
</tr>
<tr>
<td>mac</td>
<td>25.9%</td>
<td>23.9%</td>
</tr>
<tr>
<td>mpj</td>
<td>26.5%</td>
<td>24.2%</td>
</tr>
<tr>
<td>ljc</td>
<td>26.5%</td>
<td>25.6%</td>
</tr>
<tr>
<td>cmb</td>
<td>35.0%</td>
<td>29.3%</td>
</tr>
<tr>
<td>cp</td>
<td>38.7%</td>
<td>34.2%</td>
</tr>
<tr>
<td>dpwe</td>
<td>36.2%</td>
<td>37.9%</td>
</tr>
<tr>
<td>Avg</td>
<td>27.5%</td>
<td>25.8%</td>
</tr>
</tbody>
</table>

- Pass to IR for evaluation ...
TREC modspeck processing

- Modulation-filtered spectrogram (msg) features
  - helped in BN’98 eval
  - could help for TREC8 data?
  - run msg neural net for probability-combination with RNN

- Automatic wav→msg→Ina loop
  - ftp between Sheffield and ICSI
  - 2.5M weight network ran in 0.3xRT on TetraSPERT
  - 900+ files (550 hours) processed in 4 weeks
ThisILGUI updated

Date filters
Program filter
Speech input
thisILIR-0.2 & thisILIR-0.5 compatible
BBC data
Pauses & sentence breaks shown
click-to-play
Whole-show transcripts
4 General audio mixture retrieval

- Information retrieval for nonspeech audio
  - non-word events (laughter, shouting)
  - sound effects on soundtracks (engines, cars)
  - ambiances

- Approach: auditory elements, learned patterns:

- Just-submitted proposal to NSF