Using ASR in a Reading Tutor:
Speech Assessment and Remedial Tools

SPACE symposium — February 5–6, 2009

Jacques Duchateau (K.U.Leuven), Jean-Pierre Martens (U.Gent)
Overview of the talk

- Description of the recognition system
  - system overview
  - acoustic modelling
  - modelling reading errors

- ASR in a reading tutor
  - diagnostic tool: assessment of child
  - remedial tool: tracking of the child
  - remedial tool: feedback on reading errors
System overview: 2-pass recogniser

- generic phoneme recogniser
  - acoustic models
  - phoneme N−gram

  phoneme lattice

  lattice search module
  - finite state transducer (FST)

  word sentence/lattice

  post processing
  - confidence measures

word FSTs

TAFEL:

garbage FST

GARB: phoneme N−gram converted to FST

sentence FST
Acoustic modelling

- Baseline acoustic model
  - acoustic model for children (5-12 year), trained on data different from CHOREC
  - adaptive VTLN in preprocessing: avoids latency

- Improving the baseline model
  - long and short vowels: phoneme duration
  - plosives: Voice Onset Time
  - phoneme dependent frame selection preference
  - articulatory features
Modelling reading errors

Objective
- model predictable mispronunciations of a target word/sentence
- accommodate unpredictable pronunciations

Sentence FST (for sentence reading tasks)
- skipping of words, repeating part of sentence
- unexpected words (garbage)
- inserted silences

Word FST (for all reading tasks)
- word restarts, skipped phonemes
- other predictable mispronunciations
- spelled words (different spelling modes)
Diagnostic tool: assessment of child

- Assessment based on one 40 word task
- Children of 1 grade classified in 5 performance groups
- Ranking by score: time per correctly read word
- Human-machine equals human-human agreement

<table>
<thead>
<tr>
<th></th>
<th>1-syllable</th>
<th>2-syllable</th>
<th>3+4-syll.</th>
</tr>
</thead>
<tbody>
<tr>
<td>grade 1</td>
<td>0.84/80%</td>
<td>0.70/63%</td>
<td>- / -</td>
</tr>
<tr>
<td>grade 2</td>
<td>0.86/80%</td>
<td>0.94/91%</td>
<td>0.85/79%</td>
</tr>
<tr>
<td>grade 3</td>
<td>0.93/90%</td>
<td>0.85/82%</td>
<td>0.89/86%</td>
</tr>
<tr>
<td>grade 4</td>
<td>0.86/81%</td>
<td>0.93/90%</td>
<td>0.92/88%</td>
</tr>
</tbody>
</table>
Remedial tool: tracking of the child

- **How to track**
  - most likely FST state from search every 150ms
  - preprocessing: low latency
  - 1-pass recognition system

- **Use of the tracker**
  - advance to next screen (no keyboard)
  - highlight word to be read (remedy)
  - feedback by synthesis (remedy)
Remedial tool: feedback on reading errors (1)

Miscue detection rate *versus* False alarm rate

![ROC curve - Real words (7.6% child error rate)](image1)

![ROC curve - Pseudo words (26.0% child error rate)](image2)
Remedial tool: feedback on reading errors (2)

Miscue detection rate *versus* Feedback error rate

![ROC curve – Real words (7.6% child error rate)]

![ROC curve – Pseudo words (26.0% child error rate)]
## Remedial tool: feedback on reading errors (3)

School, grade and task specific results

<table>
<thead>
<tr>
<th></th>
<th>child ER</th>
<th>MD rate</th>
<th>FA rate</th>
<th>FB ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01-S03, 2-syll Real</td>
<td>7.6%</td>
<td>42.4%</td>
<td>3.0%</td>
<td>46.1%</td>
</tr>
<tr>
<td>S01-S03, 2-syll Pseudo</td>
<td>26.0%</td>
<td>43.0%</td>
<td>8.0%</td>
<td>34.7%</td>
</tr>
<tr>
<td>S01-S03, grade 1, 1-syll R</td>
<td>27.3%</td>
<td>46.3%</td>
<td>7.3%</td>
<td>29.4%</td>
</tr>
<tr>
<td>S01-S03, grade 1, 3+4-syll P</td>
<td>59.4%</td>
<td>82.1%</td>
<td>29.2%</td>
<td>19.6%</td>
</tr>
<tr>
<td>S01-S03, grade 4, 1-syll R</td>
<td>1.5%</td>
<td>6.9%</td>
<td>1.2%</td>
<td>92.0%</td>
</tr>
<tr>
<td>S01-S03, grade 4, 3+4-syll P</td>
<td>35.4%</td>
<td>54.7%</td>
<td>12.4%</td>
<td>29.3%</td>
</tr>
<tr>
<td>S05-S07, 2-syll Real</td>
<td>14.7%</td>
<td>50.1%</td>
<td>6.5%</td>
<td>43.1%</td>
</tr>
<tr>
<td>S05-S07, 2-syll Pseudo</td>
<td>46.5%</td>
<td>51.9%</td>
<td>16.8%</td>
<td>27.1%</td>
</tr>
<tr>
<td>S05, grade 2, 1-syll Real</td>
<td>23.2%</td>
<td>52.3%</td>
<td>13.0%</td>
<td>45.2%</td>
</tr>
<tr>
<td>S05, grade 2, 3+4-syll Pseudo</td>
<td>91.3%</td>
<td>84.9%</td>
<td>28.6%</td>
<td>3.1%</td>
</tr>
</tbody>
</table>
Conclusions

- Detailed reading error analysis is hard
  - improved classification of specific phoneme classes
  - improved reading error model

- Still ASR is useful in diagnostic and remedial tools
  - diagnostic: reading level classification
  - remedial: tracker for focus, for help using synthesis
  - remedial: repeat difficult words