Automatic assessment and error detection of non-native English speech using phone distance features

Konstantinos Kyriakopoulos, Kate Knill, Mark Gales
ALTA Institute / Department of Engineering, University of Cambridge

Introduction

Millions are learning English worldwide and millions yearly take tests

- **Automatic assessment** assigns grades to candidates
- **Error detection** identifies and interprets localised mistakes
- **Feedback** results to user to improve their pronunciation

ALTA system works with unstructured, spontaneous speech

Assessment Performance

- Train and test on recorded answers to BULATS speaking test
- Unstructured, spontaneous, unlabelled speech

Pearson correlations of grader output with expert grades:

<table>
<thead>
<tr>
<th>Data sources</th>
<th>Baseline features</th>
<th>Baseline + pronunciation features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only Gujarati speakers</td>
<td>0.816</td>
<td>0.872</td>
</tr>
</tbody>
</table>

Accent Evaluation and Error Detection

GPFs to relate score to distance of each phone to all others.

Phones that best predict score for different L1s:

<table>
<thead>
<tr>
<th>L1s</th>
<th>Predictive power (Pearson correlations) of top ten phones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>ɪ (0.638), ɲ(0.604), u (0.589), h (0.587), a (0.580), æ (0.558), j (0.547), æ (0.532), tʃ (0.505), b (0.502)</td>
</tr>
<tr>
<td>Gujarati</td>
<td>θ (0.419), æ (0.417), k (0.409), p (0.402), ea (0.395), tʃ (0.379), w (0.377), ej (0.366), f (0.363), t (0.350)</td>
</tr>
<tr>
<td>French</td>
<td>a (0.585), j (0.500), n (0.442), a (0.442), i (0.442), r (0.442), æ (0.442), s (0.442), j (0.442)</td>
</tr>
<tr>
<td>Thai</td>
<td>d (0.724), æ (0.711), ʃ (0.701), b (0.685), k (0.674), ə (0.651), g (0.643), æ (0.632), æ (0.607), t (0.605)</td>
</tr>
</tbody>
</table>

- Can now score speakers on pronunciation of each phone
- Use to characterise accent relative to L1 and proficiency
- Identify problem phones for feedback to the speaker
- Use to distinguish accent errors from lexical errors
  
  e.g. for a Spanish speaker:
  
  yes: ɪes => ðæs  accent error
  
  subtle: sɒtl => sɒtl  lexical error
  
  for a French speaker:
  
  near: nə(ɹ) => nə  accent error
  
  grader: ɡə(ɹ) => guə(ɹ)  accent + lexical error

- Detect lexical errors with phone substitution and insertion models

Conclusion

- Phone distance features significantly increase grader performance over baseline audio and fluency features
- Performance is stronger for known L1
- Discriminating power is greater for lower scores
- Promising potential for use in error detection