Glottalization and Vowel Nasalization Detection

Tom Wang, Georgia Tech
16\textsuperscript{th} August 2004
WS04
Outline

- Vowel Nasalization Detection
- Glottalization Detection
Vowel Nasalization

• Nasalization in Vowel Production: coupling of oral and nasal cavities during vowel production (Beddor, 2003)

• Goal: Build classifier for nasalized vowels vs. un-nasalized vowels (e.g. ae, ow,… vs. ae_n, ow_n,…)
  – “you know”

• Motivation: may be a sign of deleted nasal phoneme enhance pronunciation model
Vowel Nasalization Detection

- Data set: Switchboard (WS96, WS97)
- SVM Training Features: MFCCs, Knowledge Based Acoustic Parameters (Bitar & Espy-Wilson, 1996), Formant (Zheng & Hasegawa-Johnson, 2004), Rate Scale (Mesgarani, et al. 2004)
- Features taken per 5 ms frame.
- Classify on per frame basis
- Results: Linear SVM
  62.95% accuracy
Vowel Nasalization Detection

- Divide problem into vowel-specific classifiers (i.e. ae vs. ae\_n)

- Test common classifier on vowel specific pairs

- Finding good training parameters
  Entire Regularization Path of SVMs (Hastie, et al. 2004)
  Choose range of values
## Summary of Results

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Specific – Default</th>
<th>Specific – Optimized C</th>
<th>Common</th>
<th># Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>ey vs. ey_n</td>
<td>80.73%</td>
<td>81.30%</td>
<td>80.92%</td>
<td>524</td>
</tr>
<tr>
<td>iy vs. iy_n</td>
<td>55.12%</td>
<td>58.11%</td>
<td>75.60%</td>
<td>1406</td>
</tr>
<tr>
<td>ae vs. ae_n</td>
<td>72.28%</td>
<td>74.51%</td>
<td>68.48%</td>
<td>2024</td>
</tr>
<tr>
<td>ao vs. ao_n</td>
<td>61.44%</td>
<td>63.40%</td>
<td>73.20%</td>
<td>612</td>
</tr>
<tr>
<td>ah vs. ah_n</td>
<td>64.20%</td>
<td>65.01%</td>
<td>68.73%</td>
<td>2712</td>
</tr>
<tr>
<td>ih vs. ih_n</td>
<td>56.80%</td>
<td>54.63%</td>
<td>62.36%</td>
<td>3826</td>
</tr>
<tr>
<td>eh vs. eh_n</td>
<td>57.61%</td>
<td>60.10%</td>
<td>58.73%</td>
<td>1604</td>
</tr>
<tr>
<td>aa vs. aa_n</td>
<td>60.09%</td>
<td>52.23%</td>
<td>55.84%</td>
<td>1388</td>
</tr>
<tr>
<td>ax vs. ax_n</td>
<td>56.56%</td>
<td>55.85%</td>
<td>56.38%</td>
<td>564</td>
</tr>
<tr>
<td>er vs. er_n</td>
<td>56.19%</td>
<td>47.52%</td>
<td>54.46%</td>
<td>404</td>
</tr>
<tr>
<td>ow vs. ow_n</td>
<td>49.09%</td>
<td>55.64%</td>
<td>54.61%</td>
<td>2408</td>
</tr>
<tr>
<td>ay vs. ay_n</td>
<td>47.88%</td>
<td>51.80%</td>
<td>54.77%</td>
<td>944</td>
</tr>
</tbody>
</table>

Outputs used with SVMs of other phonetic features in different pronunciation models.

Optimized C may have caused over-fitting (Cohen & Forman, 2004)
Outline

• Vowel Nasalization Detection

• Glottalization Detection
Glottalization Detection

- Glottalization – glottal stop or deviation from canonical normal voice
  - glottal stop - “absolutely”: transcribed with ‘q’, no ‘t’
  - creaky vowel – “shepherd”: transcribed with ‘er_cr’, no ‘d’
  - creak vowel – “er_cr”:
  - non-creaky vowel – “another”: transcribed with ‘er’
  - non-creak vowel – “er”:

- Motivation:
  enhance pronunciation model
  determine if/which phoneme deleted

Shattuck-Hufnagel & Redi, 2001
Glottalization Detection

• Approach:

(1) Extract q and cr tokens and generate acoustic features per frame
   Tokens – exclusively as allophone of /t/, /d/ (not word-initial)
(2) Train SVM to detect q, cr
(3) Re-adjust acoustic features
(4) Error analysis on current /t/, /d/, /p/, detectors
(5) Evaluate added detectors

• Data – Switchboard, Switchboard transcriptions
Acoustic Features of Glottalization

- Aperiodicity – irregularity in duration of glottal pulses (Shattuck-Hufnagel & Redi, 2001)
  Cepstral Peak Prominence (CPP) (Heman-Ackah, et al. 2001)

- Creaky voice - wide pitch period, low fundamental frequency, pitch period damping (Shattuck-Hufnagel & Redi, 2001)
  Autocorrelation estimation of pitch
  Cepstral pitch determination

- Relative amplitudes of H1, F1/H2 (Klatt & Klatt, 1990; Ladefoged, et al. 1984)
  Spectral slice

- Generate in MATLAB
SVM Training and Error Analysis

• Train classifier (SVM):
  q vs. rest of utterance
  vowel_cr vs. rest of utterance
  vowel_cr vs. vowel
  Parameter selection
  Add additional acoustic features (MFCCs, Knowledge Based Acoustic Parameters, etc.)

• Test current /t/, /d/, etc. detectors
  Add glottalization detectors to pronunciation model
  Evaluate
Summary

• Glottalization may be a sign of deleted sound (/t/, /d/)
  Detection can improve pronunciation model

• Experiments similar to vowel nasalization detection

• Automated acoustic feature extraction
  SVM training and testing
  Evaluating addition of detectors