Glottalization and Vowel Nasalization Detection

Tom Wang, Georgia Tech

16th 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

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

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