Accent Characterisation and Recognition Using Self-Normalisation

Mark Huckvale
University College London

Accents Research

Top-down Research

- Specify accent groups
- Accent group for a speaker is "known"
- Specify which acoustic features to measure
- Hypothesise which phonetic/phonological characteristics important

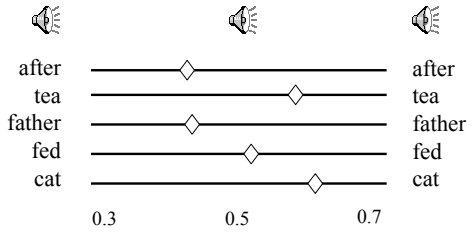
Bottom-up Research

- Accent groups emerge by clustering speakers
- Best accent group for a speaker emerges from distance to cluster centres
- Useful acoustic features discovered by comparing speakers
- Phonetic/phonological characteristics emerge by studying clusters

Bottom-up Accents Research

- Key problem in data-driven accent research is the availability of a good metric for comparing the similarity of two people's accents.
- Because any acoustic measures will be influenced by
  - vocal tract length, voice pitch, voice quality, speaking style and speaking rate
- as well as by
  - phonetic and phonological changes due to accent

Acoustic Comparison Problem



Experimental Data

- Accents of British Isles corpus
- (10 male+10 female) × 14 accent areas:
  - Birmingham | Cornwall | East Anglia | East Yorkshire | Glasgow | Inner London | Lancashire | Liverpool | Newcastle | North Wales | Dublin | Scottish Highlands | South East | Ulster
- Looking at 20 short sentences from each speaker
- Phonetically annotated by HMM forced alignment
- Generates about 130 vowel measurements from each of 270 speakers or about 35,000 data points

Formant Metric

- Divide each vowel into two halves
- Find median value of first four formants in each half
- Compare vowels in matching words across speakers
- Use Euclidean distance, average over all pairs
- Assess performance by finding accent group of nearest neighbour

Any sex	44.2%
Other sex	27.4%

Normalised Formant Metric

- Normalise formant values for each speaker independently
  - so each formant has zero mean and unit variance
- Compare vowels in matching words across speakers
- Use Euclidean distance, average over all pairs

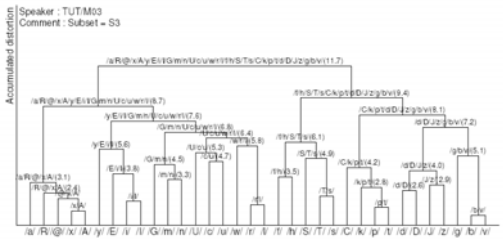
Any sex	50.4%
Other sex	34.7%

Spectral Metric

- 19-channel auditory filterbank
- Subtract mean energy, add as 20th value
- Take mean over first & second half of vowel
- Compare vowels in matching words across speakers
- Use Euclidean distance, average over all pairs

Any sex	71.2%
Other sex	43.4%

Minematsu Innovation



Analyse a speaker's accent by studying the acoustic distances between segment realisations.

Segment Distance Tables

- Compute distances between segment realisations for each speaker independently
- Then compare speaker distance tables not spectral properties of segment realisations
- "Self normalisation" accommodates speaker characteristics not due to accent

Birmingham

aa/after:	0.00	3.48	2.14
aa/father:	3.48	0.00	3.62
ae/cat:	2.14	3.62	0.00

South East

aa/after:	0.00	2.27	3.21
aa/father:	2.27	0.00	3.71
ae/cat:	3.21	3.71	0.00

Distance Table Metric

- Compute 130×130 vowel distance tables for each speaker
- Use Correlation distance on pairs of tables, between pairs of vowels from matching words
- Choose accent group of speaker with highest correlation

Any sex	80.3%
Other sex	75.9%

Accent Confusion Matrix

	brm	crn	ean	eyk	gla	ilo	lan	lvp	ncl	nwa	roi	shl	sse	uls	
brm:	17	1	0	0	0	0	1	0	0	0	0	0	1	0	= 85.0%
crn:	0	13	0	0	0	0	0	0	0	1	2	0	3	0	= 68.4%
ean:	0	2	8	0	0	2	0	0	0	0	0	0	7	0	= 42.1%
eyk:	0	1	0	11	0	0	0	0	0	6	0	0	2	0	= 55.0%
gla:	0	0	0	0	18	0	0	0	0	0	0	2	0	0	= 90.0%
ilo:	0	0	1	0	0	17	0	0	0	1	0	0	1	0	= 85.0%
lan:	0	0	0	0	0	0	19	0	0	1	0	0	0	0	= 95.0%
lvp:	0	0	0	0	0	0	0	17	1	2	0	0	0	0	= 85.0%
ncl:	0	0	0	0	0	0	1	0	19	0	0	0	0	0	= 95.0%
nwa:	0	1	0	0	0	1	0	2	1	15	0	0	0	0	= 75.0%
roi:	0	0	0	0	0	0	0	0	1	0	19	0	0	0	= 95.0%
shl:	0	0	0	0	0	0	0	0	0	1	0	19	0	0	= 95.0%
sse:	0	4	2	0	0	2	0	0	0	0	0	0	8	0	= 50.0%
uls:	0	0	0	0	0	0	0	0	0	0	0	0	0	20	= 100.0%

Most common confusions:
- East Anglia → South East
- E. Yorkshire → N. Wales

In summary, this study

- Introduced a new metric for comparing the similarity of accent of two speakers
- Calculated from an analysis of a set of sentences spoken by both speakers
- Based on the correlation of inter-segment distance tables
- Shows better accent recognition performance than a metric based on acoustic comparisons across speakers
- Is relatively unaffected by speaker characteristics
- Opens up further possibilities for bottom-up accent research