Distinguishing speakers using formant dynamics in read and spontaneous speech: a study of British English /uː/
Dynamic features of speech and speaker identification

- More information than static (e.g. vowel midpoint)
- Reflect movement of a person’s speech organs as well as dimensions
  - people move in individual ways for skilled motor activities - walking, running, … and speech
- can view speech as achievement of a series of linguistic ‘targets’
- speakers likely to exhibit similar properties at ‘targets’ (e.g. segment midpoints), but move between these in individual ways
  → examine formant frequency dynamics
Formant dynamics

- Looking beyond midpoint to formant contours for potentially speaker-distinguishing information...
Previous work

• Studies of formant dynamics → greater levels of speaker discrimination than static features
• Greisbach et al. 1995, Ingram et al. 1996, Rodman et al. 2002
• McDougall (2005) study of SSBE intervocalic /r/ sequences
Outstanding issues

• How to capture dynamic information most efficiently, e.g. characterise with polynomial or other equations using regression? (cf. McDougall 2005, 2006)
• Effect of different speaking contexts
• Effect of different linguistic conditions
• More sounds and sequences
• Larger populations
A study of Standard Southern British English /uː/

- Study of 20 speakers’ /uː/ formant dynamics (larger than previous studies)
- /uː/ undergoing change in SSBE (ref. earlier DyViS presentation, IAFPA ’07) – expect variability among speakers
- Observation of considerable between-speaker variation in formant dynamics when taking static measurements
Research Questions

• How well do the formant dynamics of /uː/ in SSBE discriminate speakers?
• Do the formant dynamics of /uː/ discriminate speakers in read and spontaneous speech? i.e. is speaker-specific information preserved across speaking conditions?
Subjects

• 20 speakers from DyViS database at Cambridge
• male
• Standard Southern British English
• aged 18-25
Data 1: read speech

• DyViS Task 4 (read sentences)
• 6 tokens of /uː/ per speaker
• nuclear-stressed in h_d context:

He hates contracting words, but he said a WHO’D today
Data 2: spontaneous speech

- DyViS Task 2
  (telephone call with ‘accomplice’)
- studio quality recording
- 6 tokens of /uː/ per speaker
- $C_{[+\text{stop}]}C_{[+\text{stop}]}$ contexts:
  Cooper, coot, Hooper, poodles, scooter, supervisor
Peter Beard
barber
visit steak house
together

Barbara Detman
hairdresser
keeps poodles
drives a scooter

Eugene Burke
barber
supervisor
lives in Dexter
play sports together

Extract from Subject 1
telephone call
Measurements

Token of /u:/ by S10 – read speech

Frequency (Hz)

+10% Step of /u:/

今天的
Results - read speech, e.g. 4 speakers - F2

- Frequency (Hz)

- 10% interval of /u:/

- S1
- S6
- S8
- S10

6 tokens per speaker
F1 and F2 contours approximated with polynomial equations to reduce the number of dimensions

**Quadratic**

\[ y = a_0 + a_1x + a_2x^2 \]

**Cubic**

\[ y = a_0 + a_1x + a_2x^2 + a_3x^3 \]

**Quartic**

\[ y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 \]
F2 of token 1, S1 - polynomial fittings

- actual values
- quadratic fit
- cubic fit
- quartic fit

Frequency (Hz)

1 2 3 4 5 6 7 8 9

900 1000 1100 1200 1300
Procedure

• fit F1 and F2 contours of each token with polynomial equations (with *Matlab*)
• test the reliability of the polynomial coefficients in distinguishing speakers using Discriminant Analysis
  → correct classification percentage (leave-one-out method)
DA on polynomial coefficients

• Discriminant analyses run to compare read speech and spontaneous data sets
• Separate tests on F1 & F2 with the predictor variables:
  Quadratic - 3 coefficients
  Cubic - 4 coefficients
  Quintic - 5 coefficients
Classification rates – read speech

% correct classification

F1

39%

F2

49%

quadratic

cubic

quartic

chance

DyViS
Classification rates – spontaneous speech

- **F1**
  - Quadratic: 19%
  - Cubic: 9%
  - Quartic: 9%

- **F2**
  - Quadratic: 9%
  - Cubic: 9%

Chance classification level: 19%
Discussion

Read speech:
- /uː/ offers good levels of discrimination (N.B. 20 speakers)
- cubic approximation best for F1 and F2

Spontaneous speech:
- large reduction in accuracy of discrimination for /uː/ formant dynamics
- discrimination still better than chance, especially in F1 (19%)
Discussion

Possible reasons for reduction in accuracy:

• increased variability due to spontaneous condition - /uː/ in different locations in utterances, receiving different levels of stress

• /uː/ in range of consonantal contexts in spont. data: *Cooper, coot, Hooper, poodles, scooter, supervisor* (all h_d in read speech)

Difficult to separate these sources of variability → controlled set of consonantal and prosodic contexts in read speech would be helpful
Discussion

• /uː/ contexts: Cooper, coot, Hooper, poodles, scooter, supervisor

Preceded by /k, h, p, sk, s/
Followed by /p, t, d/

• Hypothesise: demi-syllable or complete word is the appropriate domain for examination of speaker-specific aspects of formant dynamics

• Intuitively assumed this in earlier experiments - /aɪk/ (McDougall 2004, 2005)
Forensic Implications

• Casework situations – generally examining spontaneous speech
• Formant dynamics of SSBE /uː/ offer some speaker-specific information in spontaneous speech
• Likely to be important to compare tokens in like consonantal and prosodic contexts
Conclusion

• Formant dynamics useful to examine for speaker differences since they reflect differences among speakers in vocal tract dimensions and articulatory strategies
• F1 and F2 of /uː/ yield individual variation in shape and frequency of formant contours - contours approximated with polynomial equations (quadratic, cubic, quartic)
• Speaker-distinguishing potential of polynomial coefficients quantified with Discriminant Analysis
Conclusion

• /u:/ in read speech (hVd) achieved promising levels of discrimination
  39% F1, 49% F2 – cubic approximations
• /u:/ in spontaneous speech less successful
  19% F1 (cubic), 9% F2 (quadratic & quartic)
  - probably due to range of consonantal and prosodic contexts
• Further work to include examination of formant dynamics in spontaneous speech in more tightly controlled contexts
Acknowledgements

• UK Economic and Social Research Council for funding the DyViS Project ‘Dynamic Variability in Speech: A Forensic Phonetic Study of British English’ [RES-000-23-1248]
• BT for a donation towards the telephone investigation
• Members of the DyViS project
• Toby Hudson, Gea de Jong and Caroline Williams for assistance with recording subjects, script-writing and measurement
• Geoffrey Potter for technical assistance
Distinguishing speakers using formant dynamics in read and spontaneous speech: a study of British English /uː/