Nguni and Sotho varieties of South African English – distant cousins or twins?

F. de Wet, T.R. Niesler, P.H. Louw
Introduction

- South Africa has 11 official languages
- many more spoken on the streets
- lingua franca: English
- variety South African English accents
- consequence for ASR development
  – accent modelling required
Which accents to model?

- English English (EE)
- Black English (BE)
- Afrikaans English (AE)
- Indian English (IE)
- Coloured English (CE)
## Black South African English

- refers to English as spoken by native speakers of 9 of 11 official languages
- one homogeneous group?
- different group per mother tongue?
- possible sources of information
  - mother tongue speakers
  - intuition 🎧 (Xhosa) 🎧 (Tswana)
  - literature: many contradictory findings
Nguni & Sotho English

- 2 main language families
- **Nguni**: Zulu, Xhosa, Swati, Ndebele (45.7%)
- **Sotho**: Northern Sotho, Southern Sotho, Tswana (25.5%)
- perceptual experiment
- ASR experiment
Perceptual experiment: data

• BE & EE components of the African Speech Technology (AST) database
• 180 stimuli
  – 30 isolated words (limited context)
  – 30 phrases (variety of sounds, prosodic cues)
  – produced by native speakers of EE, NE, SE
• phonetic content: according to descriptions of BSAE in the literature
Perceptual experiment: method

- speakers: from AST database
- listeners: recruited on campus
- mesolect speakers of English
- male:female = 50:50
- Nguni:Sotho = 50:50
- praat (www.praat.org)
- “Can you identify the language group to which this speaker belongs?”
- English, Nguni, Sotho, I don’t know
Perceptual experiment: results

- 14% “I don’t know” (mostly isolated words)
- EE vs. BE: 80% correct
- NE vs. SE:

![Bar chart showing % correct for all, words, and phrases.](chart.png)
ASR experiment: data

- NE & SE selected from BE component of AST database (2.5 hours training data)
- BE training set (0.5 NE + 0.5 SE)
- mesolect speakers of English
- male:female = 50:50
- fixed line:mobile = 50:50
- development & independent test sets
ASR experiment: method

- HTK
- MFCCs & derivatives, CMN (per utterance)
- speaker-independent, crossword triphone models (3 states, 8 Gaussians/state)
- triphone clustering (600 clustered states)
- phone recognition
ASR experiment: results

![Bar chart showing PER (%) for NE, SE, and BE data.]
Conclusions

• neither perceptual nor ASR experiment revealed any discernible difference between NE & SE

• for purposes of accent modelling for AST data, BE may be treated as a homogeneous accent group
Future work

- follow-up perceptual experiments with “better” data
- no substantial changes in results
- similar ASR experiment