Speech and Language in Education

Martin Russell
School of Engineering, University of Birmingham, UK
m.j.russell@bham.ac.uk
Priors (Prejudices?)

- **Pronunciation tuition:**
  - Use techniques from ASR to decide if a spoken utterance is a ‘good pronunciation’
    - People speak to communicate. Everybody sounds different, and we don’t worry too much about ‘good’ pronunciation provided that we get the message
    - What is ‘good pronunciation’?

- **Reading tuition:**
  - Use techniques from ASR to track reading against a text and detect ‘incorrectly read words’
    - Should the emphasis be accurate reading and pronunciation or fluency and comprehension?
    - What are ‘incorrectly read words’?
Everybody sounds different: Some examples from the ‘Accents of the British Isles’ (ABI) speech corpus

- Burnley (Lancashire)
- Elgin (Scottish Highlands)
- Glasgow
- Belfast (Ulster)
- Dublin
- Liverpool
- Denbigh (North Wales)
- Truro (Cornwall)
- Newcastle
- Burnley (Lancashire)
- Hull (East Yorkshire)
- Birmingham
- Lowestoft (East Anglia)
- Inner London

Standard Southern English
Computer aided pronunciation learning system using speech recognition techniques, Sherif Mahdy Abdou, Salah Eldeen Hamid, Mohsen Rashwan, Abdurrahman Samir, Ossama Abd-Elhamid, Mostafa Shahin, Waleed Nazih

- Computer Aided Pronunciation Learning
- Arabic for non-native speakers
- Application: Correct recitation of holy Qur’an
  - Must be recited according to classical Arabic dialect
  - Little allowed variation
- Commercial product: HAFSS
  - HMM-based verification, with speaker adaptation
  - Pre-generation of likely errors, confidence scoring
- Performance
  - Detects 62% of pronunciation errors (6.6% of data)
- How is acceptable pronunciation defined?
- Are there other, similar applications?
Pronunciation verification of children’s speech for automatic literacy assessment, Jorge Silva, Abe Kazemzadeh, Hong You, Sungbok Lee and Shrikanth Narayanan

- Computer Aided Pronunciation Learning
  - When do pronunciation variants indicate poor reading skills?
- Children with Spanish language background
- Evaluation (“acceptable” vs “unacceptable”):
  - 20 human evaluators (varying degrees of Spanish)
    - Broad agreement amongst human assessors
  - Comparison of:
    - Human assessment, automatic threshold based methods, automatic decision-tree based methods (trained on ‘human voting’ and transcriptions)
- Automatic methods perform about as well as human evaluators
- Transcription good enough to train decision tree
Pronunciation verification of children’s speech for automatic literacy assessment (Continued)

- Even though ‘good pronunciation’ is not well defined, we have an automatic system which is able to replicate subjective human judgements.
- By exploring how these classifiers work, can we derive a better understanding of how people decide between acceptable and unacceptable pronunciations?
Is ASR accurate enough for automated reading tutors, and how can we tell? Jack Mostow

- Automatic reading tuition
- Discussion paper: three ASR functions considered:
  1. Tracking reader’s position in text
  2. Detecting reading mistakes that a human would correct
  3. Measuring word reading times
- Subjective conclusion is that all of these can be achieved to some extent:
  1. Can be done well enough to detect skips or end of sentence, but not always well enough to identify which word misread
  2. Can be done well enough to avoid frustration, detect some mistakes a human would correct, but not to indicate which words are wrong
  3. Can be done well enough to estimate fluency
Classroom success of an intelligent tutoring system for lexical practice and reading comprehension,
Michael Heilman, Kevyn Collins-Thompson, Jamie Callan, Maxine Eskenazi

- Automatic selection of authentic texts for lexical practice and reading comprehension in ESL teaching
- REAP system, deployed in English Language Institute (ELI), University of Pittsburgh
- Documents gathered from web:
  - Analysed for occurrence of target vocabulary, syntactic features, length and readability
  - Only about 0.5% of retrieved documents suitable
- Assessment
  - Positive user opinions
  - Analysis of email correspondence from curriculum supervisor
- Grades to be assigned according to progress with REAP
Some thoughts…

- Problems in automatic pronunciation assessment and reading tuition expose our lack of real understanding of variability in speech.
- Statistical ASR techniques accommodate variability at a local level but dodge long-term phenomena, such as accent, which provide a framework to explain this variability.
  - For example: “hood”
- Need computationally useful models of phenomena such as accent - maybe STiLL can stimulate research which contributes to better ASR?
Some more thoughts…

- If ASR techniques could accurately identify unacceptable pronunciations, how would we exploit this? What feedback could we give?
- Unified models for recognition and synthesis?
- Articulatory-based ASR?
And finally...

- Even imperfect STiLL has some serious advantages:
  - Patient
  - Non-judgemental
  - Unlimited time for one-to-one attention
  - ...

- Thanks to the authors, and to the session organisers