Speech Processing 15-492/18-492

Speech and Education
Reading Tutors
Computer Aided Language Learning
Speech Systems in Education

*Making it easier to learn*
- To learn new languages
- Learn your own language
- Learn other skills (STEM)
  - Science, Technology, Engineering Math

*Technologies used*
- ASR, TTS, Dialog etc

*Do they work?*
- People pay for them
Reading Tutors

- **Listening to reader**
  - Use ASR to follow what is being said
- **Show where they are in the text**
- **Show errors in their reading**
- **Give hints when they are stuck**
  - *(which requires detected when they are stuck)*
Reading Tutors

- **ASR through “forced alignment”**
  - You know what is to be read
  - Align given speech with expected speech
    - Mismatches could be speaker variation
    - Or reading errors
  - Expected can be strong LM not “forced”
- **Readers are typically children**
  - No fluent speech
  - Children’s speech is different (higher, multi-register)
Reading Tutors

- Can give individual aid
  - Dedicated teacher may be better but theirs is only one teacher
- All students get to read
- Can be fun
  - It's different from reading in class
- Has been shown to give some improvements
Using speech to aid language learning

- Pronunciation training
- Reading tutors
- Dialog practice
- Familiarity with language
  - Reading/listening to appropriate examples
Pronunciation training

- **Hearing well-articulated examples**
  - Golden voice recordings
  - Limited domain synthesis
  - Must be high quality

- **Correct speakers error**
  - High precision speech recognition
  - Identify issue like formant, timing, voicing issues
How to present information to user

- **Audio only**
  - Golden voice
- **Spectrogram**
  - Show spectral differences
- **Show articulatory movements**
  - Show tongue, teeth, lips etc.
- Or combinations of all of these
- Different users need different stimuli
The Golden Voice

- **Appropriate example voice**
  - “Good” voice with appropriate style/accent
  - Very well recorded
  - Can be limited domain synthesis
    - But must have no artifacts

- **Different voices for different people**
  - Same (or different genders)
  - Their own voice (cross-lingual voice conversion)
Finding appropriate material

- **Search web for appropriate articles**
  - **Subject** (something interesting)
  - **Learner-level**: appropriate reading level

- **Can be**
  - News, stories, even video

- **Hypothesis**
  - Seen relevant/appropriate level material helps

- **CMU’s Reap Project**
Language Learning through listening

- Can use TTS to play things
  - Increases familiarity with language
  - Increases understandings
  - Can skew pronunciation
  - Can skew prosody
  - But is very convenient (and very common)
Learning through Interaction

- *Use dialog system to improve interaction*
- *Leads you through a conversation*
  - Can be done multiple times
  - No human teacher required
- *May not be fully realistic*
  - Can be better than nothing
Learning through 3D worlds

- **USC/Alelo (Lewis Johnson)**
  - Game-based learning
  - Interact with characters in a scenario
    - Village elders, hotel reception, checkpoint
  - Real-time interaction
  - Gives corrective feedback (what you did wrong)
- **Cultural as well as linguistic training**
Learning through games

- **Quick responses**
  - *Travel through a maze through voice commands (in new language)*

- **Klingon game**
  - *First ASR game (around 1995)*
  - *Had to talk to characters in Klingon*
    - Bad ASR was your fault!
Non language learning

- **Intelligent tutoring systems (ITS)**
  - Mostly text based
  - Some becoming speech based
    - But ASR is hard (especially with children’s speech)

- **Sometimes game-based**
  - Adventure like, puzzle solving
  - Speech in “educational” toys (Leapfrog)