Speech Processing 15-492/18-492

Speech Processing
Current Topics and Future challenges
Commercial and Research
Current and Future

- **What are the hot topics in Speech**
  - What currently works
  - What could work soon (5-10 years)
- **What are the industry hot topics**
- **What are the research challenges**
Spoken Dialog: Now

- **Industry:**
  - *Location based querying*
    - Google: 411, smartphone
    - Microsoft Live Search: smartphone
    - Yahoo (Vlingo)
  - Blackberry, IPhone
    - (Owners have money)
  - *How do you make money out of this ...*
Spoken Dialog: Now

- **Research**
  - Error recovery
  - Adaptive systems
  - Rapid deployment
  - Learning dialog structure from data
Industry

- Moving from grammar based to N-gram based
- Broadcast news transcription of IR
- Robust speech recognition:
  - In car, outside, in noisy office
- LM adaptation from other sources
  - Using click through and search queries
- Pronunciation variants (“wrong” ones too)
- Medical transcription
Research:

- Discriminative training
  - Acoustic parameter projections to discriminate between the correct answers and competitors

- Robust recognition
  - Far field microphones
  - Blind source separation

- Out of vocabulary words

- Unsupervised training
Industry

- Building custom voices (and your voice)
- Multilingual on small devices
  - E.g. for GPS Navigation over Europe
- Easy methods to build new languages
TTS: Now

- **Research**
  - Improving statistical synthesis
  - Rapid support in new languages
  - Emotional speech synthesis
  - Automatic building of voices from data
    - Without any human intervention
  - Synthesis beyond the sentence
    - Synthesis with more text analysis
Speech to Speech Translation

- **Industry**
  - One way systems, domain limited systems
  - Simple targeted cell phone systems

- **Research**
  - Two way systems, large domains
  - One way lecture/broadcast news
VC and SID: Now

- **Voice conversion**
  - Cross Lingual Voice Conversion
  - Emotion/style conversion
  - Conversion without training data

- **Speaker ID**
  - Accuracy on large data sets (> 1000 speakers)
  - Cross channel/language ID
  - More information in ID (prosody, vocab)
CALL: Now

- **Industry**
  - Pronunciation training
  - Scenario practicing

- **Research**
  - Game based tools
  - Measuring educational contribution
Speech Processing Future

- **Hard challenges** *(PhD topics and beyond)*
- **All on the research side**
  - But maybe in Research Labs
Speech Reco without Speech

- **Using other modalities**
  - Lip movement, muscle movement
  - Silent speech
    - No generated audio
    - Just think about the words
  - Gesture recognition
Participant in a meeting

- True conversational speech
  - Appropriate non-word speech generation
  - Know when to speak, when to laugh, when to listen
  - Appropriate timing conversation
  - Able to interrupt when having something to say

- Have something to say
Summaries and Discussions

- **Describe a paper/movie/event**
  - Appropriate summary
  - Allow questions
  - Know when to use style/emotion
  - Not just speech<-text
    - Understand more of the text content
Final Notes

- Don’t forget to fill in Faculty Course Evaluation
- Final Homework due
  - Monday 8th 3:30pm
- Final exam
  - Tuesday 16th 1pm-4pm WeH 6423