Speech Processing 11-492/18-492

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**
    - On phone: Apple (Siri)
    - In home: Amazon (Echo)
  - **Smartphones, Tablets:**
    - (Owners have money)
  - **How do you make money out of this ...**
Research

- Error recovery
- Adaptive systems
- Rapid deployment
- Learning dialog structure from data
- Non-task oriented dialog
Industry

- Moving from grammar based to N-gram based
  - Adapting cloud ASR per app.
- Broadcast news transcription for IR
- Robust speech recognition:
  - In car, outside, in noisy office, far field
- 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**

- **Deep Learning (Neural Nets)**

- **End-to-end speech systems**
Industry

- Building custom voices (and your voice)
- Multilingual on small devices
  - E.g. for GPS Navigation over Europe
- Easy methods to build new languages
- Conversational Speech
  - Intent, style, expressive
Research

- Rapid support in new languages
- Emotional speech synthesis
- Automatic building of voices from found data
  - Without any human intervention
- Languages without Orthography
- Synthesis beyond the sentence
  - Synthesis with more text analysis
- Neural models
- End-to-end systems
Speech to Speech Translation

**Industry**
- One way systems, domain limited systems
- Simple targeted cell phone systems
- Youtube/Broadcast translation
- Skype translation

**Research**
- Two way systems, large domains
- One way lecture/broadcast news
- Intent transfer
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
  - Brain Computer Interfaces

- **ASR without text**
  - Find “.....” in all this audio
Beyond the Words

- Recognition of more than words
- Intent, style, emotion
- Human-Machine
  - Frustration, confidence, agreement
- Human-Human
  - Rapport, relationships, persuasion
  - Truth and lies
  - Sentiment
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

- **Answer complex questions**
  - Engage user and discuss topic
Note

- Don’t forget to fill in Faculty Course Evaluation
- Final exam
  - Friday 15\textsuperscript{th} 1:00pm-4:00pm WEH 4623
  - 18-492 – is in the same room!