Speech Processing 15-492/18-492

Using Speech with Computers

Alan W Black
August 2008
Overview

- **Practical and Theory:**
  - Understand concepts, Implement Solutions
- **Speech Recognition**
  - Speech to text
- **Speech Synthesis**
  - Text to Speech
- **Spoken Dialog Systems**
  - Interaction with machines
Course Schedule

- **MWF 3:30-4:20**
- **DH 1117**
- **Lecturer: Alan W Black (awb@cs.cmu.edu)**
- **TA: David Huggins (dhuggins@cs.cmu.edu)**
- **http://www.speech.cs.cmu.edu/15-492/**
Course Details

◆ Three lectures a week
◆ 4 Homeworks
  ● Speech Recognition
  ● Speech Synthesis
  ● Spoken Dialog System
  ● Other
◆ Final Exam
Homeworks

(Mostly) Practical

- **Build something that talks/can be spoken to**
- **Software and speech data will be provided**
  - Will run on Windows/Linux or OSX
  - Access to Linux servers if required
- **Written description of what you did**
Schedule Details

- **Week 1 (Aug 15th)**
  - Applications, Human and Computer Speech Processing
- **Week 2-4 (Sep 3rd) Speech Recognition**
  - Signal representation, acoustic modeling
  - Language modeling, applications
  - Tuning, evaluation, expectations
Course Details

- **Week 5-7 (22nd Sep) Speech Synthesis**
  - Text processing, prosody, waveform synthesis
  - Building voices, evaluations, voice conversion

- **Week 8 (13th Oct) Multilinguality**
  - Supporting new languages efficiently

- **Week 9-11 (20th Oct) Dialog Systems**
  - VoiceXML, Mixed initiative, barge-in
  - Design, installation and tuning.
Course Details

◆ **Week 12 (10\textsuperscript{th} Nov)**
  - *Speech to Speech translation*
  - *Language support, tight integration*

◆ **Week 13 (17\textsuperscript{th} Nov)**
  - *Evaluation and expectations*

◆ **Week 14 (24\textsuperscript{th})**
  - *Speaker ID, Silent Speech, Conversion*
  - *What still needs to be done.*

◆ **Week 15 (1\textsuperscript{st} Dec)**
  - *Exam*
Why Speech

- **Most natural way to communicate**
  - (For Humans)
- **Not ideal for everything**
  - Graphics and text can be better (sometimes)
- **Doesn’t compress well**
- **Hard to search**
Compression

- Alice in Wonderland
  - Text
    - 150K uncompressed
    - 43K compressed
  - Speech (2hrs 20mins)
    - 270M uncompressed
    - 600K compressed (mp3, 24KBS)
Find all NPR broadcasts mentioning Obama
  - Listen to them all

From lecture recordings
  - Find all occurrences of “this will be in the exam”

So listen to it faster …
  - Normal  2x speed
  - 2x  4x  8x
Eyes/Hands Free

- **Interaction when driving**
  - Look at screen to see next turnoff
  - “In 200 yards turn right onto Murray Ave.”

- **Blind users/ Assistive technology**
  - Text isn’t very useful

- **Alerts**
  - “Will self-destruct in 10 seconds” vs blinking light

- **Telephone dialog systems**
Speech Applications

- Command and Control
- Information Agents
- Speech to Speech Translation
- Speech summarization
  - Lecture or Meeting summarization
- Transcription/Dictation
- Speaker Identification
  - emotion/dialect/language
- Language Learning
“Hot” Commercial Applications

- **Location-based services:**
  - Yahoo GO
  - Google Maps
  - Microsoft Live Search

- **All phone/pda based**
  - Use speech-in
  - Directions speech-out
Other Speech uses

- Spoken Dialog Systems
  - Let’s Go Public 412 268 3526 evenings 412 442 2000
  - Pittsburgh bus timetables by phone
- Assistive Technologies
  - Screen readers
  - Augmentitive and assistive communication devices
- On-line Personalization
  - Blogcasts (your voice, or appropriate voice)
  - Game character customization
- Talking Heads
  - CMU’s roboceptionist
- Singing Synthesis
  - XML interface for song specification