



# Speech Processing 15-492/18-492

---

Spoken Dialog Systems  
Conversing with machines

# Spoken Dialog Systems

- ◆ *Not just ASR bolted onto TTS*
- ◆ *Different styles of interaction*
  - *IVR/Tree question/response systems*
  - *Mixed initiative systems*
  - *“How May I Help You?” open questions*
  - *True conversational machine-human interaction*
  - *Strings of characters to words*

# SDS Overview

- ◆ *Introduction*
- ◆ *Building simple dialog systems*
- ◆ *VoiceXML*
  - *A language for writing systems*
- ◆ *Beyond tree-based systems*
  - *CMU's Olympus systems*
- ◆ *Real-world deployment considerations*

# SDS Applications

- ◆ *Information giving*
  - *Flights, buses, stocks weather*
  - *Driving directions*
  - *News*
- ◆ *Information navigators*
  - *Read your mail*
  - *Search the web*
  - *Answer questions*
- ◆ *Provide personalities*
  - *Game characters (NPC), toys, robots*
- ◆ *Speech-to-speech translation*
  - *Cross-lingual interaction*

# Dialog Types

## ◆ *System initiative*

- *Form-filling paradigm*
- *Can switch language models at each turn*
- *Can “know” which is likely to be said*

## ◆ *Mixed initiative*

- *Users can go where they like*
- *System or user can lead the discussion*

## ◆ *Classifying:*

- *Users can say what they like*
- *But really only “N” operations possible*
- *E.g. AT&T? “How may I help you?”*

# System Initiative

- ◆ *Most common*
  - *Machine controls the call*
  - *Few choices in the dialog*
- ◆ *Simple form filling:*
  - *What is your bank account number*
- ◆ *Advantages:*
  - *You know what users will say (sort of)*
  - *Hard for user to get confused*
  - *Hard for system to get confused*
  - *Easy to build*
- ◆ *Disadvantages:*
  - *Limited flexibility in interaction*
  - *Fixed dialog structure*
- ◆ *Most reliable, but many turns*

# System Initiative

## ◆ *Let's Go Bus Information*

- *412 442 2000 (Evenings)*
- *Provides bus information for Pittsburgh East End (61x 5[469]x)*



## ◆ *Tell Me*

- *Company getting others to build systems*
- *Stocks, weather, entertainment*
- *1 800 555 8355*

# Mixed Initiative

- ◆ *User or system takes initiative*
  - *More interesting dialogs*
  - *“jump” through different parts of dialog state*
- ◆ *Advantages*
  - *More realistic dialog*
  - *Can do more complex tasks*
- ◆ *Disadvantages*
  - *Can get confusing*
  - *Can miss important parts*

Vera

# Classification Dialogs

## ◆ *Sort out from N things*

- *User says “anything” and system directs them*
- *Receptionist*
  - ⊗ *I have a problem with my bill*
  - ⊗ *What’s the area code for Miami*
  - ⊗ *Did you know I can see the beach from here*

## ◆ *Advantages*

- *(Apparently) complex understanding*
- *Solves a very common task*

## ◆ *Disadvantages*

- *Actually quite restrictive*
- *Needs data to train from*
- *Needs to be updated*

# Beyond Telephones

## ◆ *Telematics*

- *Voice communication in cars*
- *CPS, music selection etc*

## ◆ *Robot Interaction*

- *Robot-robot and robot-human interaction*

## ◆ *Animated talking head*

- *Non-player characters – web agents*

## ◆ *Speech to Speech translation*

# Team Talk

- ◆ *Using speech to control multiple robots*
  - *Robots have names and distinct voices*
  - *They report to each other and to you in voice*

# USI

- ◆ *Lots of different interfaces is confusing*
  - *Try to have general expectations and discover*
- ◆ *Try for some level of standardization*
  - *(like programming applications: file menu)*

# True conversation

- ◆ *Requires more than just speech*
  - *Non-verbal noises: laughing, er, um, etc*
  - *Eye gaze*
  - *Proper timing (not waiting 500ms before speaker)*
  - *Back-channeling*
  - *Movement*
  - *Talking about nothing*

# Roboreceptionist

- ◆ *Entrance to NSH*
  - *Keyboard (no ASR)*
  - *TTS, face, movement*
  - *Range finder to detect people*
  - *Significant background character*
- ◆ *Mostly talks about nothing*



