




## Interaction Styles 3


Speech and natural language interfaces



## Agenda

- Discuss midterm
- Discuss project part 2 & 3
- Recognition technologies and interfaces
  - ❖ Part 1: Speech (more on Thursday)


Fall 2003 PSYCH / CS 6750 2



## Recognition UIs

<ul style="list-style-type: none"> <li>➤ Motivation           <ul style="list-style-type: none"> <li>❖ Natural interaction</li> <li>❖ Efficient, powerful interaction</li> </ul> </li> <li>➤ Recognize or capture           <ul style="list-style-type: none"> <li>❖ Modal interaction</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>➤ Issues           <ul style="list-style-type: none"> <li>❖ Visibility</li> <li>❖ Error recovery</li> <li>❖ Thresholds for recognition</li> <li>❖ Specifying focus</li> </ul> </li> </ul>
---	--


Fall 2003 PSYCH / CS 6750 3



## Natural Language

- Natural input (written and spoken)
- Expression and ambiguity
- Recognition errors
- Interaction dialogue is really difficult


Fall 2003 PSYCH / CS 6750 4



## Natural Language Understanding

- Putting *meaning* to the words
- Input might be speech or could be typed in or written
- Holy grail of Artificial Intelligence problems

Fall 2003 PSYCH / CS 6750 5



## Advantages

- Easy to learn/remember
  - ❖ domain, not GUI
- Less transfer problems (rm or delete)
- Enormous potential (direct representation)
- Fast, efficient (expressability)
- Little screen real estate required

Fall 2003 PSYCH / CS 6750 6


## Disadvantages

- Assumes domain knowledge
- Requires confirmation/clarification
- Error-prone input (typing, voice)
- Unrealistic expectations
- Generate mistrust/anger
- Qualify focus

Fall 2003 PSYCH / CS 6750 7

## A Voice Interface

By Scott Adams



Fall 2003 PSYCH / CS 6750 8

## When to Use Speech

- Eyes busy
- Hands busy
- Mobility required
- Visual impairment
- Physical limitation
- Stressful environment
- Conditions preclude use of keyboard
  - ❖ Vibration, cold, water, hygiene, public use

Fall 2003 PSYCH / CS 6750 9

## Voice Recognition Concepts

- Speaker independent/dependent
- Discrete or continuous
- Vocabulary size
- Lots of tradeoffs


Fall 2003 PSYCH / CS 6750 10

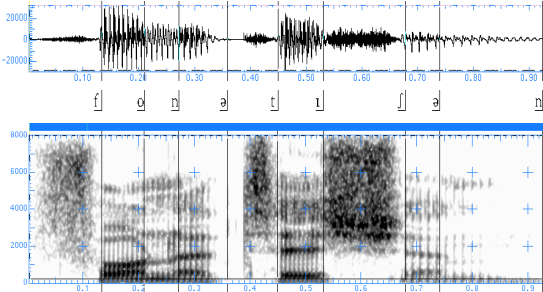
## What is Speech?

- What is speech?
  - ❖ Vibrations of vocal cords creates sound "ahh"
  - ❖ Mouth, throat, tongue, lips shape sound
- English speech
  - ❖ 40 phonemes; 24 consonants, 16 vowels
- Sounds transmit "language"

Fall 2003 PSYCH / CS 6750 11

## Waveform & Spectrogram

- Speech does not equal written language 



Fall 2003 PSYCH / CS 6750 12

## Parsing Sentences

"I told him to go back where he came from, but he wouldn't listen."

Fall 2003      PSYCH / CS 6750      13

## Speech Input

- Speaker recognition
- Speech recognition
- Natural language understanding

Fall 2003      PSYCH / CS 6750      14

## Speaker Recognition

- Tell which person it is (voice print)

- Could also be important for monitoring meetings, determining speaker

Fall 2003      PSYCH / CS 6750      15

## Speech Recognition

- Primarily identifying words
- Improving all the time
- Commercial systems:
  - ❖ IBM ViaVoice, Dragon Dictate, ...

Fall 2003      PSYCH / CS 6750      16

## Recognition Dimensions

- Speaker dependent/independent
  - ❖ Parametric patterns are sensitive to speaker
  - ❖ With training (dependent) can get better
- Vocabulary
  - ❖ Some have 50,000+ words
- Isolated word vs. continuous speech
  - ❖ Continuous: where words stop & begin
  - ❖ Typically a pattern match, no context used


Did you  
vs.  
Didja

Fall 2003      PSYCH / CS 6750      17

## Recognition Systems

- Typical system has 5 components:
  1. Speech capture device - Analog -> digital converter
  2. Digital Signal Processor - Gets word boundaries, scales, filters, cuts out extra stuff
  3. Preprocessed signal storage - Processed speech buffered for recognition algorithm
  4. Reference speech patterns - Stored templates or generative speech models for comparisons
  5. Pattern matching algorithm - Goodness of fit from templates/model to user's speech


Fall 2003      PSYCH / CS 6750      18



## Errors

- Systems make four types of errors:
  - ❖ Substitution - one for another
  - ❖ Rejection - detected, but not recognized
  - ❖ Insertion - added
  - ❖ Deletion - not detected
- Which is more common, dangerous?


Fall 2003 PSYCH / CS 6750 19



## Speech Output

- Male or female voice?
  - ❖ Technical issues (freq. response of phone)
  - ❖ User preference (depends on the application)
- Rate of speech
  - ❖ Technically up to 550 wpm!
  - ❖ Depends on listener (blind: 150-300 wpm)
- Synthesized or Pre-recorded?
  - ❖ Synthesized: Better coverage, flexibility
  - ❖ Recorded: Better quality, acceptance


Fall 2003 PSYCH / CS 6750 20



## Speech Output

- Synthesis
  - ❖ Quality depends on software (\$\$)
  - ❖ Influence of vocabulary and phrase choices
- Recorded segments
  - ❖ Store tones, then put them together
  - ❖ The transitions are difficult (e.g., numbers)
- Numbers
  - ❖ Record three versions (rise, flat, fall) ?
  - ❖ Logic to determine which version to play


Fall 2003 PSYCH / CS 6750 21





## Designing the Interaction

- Constrain vocabulary
  - ❖ Limit valid commands
  - ❖ Structure questions wisely (Yes/No)
  - ❖ Manage the interaction
  - ❖ Examples from the airline systems?
- Slow speech rate, but concise phrases
- Design for failsafe error recovery
- Process preview & progress indicator


Fall 2003 PSYCH / CS 6750 22



## Speech Tools/Toolkits

- Java Speech SDK Talking Clock 
  - ❖ FreeTTS 1.1.1 <http://freetts.sourceforge.net/docs/index.php>
  - ❖ *"For 3/4 or 75% of his time, Dr. Walker practices for \$90 a visit on Dr. Dr., next to King Philip X of St. Lameer St. in Nashua NH."* 
- IBM JavaBeans for speech
- Visual/Real Basic speech SDK
- OS capabilities (speech recognition and synthesis built in to OS) (TextEdit)
- VoiceXML

Fall 2003 PSYCH / CS 6750 23



## Recall

- A natural language interface need not be speech
  - ❖ Pen and typing are also natural
- A speech interface need not use natural language (might be more command language-like)
- Wizard of Oz evaluations are particularly useful in this area

Fall 2003 PSYCH / CS 6750 24



## Upcoming

- Pen & PDA interfaces
- Evaluation (with users)
- More evaluation 😊