Speech in User Interfaces

- Why is speech useful?
  - input and output
- Speech output technology
- Speech input technology

Notes and exercise inspired by Lai & Yankelovich, CHI'99 tutorial
Benefits of Speech

- Data entry possible without keyboard
  - mobile computing
- Excellent for hands/eyes busy situations
- Bad typists!
- Natural
- People with visual disabilities
Problems of Speech

• Controlling things, describing complex images
• High expectations
  – A three-year old is better than current technology
• Speech output sounds unnatural
• Transient
• Input is error prone
• Asymmetrical
  – speech input is faster then typing
  – speech output is slower than reading
Problems of Speech

- Public
- Noisy environments
Speech Synthesis

• Written text transformed into speech
  – text-to-speech

• Two types of synthesiser
  – parameterised
  – concatenative

• Parameterised
  – formant based - use rules based on signal from spoken input
  – articulatory - use model of vocal tract

• More like musical instrument synthesis
Speech Synthesis

• Concatenative - word
  – just record all the words you need
  – good for small sets
  – BT directory enquiries

• Concatenative - phoneme
  – phoneme - smallest unit of speech that differentiates one word from another
  – bath, bat, pat
  – makes more natural sounding speech
Speech Synthesis

- Words/phonemes sound different depending on where they occur in a sentence
- More like sampling
Speech Synthesis Problems

- Naturalness
- Understandability
- Words not in dictionary
- Rules are hard to get right
- Prosody
  - stress, pitch, intonation
Speech Recognition

• General architecture:

Acoustic Signal → Microphone/telephone → Speech Engine

Engine matches signal to stored data → Recognised Words
**Speech Recognition Issues 1**

- Continuous versus Discrete recognition

- **Discrete**
  - reduces coarticulation
  - improves accuracy
  - reduces computation

- **Continuous**
  - hard!
  - natural / fast

It is very similar to trying to read text with all the spaces removed. It is very hard to do this and it takes a lot of time to work out where each word stops and the next starts.
Speech Recognition Issues 2

• Speaker dependent versus independent
  • Dependent
    – requires training - takes time
    – can get good recognition rates
  • Independent
    – great for ‘walk up and use’ systems
    – lower recognition rates in general
Speech Recognition Issues 3

- Vocabulary size
- Smaller the size the higher the recognition rates
  - 10 - phone digits
  - 100 - menu selection
  - 10k - 60k - general dictation, email
- Current desktop SR can get around 98% on large vocab
Speech Recognition Issues 4

• Accuracy

• What is an error?
  – out of vocabulary
  – recognition failure
  – mis-recognition
    – insertion / deletion / substitution

• Hard to tell mis-recognition
Recognition Errors

- User spoke at the wrong time
- Sentence not in grammar
- User paused too long
- Words sound alike
- Word out of vocab
- User has a cold
- Over-emphasis
Recognition Errors

- Can be serious
  - mis-recognition changes the meaning of utterance
- Re-prompt with explicit choices
- Give a list of possible matches
- Allow user to spell word
  - problems with “e set”: b, c, d, e, g, p, t, v (z)
- Use visual cues if available
• **Speech engine contains**
  – vocabulary
  – language model
  – acoustic model
• **Language model**
  – grammar
  – gives likelihood of a particular sequence of words
  – differentiates between similar sounding words based on context - Wright, right, write

• **Acoustical model**
  – deals with variation in pronunciation
**Human Issues**

- Listeners will give your interface a personality and respond in human ways
- Do not violate conversational conventions and politeness!
Resources

• http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html
  — good survey of speech i/o but from 1996
• Fundamentals of speech synthesis and speech recognition: basic concepts, state of the art and future challenges, edited by Eric Keller
  — in library
Resources

• An introduction to text-to-speech synthesis, Thierry Dutoit
  — in library

• Speech FAQ
  — http://www.speech.cs.cmu.edu/comp.speech/
Demos

- www.genmagic.com/portico/demos
- www.dragonsys.com/products/video
- A really good text-to-speech engine: