

Speech in User Interfaces

DEMM54 03 S Brewster ©

Speech-1

- Why is speech useful?
 - input and output
- Speech technology
- Speech UI design
- Car design exercise



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

Benefits of Speech

DEMM54 03 S Brewster ©

Speech-2

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



Problems of Speech

DEMM54 03 S Brewster ©

Speech-3

- Remember IS3
 - 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 than typing
 - speech output is slower than reading



Problems of Speech

DEMM54 03 S Brewster ©

Speech-4

- Public
- Noisy environments



Speech Synthesis

DEMM54 03 S Brewster ©

Speech-5

- 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

Good

Medium

Poor

Good UK male

Good UK female



Speech Synthesis

DEMM54 03 S Brewster ©

Speech-6

- 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

DEMM54 03 S Brewster ©

Speech-7

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



Speech Synthesis Problems

DEMM54 03 S Brewster ©

Speech-8

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

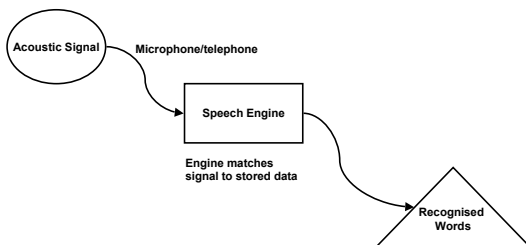


Speech Recognition

DEMM54 03 S Brewster ©

Speech-9

- General architecture:



Speech Recognition Issues 1

DEMM54 03 S Brewster ©

Speech-10

- Continuous versus Discrete recognition
- Discrete
 - reduces coarticulation
 - improves accuracy
 - reduce computation
- Continuous
 - hard!
 - natural / fast

It is very similar to trying to read text with all of 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

DEMM54 03 S Brewster ©

Speech-11

- 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

DEMM54 03 S Brewster ©

Speech-12

- 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

DEMM54 03 S Brewster ©

Speech-13

- Accuracy
- What is an error?
 - out of vocabulary
 - recognition failure
 - mis-recognition
 - insertion / deletion / substitution
- Hard to tell mis-recognition



Recognition Errors

DEMM54 03 S Brewster ©

Speech-14

- 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



Speech Engine

DEMM54 03 S Brewster ©

Speech-15

- Speech engine contains
 - vocabulary
 - language model
 - acoustic model



Speech Engine

DEMM54 03 S Brewster ©

Speech-16

- 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



Noise

DEMM54 03 S Brewster ©

Speech-17

- Background noise can have a major impact on recognition rates
- Quality of input
 - speaker phone versus throat mic
 - location
 - noise cancellation



Speech Interface Design

DEMM54 03 S Brewster ©

Speech-18

- Is there a need for speech?
- Base design on how people speak in the application domain
- Feedback
 - speech is slow
 - user needs to know if speech was recognised / heard correctly
 - is system processing data or waiting for input
- Latency
 - time taken to recognise utterance
 - pauses are confusing



Speech Prompts

DEMM54 03 S Brewster ©

Speech-19

- Let user know what to say next
- Explicit prompts
 - tell user exactly what to say
 - Welcome to the Royal Bank of Scotland. Say “check balance”, “transfer funds” or “pay a bill”
- Implicit prompts
 - open ended
 - more natural but allow more potential for error
 - Welcome to the Royal Bank of Scotland. What would you like to do?



Speech Prompts

DEMM54 03 S Brewster ©

Speech-20

- Be brief - speech is slow
- Allow ‘barge-in’
- Can use timeouts
 - Welcome to the Royal Bank of Scotland. What would you like to do?
 - <timeout>
 - Say “check balance”, “transfer funds” or “pay a bill”
- Could use text on screen if available



Recognition Errors

DEMM54 03 S Brewster ©

Speech-21

- 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



Human Issues

DEMM54 03 S Brewster ©

Speech-22

- Listeners will give your interface a personality and respond in human ways
 - The “Eliza” phenomenon
- Do not violate conversational conventions and politeness



Resources

DEMM54 03 S Brewster ©

Speech-23

- <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

DEMM54 03 S Brewster ©

Speech-24

- An introduction to text-to-speech synthesis, Thierry Dutoit
 - in library
- Speech FAQ
 - <http://www.speech.cs.cmu.edu/comp.speech/>



Demos

DEMM54 03 S Brewster ©

Speech-25

- www.genmagic.com/portico/demos
- www.dragonsys.com
- A really good text-to-speech engine:
 - <http://www.research.att.com/projects/tts/>

