Case Studies in Design Informatics 1
and
Case Studies in Design Informatics 2
Jon Oberlander

Lecture 1:
Overview *and*
Introduction to Spoken Dialogue Systems
Slides quote or paraphrase cited papers

http://www.inf.ed.ac.uk/teaching/courses/cdi1/

School of informatics

Structure of lecture

1. Overview of Case Studies Course
   - Goal
   - Structure
   - Assessment
2. Introduction to Voice Interfaces and Dialogue Systems
   - Voice user interfaces and spoken dialogue systems
     - Cohen, Giangola & Balogh (2004)
   - The trouble with speech
     - Shneiderman (2000)

Course Goal

- To address the question:
  How would you do it differently?
- Every time a design decision is made to pursue one course of action, other routes are closed off.
- The goal is:
  - to work in groups to see why specific project design decisions were taken, and
  - to envisage a different service or product that could be built from the same components.

Course Structure

- There are two core case studies:
  - Speech user interfaces and the Mymyradio project
  - Affective computing in the Help4Mood project
- This is the link:
  - How does device design elicit the “correct” human behaviour?
- Between them, there are five student case studies
  - Again, the question is: how can devices elicit “best behaviour”?
- Assessment is by term paper only; there is no final exam.
  - A1 30%: first term paper on Mymyradio (Group).
  - A2 40%: second term paper on student cases (Group).
  - A3 30%: third term paper on Help4Mood (Individual).
- Feedback:
  - Summative:
    - Term papers will be marked, and written feedback given
  - Formative:
    - All tutorials provide formative verbal feedback;
    - 1 tutorial provides formative written feedback on draft A2 reports

Course Assessment

- Assignment 1:
  - Start: Week 2, Monday, 16:00: 22nd September. Available from course page.
  - Submit: Week 4, Thursday, 16:00: 9th October
    - (30% of overall coursework grade).
  - Use Informatics submit, or DVD/thumbdrive to Informatics Teaching Office, Appleton Tower.
  - Return: Week 5, Friday, 16:00: 17th October.
- Assignment 2:
  - Start: Week 5, Monday, 16:00: 13th October. Available from course page.
  - Submit: Week 8, Thursday, 16:00: 6th November
    - (40% of overall coursework grade).
  - Use Informatics submit, or DVD/thumbdrive to Informatics Teaching Office, Appleton Tower.
  - Return: Week 9, Friday, 16:00: 14th November.
- Assignment 3:
  - Start: Week 9, Monday, 16:00: 10th November. Available from course page.
  - Submit: Week 11, Thursday, 16:00: 27th November
    - (30% of overall coursework grade).
  - Use Informatics submit, or DVD/thumbdrive to Informatics Teaching Office, Appleton Tower.
  - Return: Week 13, Friday, 16:00: 12th December.

Course Timetable

Week	Topic	Mon	Wed	Thu	Submit 16:00 Thu
1	SUI	Intro (JO)		Wired for speech (JO)	
2	SUI	Dialogue systems (JO)	Tutorial	Dialogue and error (CM)	
3	SUI	Speech synthesis (MA)	Tutorial	Mymyradio (MA)	
4	SUI	Talk, things & animals (JO)	Tutorial	<No class>	A1
5	ADI	Student cases 1 (TBC)	Tutorial	Student cases 2 (TBC)	
6	ADI	Student cases 3 (TBC)	Tutorial	Student cases 4 (TBC)	A2-draft
7	ADI	Student cases 5 (TBC)	Tutorial	<No class>	
8	AC	Affective computing (JO)	Tutorial	Affective output (JO)	A2
9	AC	Affective input (JO)	Tutorial	Affect in text (CL)	
10	AC	Affective in eyes (RH)	Tutorial	Affective agents (CM)	
11		Reflection (JO)	(Tutorial)		A3

Introduction to Spoken User Interfaces

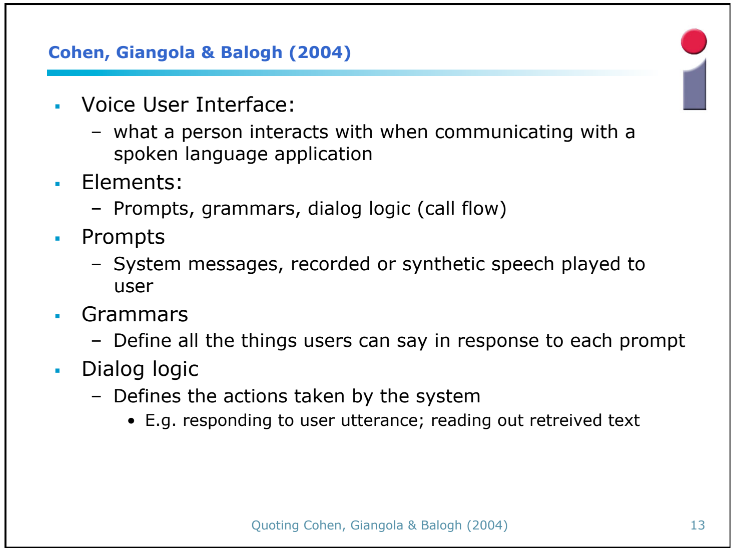
1. Voice user interfaces and spoken dialogue systems
   - Cohen, Giangola & Balogh (2004)
2. The limits of speech recognition
   - Shneiderman (2000)

Cohen, Giangola & Balogh (2004)

- Cohen, M.H, Giangola, J.P., & Balogh, J. (2004).
- *Voice user interface design*.
- New York: Addison-Wesley.

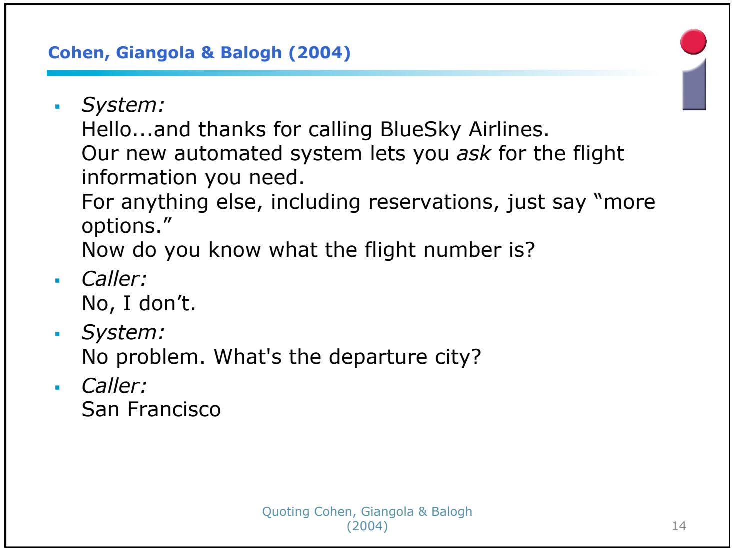
Cohen, Giangola & Balogh (2004)

- Compare Raskin (2000): User-centered design vs human-centered design
- UCD focus: task-related needs of target users in target application
- HCD focus: accordance with “universal psychological facts”
- Voice user interfaces require both



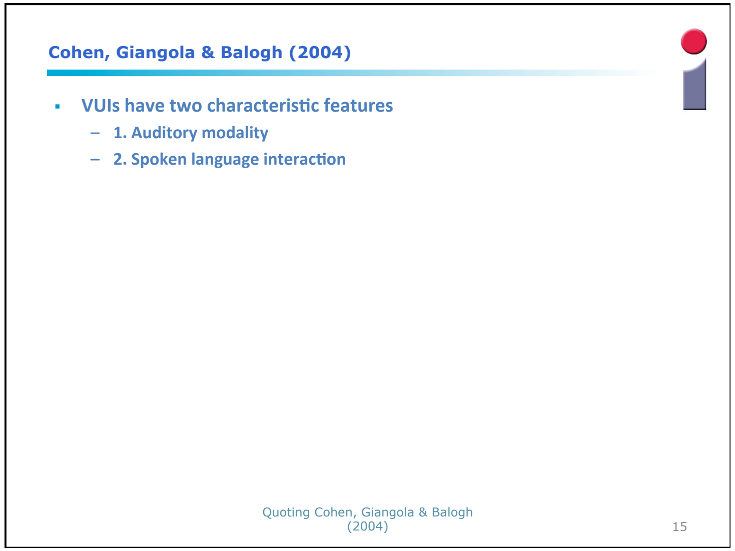
Cohen, Giangola & Balogh (2004)

- Voice User Interface:
  - what a person interacts with when communicating with a spoken language application
- Elements:
  - Prompts, grammars, dialog logic (call flow)
- Prompts
  - System messages, recorded or synthetic speech played to user
- Grammars
  - Define all the things users can say in response to each prompt
- Dialog logic
  - Defines the actions taken by the system
    - E.g. responding to user utterance; reading out retreived text



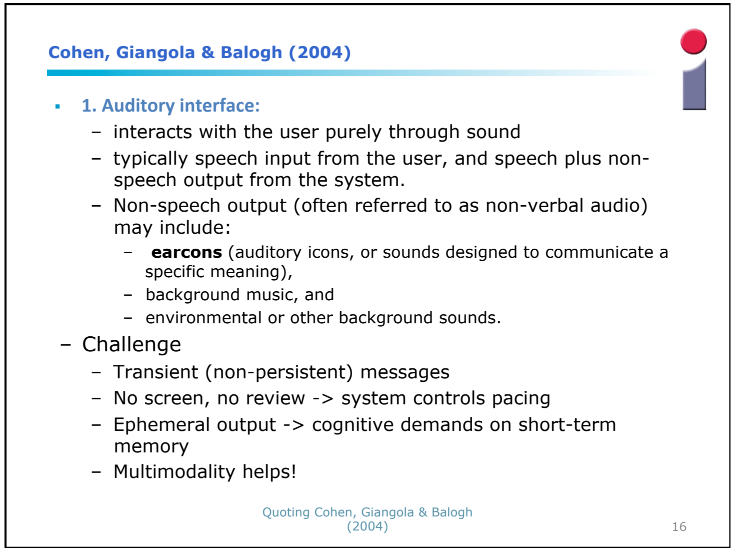
Cohen, Giangola & Balogh (2004)

- *System:*
  Hello...and thanks for calling BlueSky Airlines.
  Our new automated system lets you *ask* for the flight information you need.
  For anything else, including reservations, just say “more options.”
  Now do you know what the flight number is?
- *Caller:*
  No, I don’t.
- *System:*
  No problem. What's the departure city?
- *Caller:*
  San Francisco



Cohen, Giangola & Balogh (2004)

- **VUIs have two characteristic features**
  - **1. Auditory modality**
  - **2. Spoken language interaction**



Cohen, Giangola & Balogh (2004)

- **1. Auditory interface:**
  - interacts with the user purely through sound
  - typically speech input from the user, and speech plus non-speech output from the system.
  - Non-speech output (often referred to as non-verbal audio) may include:
    - **earcons** (auditory icons, or sounds designed to communicate a specific meaning),
    - background music, and
    - environmental or other background sounds.
- Challenge
  - Transient (non-persistent) messages
  - No screen, no review -> system controls pacing
  - Ephemeral output -> cognitive demands on short-term memory
  - Multimodality helps!

Cohen, Giangola & Balogh (2004)

- Opportunities:
  - Speech carries information via word and sentence structure …
  - And prosody, voice quality -> persona, branding
  - Non verbal audio can deliver information without breaking flow
  - Can improve navigation via sound landmarks
  - Can create “sound and feel” via msuic and natural sounds

Cohen, Giangola & Balogh (2004)

- 2. Spoken language:
  - Very natural, nearly universal
  - Shared expectations power effective human-human communication
  - Challenges
    - Violated expectations reduce comfort, disrupt flow, disrupt understanding, introduce error
    - Spoken language is learned long before your interface:
      - “As designers, we don’t get to create the underlying elements of conversation”
    - Conventions are implicit – designers must surface the conventions, in context.

Cohen, Giangola & Balogh (2004)

- Why do it?
  - **Save money:** The ROI (return on investment) for speech systems deployment is typically on the order of a few months.
  - **Improve reach:** Companies want to be available to their customers in all places (home and mobile) at all times (24x7x365).
  - **Extend brand:** When you engage in spoken interaction, you don’t get “just the facts.” Speech communicates at many levels. … we can design the “ideal employee” – with the right voice, the right personality traits, the right mood, and the right way of handling customer needs and problems.
  - **Solve new problems:** There are many instances of problems that can be solved, or services that can be offered, by speech applications that were simply impossible in the past. e.g. Call routing, personal agents
  - **Increase customer satisfaction:** Numerous surveys and deployment studies have shown high user satisfaction with speech systems.

Cohen, Giangola & Balogh (2004)

- For the end-user … speech systems, are:
- Intuitive and efficient:
  - Spoken language systems draw on the user’s innate language skills. Many tasks can be made simpler and more efficient than with touch-tones.
  - For example, in a travel application a caller may say things like “I wanna leave on June 5” rather than entering some awkward and unintuitive touchtone sequence (such as 0605) after hearing some longwinded instruction.
- Ubiquitous
- Enjoyable:
  - A well-designed system be engaging and enjoyable
- Hands-free, eyes free:
  - Activities such as driving occupy the user’s hands and eyes.
  - Speech is an ideal solution for accessing services while engaged in hands- or eyes-busy tasks.

Cohen, Giangola & Balogh (2004)

- Basic architecture of a spoken language system

1. Endpointing:
   - Detect onset and offset of user speech (slice up waveform)
2. Feature extraction:
   - Transform waveform into sequence of feature vectors (e.g. amount of energy at various frequencies), one vector per time period (e.g. 10ms)
3. Recognition:
   - Match feature vector sequence against most likely word sequence
4. Natural language understanding
   - Extract meaning from word sequence (e.g. fill slots with values)
5. Dialog management:
   - Select next system action (e.g. check database; speak to user; book flight)

Quoting Cohen, Giangola & Balogh (2004)

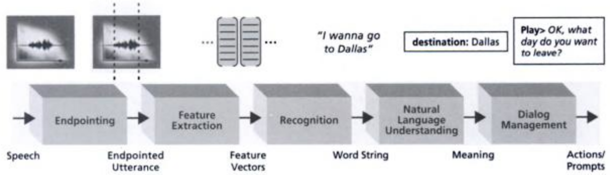


FIGURE 2-7 *The processing sequence for handling one spoken input from a caller.*

Quoting Cohen, Giangola & Balogh (2004)

Cohen, Giangola & Balogh (2004)

- Recognition requires:
- Acoustic models
  - Represent (statistically) how each phoneme (basic sound) may be pronounced
- Dictionary
  - Lists words with their (multiple) pronunciations: which acoustic models sequence into which word models
- Grammar
  - Defines all possible user inputs (word strings and meanings)
  - Rule based or statistical (SLMs are more flexible; require training)
- Recognition model
  - A search to find best-matching path(s) through the lattice
  - Returns a confidence measure (match score between input vectors and best path)

Quoting Cohen, Giangola & Balogh (2004)

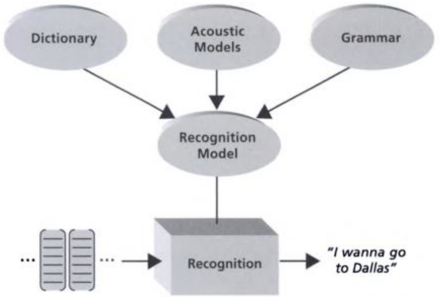
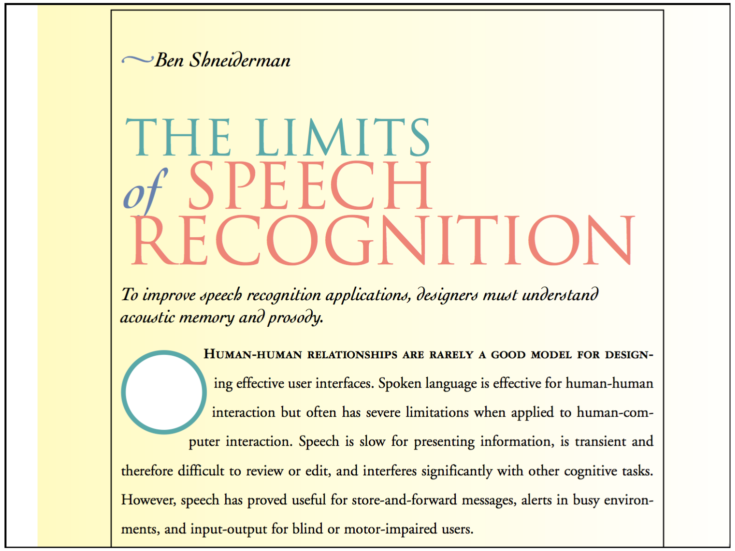


FIGURE 2-8 *The recognizer searches the recognition model to find the best-matching word string. The recognition model is built from the acoustic models, dictionary, and grammar.*

Quoting Cohen, Giangola & Balogh (2004)



~Ben Shneiderman

THE LIMITS *of* SPEECH RECOGNITION

To improve speech recognition applications, designers must understand acoustic memory and prosody.

Human-human relationships are rarely a good model for designing effective user interfaces. Spoken language is effective for human-human interaction but often has severe limitations when applied to human-computer interaction. Speech is slow for presenting information, is transient and therefore difficult to review or edit, and interferes significantly with other cognitive tasks. However, speech has proved useful for store-and-forward messages, alerts in busy environments, and input-output for blind or motor-impaired users.

Shneiderman (2000)

- Shneiderman, B. (2000).
- The limits of speech recognition.
- *Communications of the ACM, 43(9), 63-65.*

Shneiderman (2000)

- **Human-human relationships are rarely a good model for designing effective user interfaces.**
- Spoken language is effective for human-human interaction but often has severe limitations when applied to human-computer interaction.
- Speech is slow for presenting information, is transient and therefore difficult to review or edit, and interferes significantly with other cognitive tasks.
- However, speech has proved useful for store-and-forward messages, alerts in busy environments, and input-output for blind or motor-impaired users.

Shneiderman (2000)

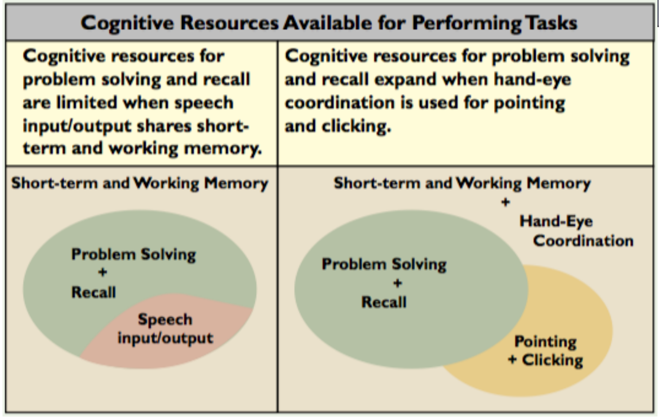
- Speech recognition and generation is sometimes helpful for environments that are hands-busy, eyes-busy, mobility-required, or hostile and shows promise for telephone-based services.
  - Telephone-based speech-recognition applications, such as voice dialing, directory search, banking, and airline reservations,
    - may be useful complements to graphical user interfaces.
  - Dictation input is increasingly accurate, but adoption outside the disabled-user community has been slow compared to visual interfaces.
    - Obvious physical problems include fatigue from speaking continuously and the disruption in an office filled with people speaking.

Shneiderman (2000)

- By understanding the cognitive processes surrounding human “acoustic memory” and processing,
  - interface designers may be able to integrate speech more effectively and guide users more successfully.
- By appreciating the differences between human-human interaction and human-computer interaction,
  - designers may then be able to choose appropriate applications for human use of speech with computers.

Shneiderman (2000)

- Now consider human acoustic memory and processing.
- Short-term and working memory are sometimes called acoustic or verbal memory.
- The part of the human brain that transiently holds chunks of information and solves problems also supports speaking and listening.
- Therefore, working on tough problems is best done in quiet environments:
  - without speaking or listening to someone.
- However, because physical activity is handled in another part of the brain, problem solving is compatible with routine physical activities like walking and driving.



Cognitive Resources Available for Performing Tasks

Cognitive resources for problem solving and recall are limited when speech input/output shares short-term and working memory.	Cognitive resources for problem solving and recall expand when hand-eye coordination is used for pointing and clicking.
Short-term and Working Memory	Short-term and Working Memory + Hand-Eye Coordination

Shneiderman (2000)

- In short, humans speak and walk easily but find it more difficult to speak and think at the same time.
  - Similarly when operating a computer, most humans type (or move a mouse) and think but find it more difficult to speak and think at the same time.
- Hand-eye coordination is accomplished in different brain structures, so typing or mouse movement can be performed in parallel with problem solving.
  - Since speaking consumes precious cognitive resources, it is difficult to solve problems at the same time.
  - Proficient keyboard users can have higher levels of parallelism in problem solving while performing data entry.
- This may explain why after 30 years of ambitious attempts to provide military pilots with speech recognition in cockpits, aircraft designers persist in using hand-input devices and visual displays.

Shneiderman (2000)

- The problem of emotive prosody
- The interfering effects of acoustic processing are a limiting factor for designers of speech recognition, but the role of emotive prosody raises further concerns.
- The human voice has evolved remarkably well to support human-human interaction.
  - A military commander may bark commands at troops, but there is as much motivational force in the tone as there is information in the words.
- Loudly barking commands at a computer is not likely to force it to shorten its response time or retract a dialogue box.

Quoting Shneiderman (2000) 33

Shneiderman (2000)

- Promoters of "affective" computing, or reorganizing, responding to, and making emotional displays,
  - may recommend such strategies,
  - though this approach seems misguided.
- Many users might want shorter response times without having to work themselves into a mood of impatience.
- Secondly, the logic of computing requires a user response to a dialogue box independent of the user's mood.
- And thirdly, the uncertainty of machine recognition could undermine the positive effects of user control and interface predictability.

Quoting Shneiderman (2000) 34

Shneiderman (2000)

- Human emotional expression is so
  - *varied* (across individuals),
  - *nuanced* (subtly combining anger, frustration, impatience, and more), and
  - *situated* (contextually influenced in uncountable ways)
- that accurate simulation or recognition of emotional states is usually impractical.
- For routine tasks with limited vocabulary and constrained semantics, such as order entry and bank transfers,
  - the absence of prosody enables limited successes,
  - though visual alternatives may be more effective.
- Speech systems founder when designers attempt to model or recognize complex human behaviors.
  - Comforting bedside manner, trusted friendships, and inspirational leadership are components of human-human relationships not amenable to building into machines.

Quoting Shneiderman (2000) 35

Taking stock

- Shneiderman's view has been very influential in the HCI community
  - Speech is seen as having significant problems, and only niche applications
- Now, voice user interfaces – and spoken dialog systems – are becoming more common
- But they still raise quite specific design challenges
- Looking ahead:
  - We're going to look some more at speech output, dialog systems, and more-or-less applications built on in these foundations

36