The Design of Interactive Computational Media

Class 9: 11 Mar. 2003

## Interaction Design Hour 3: Audio Interaction

## System and Interface Evaluation Hour 1: User Testing



## Interaction Paradigms and Styles (Review)

- Simple command languages
- Query languages, conversational languages
- Natural language input
- Menus
- Forms
- Icons
- Windows
- Direct manipulation
- Gestural interaction
- 3D interaction
- (now) Audio I/O, including voice and non-speech audio



## Audio Interaction

- Speech as data
- Input versus speech recognition
- Output versus speech synthesis
- Non-speech audio output

- Multimodal interaction



## Speech as Data

- Example: Voice messaging as in Olympic Message System
- Example: Voice annotations as in Wang Freestyle



## Input via Speech Recognition

- Ideal applications
  - Hands busy or covered in “gunk”
  - Manual input already overloaded
  - Disabled users
- Simplified method of operation
  - Recognition vocabulary represented as stored patterns
  - Speech sampled and digitized
  - Waveforms or their parameters compared against patterns



## Speech Recognition: Dimensions of Success

- Size of vocabulary: A few words to tens of thousands
- Accuracy, recognition percentage
  - >>95%, >99%, around 75 or 80%
- Repeatability of performance
- Cost
- Speaker-dependent vs. speaker-independent
- Training not required or easily trainable
- Location of microphone
- Acoustic environment, quiet or noisy environment
- Discrete words or continuous speech

## Speech Recognition Example

- *VIDEO — The OM System Spoken Language Interface (CMU, SGVR 64, CHI'91)*
  - Methods of user error correction
  - Recognition architecture
  - Usage of lexical and syntactic information (certain words and sentence structure are legal and therefore expected)

QuickTime™ and a Sorenson Video decompressor are needed to see this picture.



## Output via Speech Synthesis

- Why is the problem hard? Examples:
  - How to pronounce "gh"?
    - No sound in "thorough"
    - "f" in "enough"
    - "g" in "ghost"
  - How to pronounce "invalid"?
    - Not valid ==> Accent on second syllable
    - Disabled ==> Accent on first syllable
  - How to stress (intonation)?
    - "I told you" meaning different depending upon word that is stressed
  - Method of operation
    - Digitized (stored) versus synthesized speech
    - Synthesize speech: Phoneme-to-speech, text (ASCII)-to-speech
    - Retrieve or generate waveform, convert to analog, output



## Speech Synthesis: Dimensions of Success

- Size of vocabulary
- Bandwidth, data rate
- Intelligibility
- Cost
- Naturalness
- Discrete words versus connected speech



## Speech Synthesis Example

- *VIDEO — Talking to Machines (University of Wales, SGVR 88, InterCHI'93)*
  - Speech input and output
  - Consequences of failure to anticipate user errors
  - Example of a real application
  - Design "principles"



## Roles for Auditory Displays

- Alarms and warnings
- Status and monitoring indicators
  - e.g., feedback from control inputs
- Messages and data (perhaps encoded)
  - e.g., responses to user queries



## Visual versus Auditory Displays

- From Deatheridge, 1972, p. 124, in BGBG, p. 532

| *Use auditory presentation if:* | *Use visual presentation if:* |
|---|---|
| Message is simple | Message is complex |
| Message is short | Message is long |
| Message will not be referred to later | Message will be referred to later |
| Message deals with events in time | Message deals with location in space |
| Message calls for immediate action | Message does not call for immediate action |
| Visual system of person overburdened | Auditory system of person overburdened |
| Location too bright or dark | Location too noisy |
| Person moving continually | Person stationary |

## Non-speech Audio Output

- Motivation
  - Consider role of sound in video games, driving
  - Warnings (e.g., sound of blowout)
  - Status indicators (e.g., revving engine)
  - Feedback (e.g., grinding gears)
- *VIDEO: Example of Communicative Sounds (NCSA)*

QuickTime™ and a Cinepak decompressor are needed to see this picture.



## Non-speech Audio

- Example — Sonic Finder (Gaver, UCSD, mid-80s)
  - Hear the trash can through a “tinny crash”
  - Hear amount of space on disk through reverberation
  - Hear status of scrolling through ascending or descending tones
- Issues
  - Appropriate acoustic design
  - Storage requirements or real-time processing
  - Acoustic pollution



## Examples of Non-speech Audio

- *VIDEO — LogoMedia (DiGiano, Baecker, UofT, 1993)*
  - Use of sound in software visualization
- *VIDEO — Arkola Simulation (Smith et al., Xerox EuroPARC, early 90s)*
  - Sound for understanding complex system, collaborating

QuickTime™ and a Sorenson Video decompressor are needed to see this picture.



## Multi-modal Interaction

- *VIDEO — Put That There (MIT, CHI’83 Videos)*

QuickTime™ and a Sorenson Video decompressor are needed to see this picture.



## Other Advances

- Word spotting for speech skimming
- Speeding up digitized speech output
- Uses of “good enough” recognition



## *Questions and Discussion*

## *Break*



## User Testing

- Usefulness and usability
- User testing in the design process
- Empirical evaluation more generally
- Observing scenarios and prototypes
- How to do user testing
- Thinking aloud
- Data capture and analysis
- Asking users as well as testing them
- Ethical issues

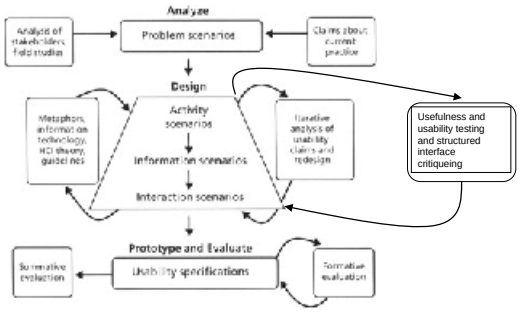


## Usefulness and Usability

- *Usefulness* refers to the extent to which the system's functionality meets the needs of the user and of the tasks he or she is carrying out.
- *Usability* refers to the degree to which the interface facilitates carrying out of the task, and incorporates such criteria as ease of learning, ease of use, protection against catastrophic errors, and provision of user support.



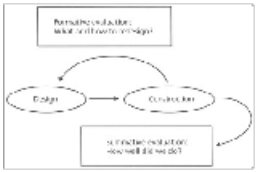
## Rosson and Carroll Design Method






## User Testing in the Design Process

- Empirical evaluation can happen at every stage
- Formative evaluation
  - Happens throughout the design process
  - Can evaluate scenarios, sketches, models, prototypes
- Summative evaluation
  - Typically happens at the end
  - Assesses system and interface design quality, i.e., how well have we done?





## Goals for Empirical Evaluation

- Understanding what happens in practice
  - Interface quality
  - Interface problems
  - What factors seem to affect quality — e.g., metaphors, mental models, look & feel, documentation
- Understanding why it happens
  - Often users will tell us why, or we can deduce it from observations and analysis
  - We also build models to achieve a deeper understanding of casual factors (CSC428)
  - Since this has been demonstrated to be cost-effective in only scattered cases, we usually rely on empirical approaches

## Four Dimensions of Evaluation
(Ray and Ravizza, 1985)

- Methodology
  - Naturalistic observation (watching and recording)
  - True experiments (manipulating and measuring)
  - Actually, extremes with points in between
- Setting — Field research or laboratory research
- Experimental role — Scientist as participant or observer
- Size: one, few, or many subjects



## Naturalistic Observation vs. True Experiments

| Naturalistic observation | True experiments |
|---|---|
| Noninterference with phenomena | Manipulation, control |
| Observations of patterns and invariants | Measurements of observed patterns |
| Big picture insights | Detailed results |
| Qualitative, descriptive | Quantitative |

- Will talk about experiments next week



## Observing Scenarios and Prototypes

- Method
  - Produce facsimile of interface
    - Construct scenario — words, pictures, animation, video
    - Build physical mockups
    - "Program" interactive prototypes
    - Write manual in advance of system
  - Design and produce situation
  - Observe (or test) behaviour of users
- Roles
  - Elicits initial reactions to, problems with user's model, interface
  - Engages potential users, gets them excited and involved, makes them see that they can contribute



## Observing Scenarios and Prototypes

- Advantages
  - Can be done in advance of actual system construction
- Disadvantages
  - Facsimile will not embody all characteristics of system
- Examples
  - Iterative design of CHI'89 Info Kiosk (Salomon, BGBG, 23-34)
  - Wizard of Oz simulation of an intelligent adaptive interface to Microsoft Word (McGrenere, 2002)



## User Testing

- Methods
  - Design and implement scenario or prototype
  - Record user behaviour
    - Typical usage, or critical incidents
    - Keystroke recording
    - Thinking aloud protocols
    - Videotape protocols
  - Interviews for subjective impressions
  - Analyze user behaviour
- Roles
    - Understanding user methods
    - Understanding user problems
    - Discovering user thought processes



## User Testing

- Advantages
  - Observation of real usage of real system
- Disadvantages
  - Individuals often know only limited aspects of a system
  - Usually done in a lab on "toy problems" with invented data
- Examples
  - Mack, et al. studies of text editor learning (BGBG, Ch. 10)
  - Usability tests of SASE, SASSE tools for collaborative writing (BGBG, Ch. 11)

## Typical Steps in User Testing (Gomoll, in Laurel, 85-90)

- 1. Set up the observation
- 2. Describe the purpose of the observation
- 3. Tell the user that it's OK to quit at any time
- 4. Talk about and demonstrate the equipment in the room
- 5. Explain how to "think aloud"
- 6. Explain that you will not provide help
- 7. Describe the task and introduce the system
- 8. Ask if there are questions before you start; then begin observation
- 9. Conclude the observation and debrief (interview) subjects
- 10. Analyze the data
- 11. Tabulate the results
- 12. Interpret the results in the context of other results



## User Testing (BGBG, Fig. 2.8, p. 85, adapted from Neilsen, 1992)

- Practical study design
  - Reflect on the participants' backgrounds and how they might affect the study
  - Be aware of problems that arise when experimenters know the users personally
  - Prepare for the study carefully (avoid last minute panic)
  - Select the tasks carefully to be representative and to fit the allotted time
  - In general, start with an easier (but not frivolous) task
  - Write down features of system *not* being tested as well as those that are!
  - Define the start-up state for the study precisely
  - Define precise rules for when and how users can be helped during the study
  - Plan timing and cut-off procedure (if subject gets stuck) for each part of study
  - Include provisions for data collection (e.g., audio, video, or keystroke capture)
  - Plan data analysis techniques in advance
  - Carry out a pilot study (important but often overlooked)
- Written materials
  - Participant release form
  - Questionnaire covering prior experience etc. (if relevant)
  - Introduction to the study for users, including scenario of use
  - Checklist for experimenters
  - Evaluation survey (if relevant)



## User Testing (BGBG, Fig. 2.8, p. 85, adapted from Neilsen, 1992)

- Carrying out the study
  - Let users know that complete anonymity will be preserved
  - Let them know that they may quit at any time
  - Stress that the system is being tested, not the participant
  - Indicate that you are only interested in their thoughts relevant to the system
  - Demonstrate the thinking-aloud method by acting it out for a simple task, e.g., figuring out how to load a stapler
  - Hand out instructions for each part of the study individually, not all at once
  - Maintain a relaxed environment free of interruptions
  - Occasionally encourage users to talk if they grow silent
  - If users ask questions, try to get them to talk (e.g., "What do you think is going on?" and follow predefined rules on when to help or interrupt to help.
  - Debrief each user after the experiment
- Improving the study
  - The pilot study should "debug" the study. This minimize changes during the study, allowing quantitative data analysis. But improvements may be warranted.
  - Experimenters' role can be improved
  - Tasks given to participant can be improved
  - Written materials can be improved



## Thinking Aloud

- Attempt to elicit thought processes of subject, thereby yielding valuable insights (although process is slowed down and may be changed)
- Subject talking while they are doing
  - Problems they are having
  - Solutions they are considering
  - Why they are having trouble
  - Insights that they have
  - Wishes that they have
- Pairs of subjects conversing (Co-Discovery Learning, Kennedy paper in BGBG, pp. 182-185)



## Data Capture and Analysis

- Keystroke logging (sometimes known as "dribble files")
  - Record precise user behaviour
  - Record times to carry out actions
  - Record user errors
- Observation and notetaking by observers, especially of user problems and critical incidents
- Audio and video recordings
  - Can't observe and record all behaviour in real-time
  - Preserve behaviour for review
  - Non-verbal behaviour
  - Behaviour in context



## Data Capture and Analysis

- Data extraction and analysis
  - Enormous amounts of "data" in video tape — how to convert to information?
  - Review of video & audio tapes, video annotation & analysis
  - Classifying events and counting events — construction of *analysis protocols*
- Interaction analysis (see also Lecture 3)
  - Interaction analysis..."to uncover the regularity and efficacy of peoples' relations with each other and their use of the resources that their environment affords"
  - Typically used in ethnographic observations of real work contexts
  - Can also apply to analysis of user testing or other lab studies

## Asking Users in Addition to Observing Them

- Methods
  - *Questionnaire* design
    - Formulating & asking questions, & analyzing answers
    - Hard to avoid bias in the phrasing of questions
    - Therefore requires pre-testing ("pilot testing")
  - *Surveys* — (possibly large-scale) administration of questionnaires to appropriate *samples* of individuals chosen from a *population*
  - Administration of questions through *interviews*
- Roles
  - Relating user behaviour to user needs
  - Carrying out a task analysis
  - Understanding the work context



## Asking Users in Addition to Observing Them

- Advantages
  - Easy to administer, efficient time-wise
  - Compared to surveys, interviews can be free-ranging, adapting to conversation flow, yielding novel insights
- Disadvantages
  - Subjects may not be able to conceptualize new technology
  - Abstract, may not be well grounded in real application
  - Danger of bias – Putting words into people's mouths
- Examples
  - Gould & Boies interviews prior to design of Speech Filing Syst.
  - Posner's collaborative writing interviews (Baecker et al.,BGBG, 775-782)
  - McGrenere's interviews with users of Word and Word Personal



## Ethical Issues

- Basic principles
  - Do no harm
  - Voluntary participation
  - Informed consent
  - Right to privacy
- Use of research protocols and consent forms
  - Explanation of study and purpose
  - Anonymity
  - Ability to withdraw at any time
  - For example, see p. 256 of Rosson & Carroll
  - As discussed earlier in term
- Difficult issue — Uses of video data



## *Questions and Discussion*