Chapter 9

Interaction Devices

9.4 Speech & Auditory Interfaces
9.5 Displays

Auditory interfaces

• Audio tones, audiolization, and music
  – Sound feedback can be important:
    • to confirm actions
    • offer warning
    • for visually-impaired users
    • music used to provide mood context, e.g. in games
    • can provide unique opportunities for user, e.g. with simulating various musical instruments

Speech Interfaces

• Human Language Processing
  – Stages
    • Speech Recognition
    • Natural Language Understanding (NLP)
    • Natural Language Production (NLP)
    • Speech Synthesis
  – Speech processing totally separate from NLP
  – Speech recognition and NL understanding much harder than NL production and speech synthesis
  – Speech processing is usually real-time interaction
  – NLP is usually text processing and not real-time interaction
Speech Recognition- A very hard problem!

“We saw you a to-ah.” “I know your mom.”

Speech recognition

- Speech recognition still does not match the fantasy of science fiction:
  - Only successful for limited vocabulary tasks with acceptable failures
  - Most suitable when hands cannot be used
  - Voice-controlled editor versus keyboard editor
    - lower task-completion rate
    - lower error rate
  - May be disturbing in some environments
  - Does not require natural language systems
  - Most useful in specific applications, such as to benefit handicapped users

Speech recognition

- Dimensions of difficulty
  - Isolated (discrete) words vs. continuous speech
    - Discrete: 95- to 98-percent reliability for 20 to 200 word vocabularies
    - Continuous
      - Difficulty in recognizing boundaries between spoken words
      - Words acoustically confusable
        - “see care you a prin’”
        - “recognize speech” or “smack a nice beach”
  - Vocabulary size
    - Search increases exponentially with vocabulary size
  - Speaker dependent vs. independent
    - Speaker dependent must be trained: go through vocabulary twice
    - Speaker independent: very limited application
  - Noisy environment
Speech Processing

• Stored Speech systems
• Speech Synthesis systems

Stored Speech synthesis

• Voice information systems
  – Stored speech commonly used to provide information about tourist sites, government services, after-hours messages for organizations
  – Low cost
  – Voice prompts
  – Deep and complex menus frustrating
  – Slow pace of voice output, ephemeral nature of speech, scanning and searching problems
• Applications
  • Voice mail
  • Handheld voice recorders
  • Audio books
  • Instructional systems

Speech Synthesis

• Converts text to language sounds (phonemes)
• Can choose pitch, speed, type of voice
• Does not handle continuous speech well
  – Conversion is one word to a sequence of sounds
  – Lacks cadence
  – Lacks emphasis in loudness and speed of delivery
Speech synthesis

- Speech generation
  - Michaelis and Wiggins (1982) suggest that speech generation is "frequently preferable" under these circumstances:
    - The message is simple.
    - The message is short.
    - The message will not be referred to later.
    - The message deals with events in time.
    - The message requires an immediate response.
    - The visual channels of communication are overloaded.
    - The environment is too brightly lit, too poorly lit, subject to severe vibration, or otherwise unsuitable for transmission of visual information.
    - The user must be free to move around.
    - The user is subjected to high G forces or anoxia.

Displays

- The display has become the primary source of feedback to the user from the computer
  - The display has many important features, including:
    - Physical dimensions (usually the diagonal dimension and depth)
    - Resolution (the number of pixels available)
    - Number of available colors, color correctness
    - Luminance, contrast, and glare
    - Power consumption
    - Refresh rates (sufficient to allow animation and video)
    - Cost
    - Reliability
Displays

Human Factors Issues:
• Realism and Quality (Psychophysics)
• Portability
• Privacy
• Simultaneity (Screen Real Estate)

Novel Display technology

• Electronic ink
  – Paper like resolution
  – Tiny capsules with negatively and positively charged particles
• Braille displays
  – Provides output for the blind
• 3D Display with Stereo Glasses
  – Two images displayed, one for each eye
  – Depends on brain of viewer to “fuse” the image as 3D (depth)

3D Display with Stereo Glasses
Displays – Large

- Large displays
  - Multiple desktop displays
  - Informational wall displays
  - Interactive wall displays

Multi Display (Desktop)

Multiple Displays (Avionics)
Princeton Wall Display

“Walkaround” display with Stereo Glasses

Displays Head Mounted

• Heads-up and helmet mounted displays
  – A heads-up display can, for instance, project information on a partially silvered widescreen of an airplane or car
  – A helmet/head mounted display (HMD) moves the image with the user
  – 3D images
Head Mounted Display

Head Mounted Display & Data Glove (Virtual Reality)

Small displays (Mobile phone)

- 640 x 480 is large display!
  - Custom designs to take advantage of every pixel!
  - Okay for linear reading, but making comparisons can be difficult
- Currently mobile devices used for brief tasks, except for game playing
- Multi-media (and function)
  - Camera phones
  - MP3 players
  - Web browsing difficult
- Optimize for repetitive tasks
### Animation, image, and video

- Accelerated graphics hardware
- More information shared and downloaded on the web
- Scanning of images and OCR
- Digital video
- CDROMS and DVDs
- Compression and decompression through MPEG
- Computer-based video conferencing