cisc3650
human-computer interaction
spring 2012
lecture # IV.2
multimodal interfaces

topics:

- multimodal interfaces

references:

- *Multimodal interfaces that flex, adapt, and persist*, by Sharon Oviatt, Trevor Darrell, Myron Flickner, Communications of the ACM, volume 47, number 1, January 2004.
- *Guidelines for multimodal user interface design*, by Leah M. Reeves, Jennifer Lai, James A. Larson, Sharon Oviatt, T. S. Balaji, St'ephanie Buisine, Penny Collings, Phil Cohen, Ben Kraal, Jean-Claude Martin, Michael McTear, TV Raman, Kay M. Stanney, Hui Su, Qian Ying Wang, Communications of the ACM, volume 47, number 1, January 2004.

multimodal interfaces

- multimodal interfaces are designed for:
  - compatability with users' work practices
  - flexibility
- design criteria:
  - *robustness* increases as the number and heterogeneity of modalities increase
  - *performance* improves with adaptivity of interface
  - *persistence* of operation despite physical damage, loss of power, etc.
- classes of multimodal interfaces
  - speech with manual (pen or touch) input
  - audiovisual speech with lip movement
  - multibiometric input (e.g., speech, vision, physiological inputs)
- technologies:
  - computer speech
  - vision processing

- multimodal $\rightarrow$ hybrid
- multimodal communication parallels human's natural modes of communicating
- natural structures are generally more flexible than engineered structures
- adaptivity: interfaces learn from users and environments
- *M3 systems*: Multibiometric, Multimodal, Multisensor
- big challenges:
  - natural language communication
  - vision processing
  - multimodal/sensory fusion

guidelines for multimodal user interface design

- two goals:
  1. achieve more natural interaction, like human-human interaction
  2. increase robustness by providing redundant and complementary information
- six guidelines:
  1. requirements specifications
     - design for broad range of *users* (experience, abilities, etc.) and *contexts* (home, office, changing environments like car)
     - address privacy and security issues
       - don't remember users by default
       - use non-speech input for private information, like passwords
  2. designing multimodal input and output
     - guidelines stem from *cognitive science*:
       - maximize human cognitive and physical abilities
         - e.g., don't require paying attention to two things at once
         - reduce memory load

        - multiple modes should complement each other, enhance each other
      * integrate modalities to be compatible with user preferences, context and system functionality
        - e.g., match input and output styles
        - use multimodal cues, e.g., look at speaker
        - synchronize modalities (timing)
        - synchronize system state across modalities
3. adaptivity
   - adapt to needs/experiences/skill levels of different users and contexts
   - examples:
     * gestures replace sounds in noisy settings
     * accommodate for slow bandwidth
     * adapt quantity and stye of information display based on user's perceived skill level
4. consistency
   - use same language/keywords for all modalities
   - use same interaction shortcuts for all modalities
   - support both user and system switching between modalities
5. feedback

   - users should know what the current modality is and what other modalities are available
   - avoid lengthy instructions
   - use common icons, simple instructions and labels
   - confirm system interpretation of user's commands, after fusion of all input modalities has completed
6. error preventing and handling
   - clearly mark "exits" from:
     * task
     * modality
     * system
   - support "undo"
   - include help
   - integrate complementary modalities to improve robustness: strengths of one modality should overcome weaknesses of others
   - let users control modality selection
   - use rich modalities that can convey semantic information beyond simple point-and-click
   - fuse information from multiple sources