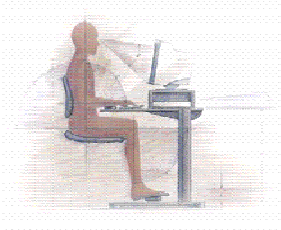
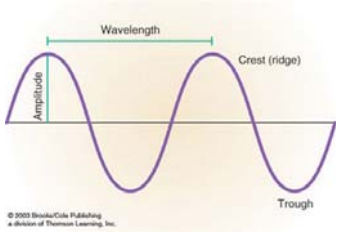
CMPE 233: Human Factors



Auditory and Tactile Systems

Sound Waves

- sound waves are created by a vibrating object
  - any type of molecules (gas, liquid, solid) that can move and create a pressure wave can produce sound
- waves received by ears, transduced into neural signals
- amplitude corresponds to volume
  - Larger/smaller amplitude = louder/softer sound
- wavelength corresponds to pitch
  - Longer/shorter wavelength = lower/higher pitch



Sound Waves

- Pitch
  - 20Hz – 15KHz, tuned to 3Khz by shape of outer ear
  - Human is less accurate in distinguishing high frequencies than low ones
- Timbre
  - 'signature' of sound source
  - complex set of resonance overlaying the fundamental frequency
- Amplitude and loudness
  - *Loudness* is a psychological property of sound
  - Our ears are capable to cope with 0 to 160db (pain at 130db!)
- Spatialisation – positioning of a sound in 2D or 3D space

Sound Waves

- sound intensity are measured in *dB*
  - logarithmic measure of the volume of different stimuli as compared to a reference point
- threshold – ambient sound intensity above which sounds stand out
- prolonged exposure above 85 dB can cause hearing damage → noise-induced hearing loss (NIHL)
- Types of auditory 'damage'
  - *nerve damage*: occurs when the hair cells are destroyed by loud sounds
  - *conduction damage*: physical damage of the outer or middle ear, e.g. broken eardrum

Hearing Impairment

- 28+M Americans have hearing problems
  - can be inherited or acquired
  - more than 30 genes have been linked to deafness
- Marginal, mild, and moderate losses: 2-60 dB loss
- Profoundly impaired/deaf: 60-75 dB loss in hearing capacity in the better ear
- Causes: 50-75% prenatal, 10-20% perinatal (rubella), 20-30% postnatal (aging)
- Presbyacusis: aging-related progressive hearing loss of higher frequency, more common in men
  - Check mosquito ringtone → annoying for under 30
- *Tinnitus*: constant ringing or noise in the ears
  - Causes: nerve damage, ear infections, even aspirin!

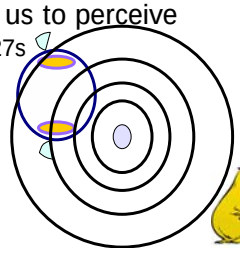
Auditory Transduction

- Two explanations of how basilar membrane converts pressure waves to perceived sound (i.e. how we perceive pitch):
  1. *place theory*: different frequencies activate different parts of the basilar membrane
  2. *frequency theory*: higher frequencies = greater neural firing

  But neurons can fire, at most, 1000 times per second. How do we hear sounds that are at a much greater frequency? (e.g. the upper third of a piano's keyboard) → *volley principle*
- primary auditory cortex in temporal lobe
  - different pitches registered by different neurons within auditory cortex (like feature detectors in vision)

Locating Sounds

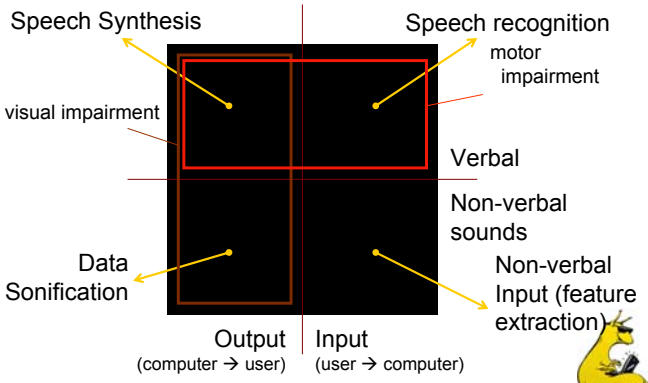
- two ears work together to locate the source
  1. difference in phase: sound waves reach ears at slightly different points in wave cycle
  2. difference in loudness: ear closer to sound source registers louder signal
  3. difference in onset: ear closer to sound source registers signal slightly sooner
- tiny differences, but enough for us to perceive
  - e.g. difference in onset of 0.000027s can be distinguished



Hearing Without Awareness

- *cocktail party phenomenon*: the effect of not being aware of the content of other people's conversations until your name is mentioned
  - processing signal bottom-up, but top-down awareness isn't drawn until self-relevant information is introduced
- *dichotic listening task*: different stimuli delivered to two ears via headphones; instructed to monitor only one signal
  - still perceive some information (e.g. speaker's gender) from ignored ear (Treisman, 1964)

Interacting with Sounds

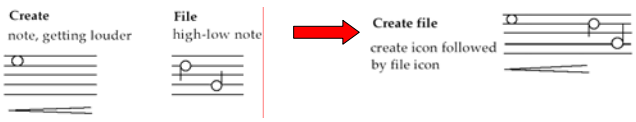


Non-Speech Outputs

- Language independent
- Evidence to show they are useful
  - fewer typing mistakes with key clicks
  - video games harder without sound
- Especially good for transient/background status info
- Use of stereo allows positional information → sonification
- Auditory icons
  - Natural sounds with associated semantics which can be mapped onto similar meanings in the interaction, e.g. throwing something away → the sound of smashing glass
  - Problem: not all things have associated meanings

Non-Speech Outputs

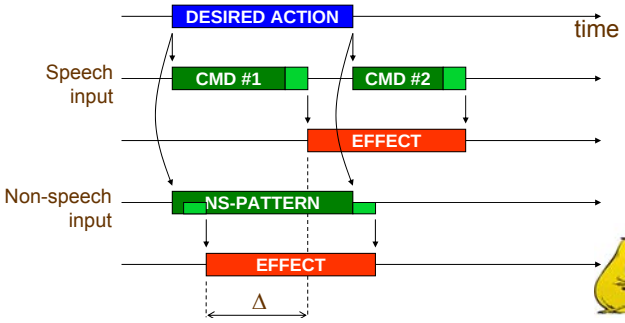
- Earcons
  - Synthetic sounds used to convey information
  - Structured combinations of notes (motives) represent actions and objects



  - Easily grouped and refined due to compositional and hierarchical nature
  - Harder to associate with the interface task since there is no natural mapping

Non-Speech Inputs

- Interpretation of non-speech patterns → faster
- Usable by people with speech/language impairment
- But highly affected by background noise as it relies on FFT or autocorrelation for some



Speech

- Humans have a great & natural mastery of speech
  - makes it difficult to appreciate the complexities, but
  - it's an easy medium for communication
- Structure of Speech
  - Phonemes → 40 basic atomic units; sound slightly different depending on the context they are in
  - Allophones: 120-130; all the sounds in the language
  - Morphemes: smallest unit of language that has meaning
- Other aspects of speech
  - Syntax (structure) & semantics (meaning) of sentences
  - Prosody: alteration in tone; variations in emphasis, stress, pauses and pitch → impart more meaning to sentences.
  - Formant: a combination of tones produced in the articulation of vowels and some consonants

Speech Systems

- Voice is being accepted as a natural way to interact with (mobile) devices and more recently agents
  - Windows Vista provides embedded speech recognition / dictation technology
- Speech Recognition
  - Single user, limited vocabulary systems widely available, e.g. dictation, voice commands (continuous speech)
  - Open use, limited vocabulary systems can work satisfactorily, e.g. phone banking
  - No general user, wide vocabulary systems are commercially successful, yet (state-of-the-arts = 60K words)
  - But useful in hands-busy situations, for people with motor impairment or in mobile applications

Speech Synthesis

- Speech Synthesis and Text-to-Speech
  - Successful in screen reader applications for blind people and in warning systems (in both cases being natural is hardly the main issue)
- First system was in 1804 (von Kempelen)
  - Leather resonator manipulated by the operator to try and copy vocal tract configuration
- Modern speech synthesis
  - Articulatory Synthesis: Model movements of articulators and acoustics of vocal tract
  - Formant Synthesis: Start with acoustics, create rules/filters to create each formant → when computers were relatively underpowered
  - Concatenative Synthesis (a.k.a. unit selection synthesis): Use databases of stored speech to assemble new utterances.

Problems with Speech Systems

- Speech Recognition
  - Different people speak differently: accent, intonation, stress, idiom, volume and so on can all vary.
  - The syntax of semantically similar sentences may vary.
  - Background noises can interfere.
  - People often "ummm....." and "errr....." → meaningful
  - Words are not enough - semantics needed as well
- Speech Synthesis
  - Problems - similar to recognition: prosody particularly
  - Intrusive - needs headphones, or creates noise in the workplace
  - Transient - harder to review and browse → check Whitaker's work on chitty chatty

Speech interface approach:	Unconstrained Natural Language	Dialog trees	Command and Control
User's effort	low	moderate	high (moderate with use)
Developer's effort	very high	moderate	moderate
User training	none	none	required for each application
Supports discovery?	no	yes, inefficiently	possibly
Scales to complex tasks?	unknown	no	no
Scales to hundreds of applications?	yes, but effort not amortized	yes	no
Stress on speech recognizer & parser	high	low	moderate

Table 1 summarizes the properties of the three approaches to speech interfaces discussed above. As can be seen, no approach is satisfactory along all dimensions.

Speech Systems Comparison

Voice Recognition Systems

- Not speech recognition! These are biometric systems.
- Utilizes the distinctive aspects of the voice to verify the identity of individuals.
- Measures vocal qualities not detectable by humans
  - Pitch and frequency are key features measured
- Voice scan algorithms also measure
  - gain or intensity
  - short time spectrum of speech,
  - formant frequencies,
  - Spectrograms
  - Nasal co-articulation
- Replicable only by human voice and therefore more secure
  - Based on statistics based pattern matching called Hidden Markov Models (HMM)

Music

- Influence emotion and decision making (gender and age specific)
- Music in commerce is a multi-million pound industry
  - Many effects of background music in retail and leisure
  - Arousing music speeds up customer activity
  - Relaxing music increases on-hold waiting time
- Music is also used in healthcare
  - Conventional music therapy → treats chronic psychological and physical disorders (e.g. learning difficulties)
  - Music and physical health → There is a link between music and pain, stress, immunity, anaesthetic use, length of labor, weight gain in premature babies etc.
- But very under-explored in computer interaction

Body Motion

High Speed & Precision

Low Speed & Precision

- **1st class**: *finger* motions, e.g., typing, clicking.
- **2nd class**: also *wrist* motions, e.g., moving mouse
- **3rd class**: also *forearm* motions, e.g., using wii
- **4th class**: also *upper arm* and *shoulder* motions, e.g. arcade games – first person shooter
- **5th class**: *whole body* motions: leg, trunk, e.g., more immersive VR

Low Force

High Force

Gilbreth's 17 Basic Therbligs

Effective Therbligs

1. RE = Reach
2. M = Move
3. G = Grasp
4. RL = Release
5. PP = Pre-position
6. U = Use
7. A = Assemble
8. DA = Disassemble

Ineffective Therbligs

9. S = Search
10. SE = Select
11. P = Position
12. I = Inspect
13. PL = Plan
14. UD = Unavoidable delay
15. AD = Avoidable delay
16. R = Rest
17. H = Hold

Effective: advance work
Ineffective: do not advance work – consider whether they can be eliminated.

Tactile Perception

- Touch is complex: tying a shoelace
- Only bi-directional communication channel – both input & output
- Provides information about our environment
  - e.g. hot, cold, smooth, rough
- Provides feedback
  - e.g. when trying to lift an object, press buttons, etc.
- Mechanoreceptors detect skin deformations
- Tactile acuity is determined by how close the mechanoreceptors are to each other and by the size of the receptive field

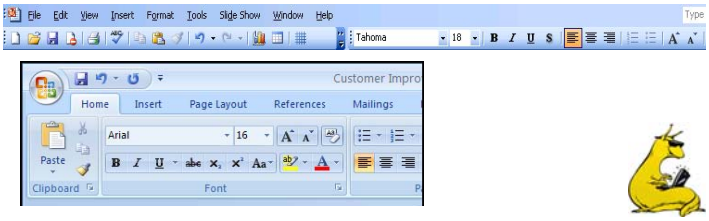
Haptics

- *adj*. Of or relating to the sense of touch; tactile. [Greek *haptikos*, from *haptesthai*, to grasp, touch.]
  - Haptics involves both proprioceptive and tactile senses, in concert with other senses.
- Haptic interfaces
  - Fully duplex channel. You can both transmit and receive information simultaneously.
  - Requires very high refresh rates of approx. 1000 Hz for realistic feel.
  - Requires very high spatial resolution.

- Tactile technology: information is produced by perturbing the skin → Braille devices
- Force feedback technology: Kinesthetic info is produced by exerting mechanical forces

Motor subsystem: Fitts' Law

- Fitts' Law predicts that the time to point at an object using a device is a function of the distance from the target object & the object's size.
- The further away & the smaller the object, the longer the time to click on it.
- Fitts' Law is useful for designing systems for which the time to click on an object is important



Fitts' Law

- *MT* is the movement time
- *a* and *b* are empirically determined constants, that are device dependent.
- c is a constant of 0, 0.5 or 1
- *A* is the distance of movement from start to target centre
- *W* is the width of the target, which corresponds to "accuracy"
- It has an assumption that the most time used is for homing (i.e. better to locate objects on the edges of the screen even if it's further)
- BUT, only accounts for *direct line* movements

$$MT = a + b \log_2\left(\frac{2A}{W} + c\right)$$

Motor Impairment

- Paralysis – usually due to spinal injury, the higher the damage the greater the degree of paralysis
  - tetraplegia/quadriplegia – all four limbs
  - paraplegia – lower limbs only
- Lack of strength (aging = reduced grip strength)
- Tremor/lack of accuracy (Parkinson's disease)
- Slowness (age-related)
- Cerebral palsies: a group of disorders in the development of postural control and mobility
- Some input devices to address motor impairment:
  - HMD and eye-trackers
  - Blow-suck tube, tongue joystick
  - Voice recognition systems
  - Sticky keys, slow keys, gravity well

Cumulative Trauma Disorders (CDT)

- A.k.a. repetitive motion/strain injuries (RSI)
- E.g. carpal tunnel syndrome, tennis elbow, bursitis, trigger finger
- National Safety Council (1997) says that 15 – 20% of workers in key industries are at risk for CTD (meat packing, poultry, auto and garment industries) → 2 claims per 1000 workers @ $30K pre case
- Symptoms: swelling, numbness, pain, restricted movement
- Causes:
  - Excessive force; high repetition
  - Awkward or extreme joint motions
  - Long work without break

Avoiding CTD – Relevant Guidelines

- From the work design guidelines for hand tools

1. Maintain a straight wrist
2. Avoid tissue compression
3. Design tools to be used by either hand
4. Avoid repetitive finger motions
5. Use strongest working fingers: middle and thumb
6. Design appropriately shaped handles
7. Design grip surface to be compressible and non-conductive
8. Keep the weight of the tool below 5 lbs.
9. Use proper configuration and orientation of tools

Multimodal versus Multimedia

- Multimodal systems
  - use more than one sense/mode of interaction
  - e.g. visual and aural senses: a text processor may speak the words as well as printing them to the screen
- Multimedia systems
  - use a number of different media to communicate information
  - e.g. an e-learning system may use video, animation, text and still images and use speech and non-speech
- Interaction with these systems
  - The combination of human output channels effectively increases the bandwidth of the human-machine channel (e.g., Oviatt, 1999).
  - But can cause deterioration in performance (e.g. driving and talking on cell phones?)

Paradigms of Multimodal Interaction

1. Computer as a tool
   - multiple input modalities are used to enhance direct manipulation [Shneiderman, 1982] system behavior
   - the computer is a passive tool and tries to understand the user through all the different input modalities that the system recognizes
   - the user is responsible for initiating the actions
2. Computer as a dialogue partner
   - the multiple modalities are used to increase the anthropomorphism of the user interface
   - agent-based conversational user interfaces
   - multimodal output is important: talking heads and other human-like presentation modalities
   - speech recognition is a common input modality in these systems, and speech synthesis is used as an output modality

Paradigms of Multimodal Interaction

3. Proactive computing
   - used in ambient intelligence, Perceptual User Interface
   - the multiple modalities are used to sense the user and the environment
   - multimodal (multisensory) input is important
   - the functionality of the system depends on the level of deduction the system is capable of
   - proactive functionality is often in the background and only indirectly visible (transparent) for the user, predicting his/her actions and needs.