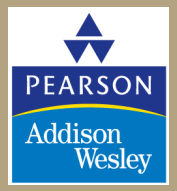
Lecture 31

Based on Chapter 13: Speech and Hearing

Lecturer: Jim Warren

The Resonant Interface
HCI Foundations for Interaction Design
First Edition
by Steven Heim



Chapter 13 Speech and Hearing

- The Human Perceptual System
  - Hearing, speech, non-speech
- Using Sound in Interaction Design
- Technical Issues Concerning Sound

The Human Perceptual System

- Hearing

MAXIM

Our ears tell our eyes where to look

People become habituated to continuous sounds

The Human Perceptual System

- Hearing
  - We can respond to audio input more quickly than we can to visual stimuli
  - We have the ability quickly to locate the source of a sound
    - Interaural time difference (ITD)
    - Interaural intensive difference (IID)

The Human Perceptual System

- Hearing
  - Sound plays a vital role in our sense of connectivity to our environment (Auditory Presence)
  - Immersive and realistic virtual auditory environments require high quality sound
    - To create a more realistic virtual auditory environment, measurements must also be taken of the auditory signals close to the user's eardrum, yielding Head-related transfer functions (HRTFs)

The Human Perceptual System

- Speech
  - Speech is a significant part of our interaction with the world
- Advantages of Speech
  - People gravitate to verbal modes of communication
  - It is easier to speak than to write
- Disadvantages of Speech
  - It requires the knowledge of a language
  - It is more efficient to read than to listen

The Human Perceptual System

- Speech

MAXIM

We can speak faster than we can write

We can read faster than we can listen

- The most efficient method of communication depends on the context

The Human Perceptual System

- Nonspeech Sound
  - We monitor our nonspeech auditory environment habitually and, to some degree, unconsciously
- Advantages of Nonspeech Sound
  - It informs us about the success of our actions
  - It can be processed more quickly than speech
  - It does not depend on the knowledge of a language

The Human Perceptual System

- Disadvantages of Nonspeech Sound
  - It can be ambiguous
  - It must be learned
  - It must be familiar
  - It does not have high discrimination
  - It is transitory
  - It can become annoying

The Human Perceptual System

MAXIM

We often judge the success of an action by auditory feedback

Auditory stimuli are transitory

Sound can be annoying or inappropriate

Chapter 13 Speech and Hearing

- Using Sound in Interaction Design
  - Redundant Coding
    - Reinforces visual feedback on actions
  - Positive/Negative Feedback
    - Maybe we should use sound to confirm success more, instead of always using it to indicate failure
  - Speech Applications
    - E.g., voice tags annotation to a text document
  - Nonspeech Applications
    - E.g., ‘auditory icons’

Using Sound in Interaction Design

- Auditory Icons - Gaver applied the concept of ecological listening to the computer interface
  - Recordings of everyday sounds
  - Exploited analogies with real-world objects and events
    - File types related to different materials
    - File size related to volume or pitch
- SonicFinder
  - A redundant auditory layer that reinforced essential feedback about tasks

Using Sound in Interaction Design

- Benefits of Auditory Icons
  - Disperses some of the cognitive processing over multiple channels
  - Allow users to interact simultaneously with screen objects and with objects beyond the view of the screen
    - Or on other users' screens in a collaborative computing environment

Using Sound in Interaction Design

- Concerns for Auditory Icons
  - Learnability of the mapping between the icon and the object represented
    - "Oink" and "bow wow" have high articulatory directness
    - A swishing sound accompanying a paintbrush tool also has high articulatory directness
    - A system beep carries no information about the error it represents

Using Sound in Interaction Design

- Auditory Icons – Formal Guidelines *(Mynatt)*
  - **Identifiability—**The user must be able to recognize the sound's source. Familiar sounds will be more easily recognized and remembered.
  - **Conceptual Mapping—**How well does the sound map to the aspect of the user interface represented by the auditory icon?
  - **Physical Parameters—**The physical parameters of the sound, such as length, intensity, sound quality, and frequency range, can affect its usability. No one parameter should be allowed to dominate; the user may infer significance.

Using Sound in Interaction Design

- Auditory Icons – Formal Guidelines *(Mynatt)*
  - **User Preference—**How the user responds emotionally to the auditory icon is also important. Is the sound harsh or too cute?
  - **Cohesion—**The auditory icons used in an interface must also be evaluated as a cohesive set. For example, each auditory icon must be relatively unique. They should not sound too similar to each other.

Using Sound in Interaction Design

- Auditory Icons – Procedural Guidelines *(Mynatt)*
  - Use sounds that are:
    - Short
    - Of wide frequency range
    - Equal in length, intensity, and sound quality
  - Use free-form questions to determine how easy it is to identify the sounds
  - If it is not easy to identify the sounds, evaluate how easy it is to learn them

Using Sound in Interaction Design

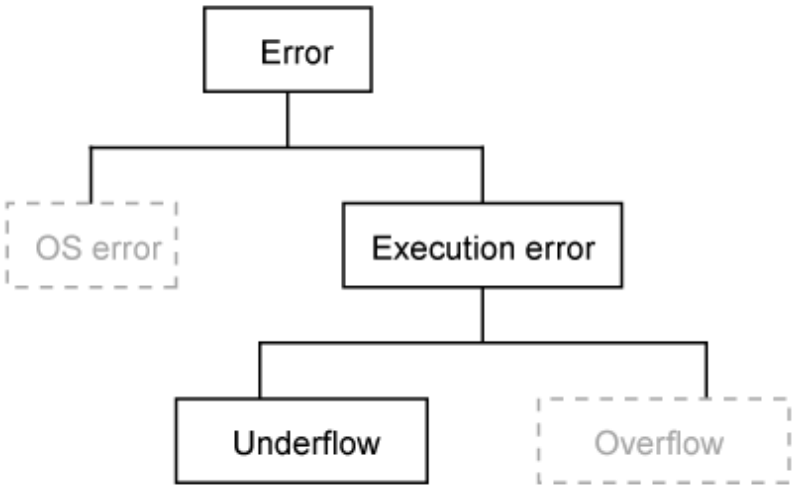
- Earcons (Blattner, Sumikawa, and Greenberg ,1989)
  - "Nonverbal audio messages used in the user–computer interface to provide information to the user about some computer object, operation, or interaction"
  - Short musical phrases that represent system objects or processes
  - Involve musical listening

Using Sound in Interaction Design

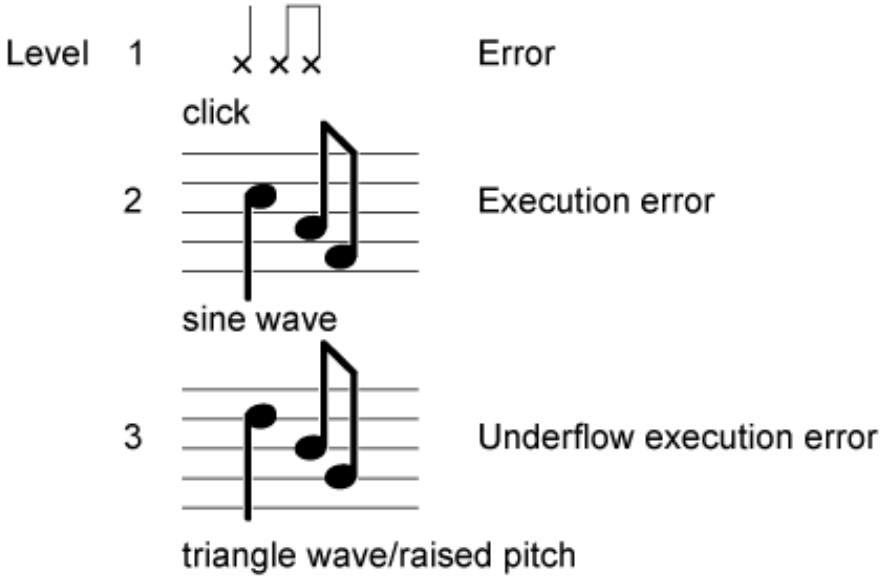
- Earcons (Blattner, Sumikawa, and Greenberg ,1989)
  - Earcons can be used to:
    - Reinforce icon family relationships
    - Support menu hierarchies
    - Support navigational structures

Using Sound in Interaction Design

- Hierarchical Earcons
  - Each node in a hierarchy inherits the attributes of the previous level



Hierarchical structure



Hierarchical earcons

Using Sound in Interaction Design

- Earcons versus Auditory Icons
  - Earcons and auditory icons need not be mutually exclusive
  - Consider the entire structure of an interface, and design its auditory layer with a consistent sound ecology

Using Sound in Interaction Design

- Globalization-Localization
  - Both concrete (real-world) and abstract (Musical) sounds involve cultural biases

MAXIM

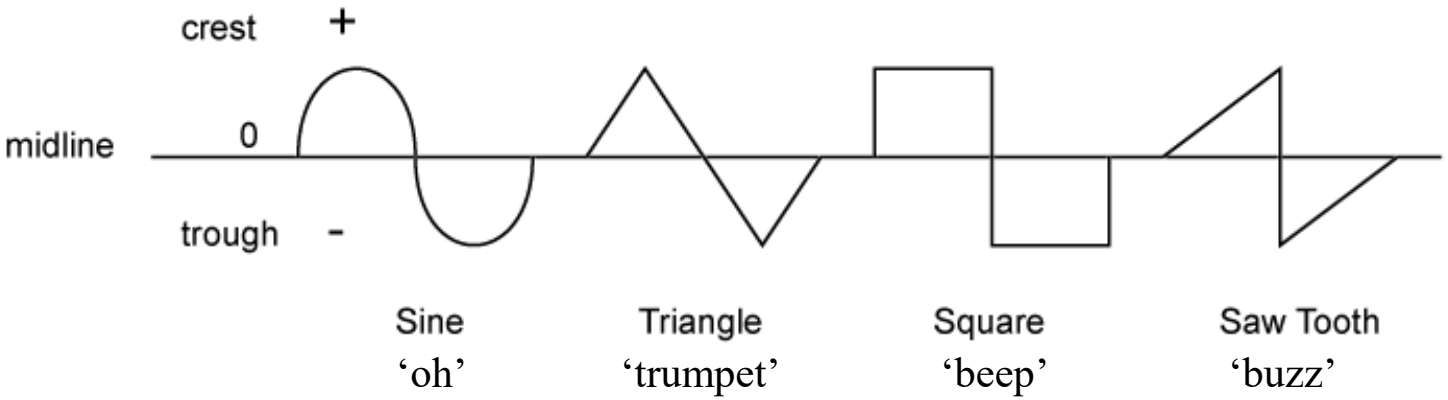
Musical sounds are culturally biased

Chapter 13 Speech and Hearing

- Technical Issues Concerning Sound
  - Sound Waves
  - Computer-Generated Sound
  - Speech Recognition

Technical Issues Concerning Sound

- Sound Waves
  - Sound is made up of waves and can be described in terms of frequency (pitch) and amplitude (volume, power, decibels [dB]), as well as waveform (timbre)



  - The human ear can perceive sound in the range of 20 to 20,000 Hz (20 kHz)
    - With age and noise exposure, though, we tend to lose the upper range

Technical Issues Concerning Sound

- Computer-Generated Sound
  - Synthesis
  - Sampling
  - MIDI
  - Speech Generation
  - Speech Recognition

Technical Issues Concerning Sound

- Synthesis
  - Digital signal generators use software to create sound waves
  - Once the wave is generated, it can be processed to produce an almost unlimited range of sounds
  - Frequency modulation (FM) synthesis
    - The frequency of one sound wave (the modulator) affects the parameters of a second wave (the carrier)
    - We can also apply filters, e.g., to allow only part of the frequency range
    - However, it is difficult convincingly to imitate acoustic instruments

Technical Issues Concerning Sound

- Sampling
  - High-fidelity sounds can be obtained by using digital samples of actual instruments
  - A sample is basically a snapshot of a sound wave at a certain point in time that captures its amplitude information
    - The wave must be sampled at twice the rate of its highest frequency (Nyquist-Shannon sampling theorem)
    - CDs are sampled at a rate of 44.1 kHz, slightly greater than twice the human threshold of 20 kHz

Technical Issues Concerning Sound

- MIDI (Musical Instrument Digital Interface)
  - MIDI files are analogous to the piano roll on a player piano
  - MIDI file contains information about pitch, duration, and intensity
  - MIDI files contain no timbre information
  - Small file sizes
  - Depend on the sounds embedded in the target device

Technical Issues Concerning Sound

- Speech Generation
  - Computers can generate synthetic speech
    - A significant benefit to people with visual handicaps
  - Applications that convert text to verbal output are called “text to speech” (TTS) systems

Technical Issues Concerning Sound

- Speech Generation
  - TTS systems have been used for:
    - Information access systems that facilitate remote access to databases
    - Transactional systems that process customer orders
    - Global positioning system–based mobile navigation systems that output driving directions
    - Augmentative systems that aid disabled users
  - The quality of speech generation is getting better (and computers are increasingly easily able to store digitisation of human speech) so speech emanating from computers is increasingly available for UI designers

Speech Synthesis Markup Language

- Designed by W3C to provide a rich, XML-based markup language for assisting the generation of synthetic speech in Web and other applications

```
<?xml version="1.0"?>
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.w3.org/2001/10/synthesis
    http://www.w3.org/TR/speech-synthesis/synthesis.xsd" xml:lang="en-US">
      Take a deep breath <break/> then continue.
      Press 1 or wait for the tone. <break time="3s"/>
      I didn't hear you! <break strength="weak"/> Please repeat.
</speak>
```

- Tags for controlling pitch, pace, emphasis, etc.
  - See http://www.w3.org/TR/speech-synthesis/

Technical Issues Concerning Sound

- Speech Recognition
  - Two distinct applications:
    - Transcription
    - Transaction
  - Automatic speech recognition (ASR) systems allow users to speak in real time and this input is converted into text that is displayed on the screen
  - E.g., Dragon Systems' NaturallySpeaking® and IBM's Via Voice®

Speech Recognition

- Full vocabulary speech recognition is still hard
  - Requires system to be trained to user
  - Requires user to speak distinctly
  - Best with optimal mic placement
- Limited vocabulary is easier
  - E.g., for limited-domain commands
  - Can understand unrestricted speakers for simply domains like digits

Technical Issues Concerning Sound

- Speech Recognition Concerns

MAXIM

Speech can interfere with problem-solving activities

Verbal input can be inappropriate in certain situations

Technical Issues Concerning Sound

- Searching Speech
  - Speech files do not afford easy opportunities for indexing and searching
  - ASR systems can be used to transcribe speech files and create transcripts that can be searched like any other text file

Technical Issues Concerning Sound

- Multimedia Indexing
  - Large collections of multimedia documents are being created in domains as diverse as medicine, entertainment, and education
  - Archiving speech according to content can be difficult:
    - The system must not only recognize the meaning of spoken language, it also must create relationships according to content

Technical Issues Concerning Sound

- Multimedia has become a common element in contemporary computing environments, however, we have only begun to understand how to take advantage of its potential

Summary

- Hearing is an important human sensory channel providing context and feedback
- Non-speech sounds can be used like icons to create an auditory language for HCI
- Speech processing – generation and recognition – is just beginning to enrich our user interfaces