Chapter 13 Speech and Hearing

• The Human Perceptual System
• Using Sound in Interaction Design
• Technical Issues Concerning Sound
Chapter 13 Speech and Hearing

Sound exists in time and over space, vision exists in space and over time.

(Gaver, 1989)
Chapter 13 Speech and Hearing

• The Human Perceptual System
  – Hearing
  – Speech
  – Nonspeech
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
    • 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
  – Positive/Negative Feedback
  – Speech Applications
  – Nonspeech Applications
Using Sound in Interaction Design

• Redundant Coding
  – Research shows that redundant coding has certain benefits:
    • It aids memory by adding additional associations.
    • It increases efficiency by allowing the most efficient mode for a particular task to be chosen.
    • It allows users with perceptual deficits to take advantage of their strengths in other channels.
Using Sound in Interaction Design

- Positive/Negative Feedback
  - How we use auditory feedback in interaction design must be determined by the user’s task
    - Redundant auditory alarms might be crucial to the safe operation of computer-operated machinery in mission-critical environments
    - General computing interfaces that announce every mistake the user makes should be seriously reconsidered
Using Sound in Interaction Design

• Speech Applications
  – Composition
    • Digitized speech can be used to facilitate and augment the process of composing documents
  – Transcription
    • A searchable record of the spontaneous conversations that occur in business meetings could prove advantageous
  – Transaction
    • Computers are often used to carry out tasks that are initiated by a user; this is a form of transaction
  – Collaboration
    • Telephone system
    • Audio IM
Using Sound in Interaction Design

• Nonspeech Applications
  – Concrete (auditory icons)
    • Ecological listening (natural sounds)
      – Distal stimulus (sound source)
  – Abstract (earcons)
    • Musical listening (synthetic sounds)
      – Proximal stimulus (physical sound properties)
Using Sound in Interaction Design

• Auditory Icons

MAXIM
People generally attend to the source of a sound rather than the acoustic properties of the sound wave

• People do not attend to the “physics of the sound waves” (the proximal stimulus), they listen for the sound source (the distal stimulus) (Gaver, 1989)
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
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.
  – Tasks could be distributed immediately on arrival of the auditory output in collaborative computing environments
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

- Auditory Icons – Procedural Guidelines (Mynatt)
  - Evaluate sets of icons to determine whether they conflict with each other
    - Do they mask each other?
    - Do the mappings conflict?
    - Are they easy to tell apart?
  - Conduct usability tests
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
  - Based on the concept of the musical motive
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

• **Compound Earcons**
  – Concatenated motives

```
Level 1
  \[\text{Bass Drum} \times\]

Level 2
  \[\text{Medium Violin Note}\]
  \[\text{Medium Flute Note}\]

Level 3
  \begin{tabular}{ccc}
  1 xly. note & 2 xly. notes & 3 xly. notes \\
  \[.\] & \[\ldots\] & \[\ldots\]
  \end{tabular}
  \begin{tabular}{ccc}
  1 vib. note & 2 vib. notes & 3 vib. notes \\
  \[.\] & \[\ldots\] & \[\ldots\]
  \end{tabular}

Result \[\times_{\ldots}\] \[\times_{\ldots}\] \[\times_{\ldots}\] \[\times_{\ldots}\] \[\times_{\ldots}\] \[\times_{\ldots}\] \[\times_{\ldots}\] \[\times_{\ldots}\]\]
Hierarchical Earcons

- Each node in a hierarchy inherits the attributes of the previous level
Using Sound in Interaction Design

• Parallel Earcons
  – Represent complex hierarchical structures by presenting various complete motives concurrently
    • Create Earcon—A long note that starts softly and gradually gets louder.
    • File Earcon—Two notes, the first higher than the second, and each half the length of the Create earcon.
Earcon Guidelines

• Timbre
  – Timbre is the most important grouping factor and the most easily recognizable parameter.
  – Use sounds with multiple harmonics.

• Pitch and Register
  – Do not use only pitch and register as a cue if absolute judgment is required.
  – All pitches in a motive should span only one octave.
  – Use major and minor scales for pitches within a motive.
  – Use two- or three-octave differences between earcons when using register only.
  – Use pitches within the 125 Hz to 5 kHz range.
Using Sound in Interaction Design

Earcon Guidelines

• Rhythm, Duration, and Tempo
  – Rhythm is the most prominent characteristic of a motive.
  – Make rhythms as different as possible.
  – Use different numbers of notes.
  – Do not use notes shorter than 0.03 second.
  – Use different tempos to differentiate earcons.
  – A motive should be long enough to convey meaning but no longer.
Using Sound in Interaction Design

Earcon Guidelines

• Intensity/Dynamics
  – Use intensity and dynamics sparingly due to the potential annoyance factor.
  – Threshold limits are from a maximum of 20 dB to a minimum of 10 dB.

• Spatial Location
  – Use stereo or full three-dimensional spatialization to differentiate families and
  – parallel 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 and amplitude

  ![Sound Wave Diagram](image)

  - The human ear can perceive sound in the range of 20 to 20,000 Hz (20 kHz)
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)
    • 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
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
  – Dragon Systems’ NaturallySpeaking® and IBM’s Via Voice®
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
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

• Searching Speech
  – There are times when the visual display of data is either not appropriate or simply not possible
  – Researchers have explored ways of compressing speech data so that they take less time to parse
    • Speed up the sound file
    • Sampling, involves the removal of redundant information from the signal
    • Shorten or remove the small pauses between words
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