Audio- Speech

Joemon Jose
http://www.dcs.gla.ac.uk/~jj/teaching/demms4

Non-Speech Sound

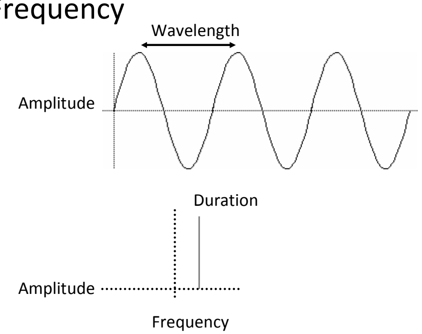
- What is sound?
- Sampling sounds
- Synthesizing sounds
- MIDI
- Sound in interface design

Audio

- Differences between images and audio
  - We can drop frames from video
  - Not with audio! Why?
- What is sound
  - Wave phenomenon
  - Without air there is no sound
  - Sound is a pressure wave- it takes on continuous values
  - We must digitise it to get process tem in a computer

What is Sound?

- Frequency



200 Hz
500 Hz
1500 Hz

How Do We Hear?

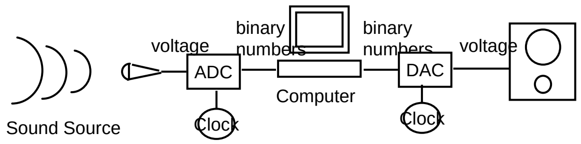
- Ear has three parts
  - inner ear
  - middle ear
  - outer ear
- Outer ear
  - Pinnae and ear canal
- Middle ear
  - ear drum connects to inner ear
  - amplification
- Inner ear
  - Cochlea - vibrations stimulate auditory nerve

How Do We Hear?

- Humans can hear 20Hz to 20kHz frequency range
- Can hear differences of around 1.5Hz
- Hearing loss occurs with age
  - age 50 max 14kHz, at age 70 max 10kHz
  - important for interface design

Digital Sound Recording and Playback

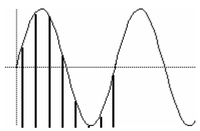
- Convert *analogue* signal into *digital* signal for storage and manipulation ("sampling")
- Convert digital signal to analogue for playback
- Sampler may be PC or dedicated h/w sampler



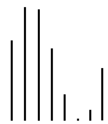
Sampling - Sample Rate

- Signal measured at a set of distinct times - sample rate

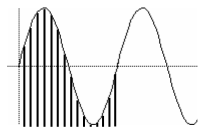
- 8 samples per cycle



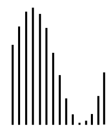
- Resulting sample data



- 16 samples per cycle



- Resulting sample data



Sampling - Aliasing

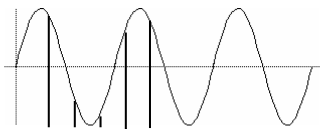
- Higher the sampling rate, the higher the match between the original signal and that reconstructed from sample data
- Sample rate must be greater than 2 x frequency
  - aliasing problems
  - distortion

Aliasing

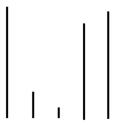
- Good sampling rate is 44,100 samples/second
  - maximum frequency is therefore 22,050Hz
  - humans can only hear up to ~20,000Hz
  - can therefore filter sound source to avoid aliasing
- New DVD Audio specification
  - 192kHz sample rate / 24bit sample size (max)
  - max frequency 96,000Hz

Aliasing

- 10Hz sound, 5 samples/Sec.



- Resulting sample data



Sampling - Sample Size

- Larger the bit size the better the amplitude range (*dynamic range*)
  - usually 8 bit (only 256 possible values)
  - or 16 bit (65,535 possible vales)
- Quantization problems
  - with 8 bit numbers cannot represent many amplitude levels
  - can result in *noise*
- Lower quality can affect usability

Sample Formats

- Many, many formats for sampled sound
- CD quality 44.1kHz, 16bit (per channel)
  - 44,100 x 2 x 60 = 5.2 MB/minute
  - .wav for PC, AIFF for Mac
- Medium quality 22kHz, 8 bit
- Low quality 11kHz, 8 bit, mono
  - 11,000 x 60 = 660KB/minute

44.1kHz, 16bit
11kHz, 8bit
5.5kHz, 8bit

Sample Formats

- New DVD Audio specification
  - 34MB/min per channel (max)
- Telephony quality 8kHz, 8bit, mono
  - mu-law : samples encoded logarithmically in 8 bits -> 12 bit linear range

Audio Compression

- Want high quality sounds but need to reduce size
- MP3
  - MPEG-1 (Motion Pictures Expert Group) Audio Layer 3
- Uses a range of perceptual coding techniques to reduce file size but keep quality
  - lossy compression

MP3 Compression

- Can reduce file sizes by a factor of 12
- Bitrate - number of bits for one second of audio

Bitrate	Quality	MB/min
1411	CD	10.584
192	good CD quality MP3	1.440
112	near CD quality MP3	0.840
64	FM quality MP3	0.480
32	AM quality MP3	0.240
16	Shortwave quality MP3	0.120

Sound Synthesis

- Important factor is number of sounds generated together
  - multitimbral
  - 64 or 32 voice are common

Sound Synthesis

- Many synthesisers available
  - mostly controlled by MIDI (see below)
  - maybe separate h/w, a sound card or s/w synthesiser built into PC
- Many different types of synthesis techniques
  - additive synthesis
  - FM synthesis
  - wavetable
  - physical modelling

MIDI - Musical Instrument Digital Interface

- Allows real-time control of electronic musical instruments
  - synthesisers, samplers, etc.
- Specifies a h/w interconnection scheme + protocol for data communications + grammar for encoding musical performance data

MIDI

- MIDI data is like a ‘piano roll’ - gives note on, off, instrument info
- Much less information is contained than for samples
  - MIDI files much smaller than sample files
- MIDI does not encode timbre so synthesis is left up to synthesiser
  - things may sound different when played back

Advantages of MIDI

- Allows play back of sounds on many different types of synthesiser
  - almost!
  - General MIDI
- Separates i/p device from sound generator
  - one keyboard can play many synths
- Many different types of i/p device can be used
  - piano keyboard, computer, MIDI guitar, etc.
- MIDI can control a wide variety of devices
  - synthesiser, sampler, audio effects, lighting

Basic MIDI Commands

- MIDI messages used to pass data between MIDI devices
- Two types of command
  - channel and system
- Channel commands allow
  - turning notes on and off, setting instruments to channels, etc.
- System commands allow
  - synchronisation, system exclusive

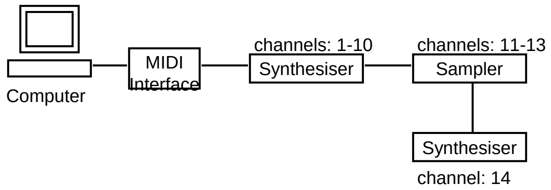
MIDI Setup

- PC (maybe piano style keyboard) used to control the MIDI system
- S/w running on PC will generate MIDI messages
  - might be a sequencer or other s/w
  - these sent to synthesiser (or sampler) to play sound

MIDI synthesizer with MEGA32 microprocessor

MIDI Setup

- A standard MIDI setup might look like:



Sound in Interface Design

- Historically used for background music and sound effects in theatre, TV, films
- Can be used in this way in computers
  - entertainment/educational products
- Emerging modality in multimedia
  - strong technical infra-structure
  - weak design/research infrastructure

Sound in Interface Design

- More natural information representation
  - 'sonification' of high dimensional data
  - spatialized audio for more immersive displays
- Essential for eyes-free applications
  - aids for blind people
  - compact mobile devices
  - telephone displays
- **Sonification** is the use of non-speech audio to convey information or perceptualize data.
  - Due to the specifics of auditory perception, such as temporal and pressure resolution, it forms an interesting alternative to visualization techniques, gaining importance in various disciplines.

Why Use Sound in User Interfaces?

- Why use non-speech sound ?
  - interdependence of vision and hearing
  - natural means of presentation
  - reduce the load on the visual sense
  - increase the bandwidth of communication
  - omni-directional
  - attention grabbing
  - your eyes can only do one thing at once

Why use Sound in User Interfaces?

- Who might benefit from sound?
  - users of graphical interfaces
  - visually disabled people
  - users of telephone-based interfaces
  - users of interfaces where eyes are busy (planes, cars)
  - mobile computer users (lack of screen space)

Sound in User Interfaces

- Technology is available now
  - DSP / MIDI / Sound cards in every PC
  - Many mobile computers make some sounds
  - Only used in games not everyday interactions
    - games very sophisticated
- Two main types of sounds
  - auditory icons
  - earcons

Auditory Icons

- Developed by Bill Gaver
- Everyday, natural sounds represent objects and actions in the interface
- Sounds have an intuitive link to what they represent
- Sounds are multi-dimensional

Earcons

- Structured audio messages based on abstract sounds
  - A brief structured sound pattern used to represent specific item or event
  - 1989
- a five-day weather forecast on a local news program where each day's temperatures set the pitches in a five tone sequence.
- Earcons are abstract rhythmic / melodic patterns rather than everyday sounds.
- More info
  - http://www.dcs.gla.ac.uk/~stephen/generalearcons/generalearcons1.shtml

Use of Sound in User Interfaces - Mobile Computers

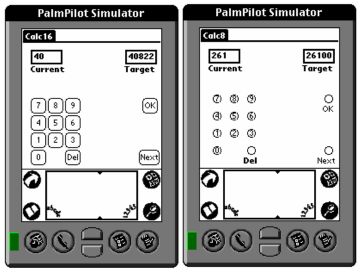
- Limited screen - need other forms of output
- Sound can present information about what is on screen
  - could allow visual widgets to be reduced in size
  - reduce visual clutter / get more on screen
- Experimental design
  - 16 participants
  - 2-condition, within-groups design, fully counterbalanced

Sound in Mobile Computers

- Conditions: Large buttons / Small buttons
- Two treatments per condition: Sound / No sound
- Hypotheses
  - sounds should allow more data to be entered for both button sizes
  - should be no increase in annoyance due to sounds
  - people should be able to walk further with sounds

Mobile Computers

- Large buttons
  - 16 x 16 pixels
  - standard size
  - highlight by reverse video
- Task
  - entering 5 digit codes



- Small buttons
  - 8x8 pixels

Earcons Used

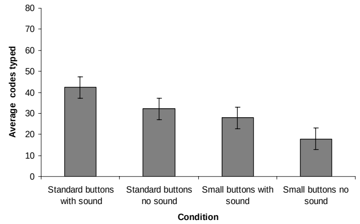
- Sounds constrained by device capabilities
- Silent condition - no sounds
- Sound condition: Standard Palm III sounds plus enhancements
  - pen down - medium pitch
  - pen release - higher pitch
  - mis-press error - lower pitch

Location

- Pathway by the University
- Participants had to walk 10m laps whilst entering data on 3Com Palm III

Results – Numbers of Codes Typed

- Number of codes typed
  - in both conditions significantly more codes entered with sound than without



  - Sound can overcome limitations of small screens

Results – Distance Walked

- More laps walked with sound
- Small buttons with sound as effective as standard silent buttons