Sound Design

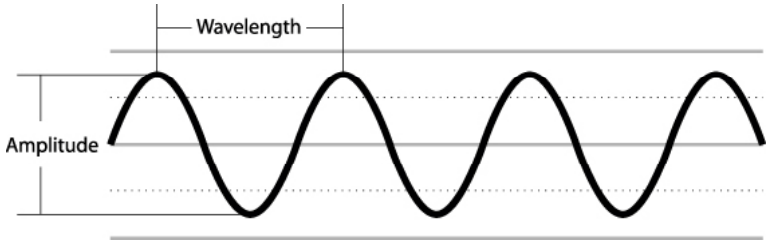
305900
Fall 2010
10/11/2010
Kyoung Shin Park

What Good is Sound?

- It enhances the visual (and vice versa)
- It is all around us (surround)
- It is a rich source of information (and timing)
- It can have psychological effects (psychoacoustics)

The Physics of Sound

- Sound is a wave that interacts with the environment
- Sound has some basic components
  - Frequency, Amplitude, Tone



The Physics of Sound

- Frequency
  - The number of cycles (or periods) that a vibration completes per unit time
  - Frequency determines the height of a sound wave (i.e., amplitude)
    - Higher frequency, louder sound; Lower frequency, softer sound
  - Frequency has an inverse relationships to the concept of wavelength (lambda)
    - $f = v/\lambda$ where v is the phase speed, λ is the wavelength
  - Measurement unit:
    - The unit of frequency is Hertz/Hz or cps (cycles per second)
    - For humans, hearing is limited to frequencies between about 20 Hz and 20 kHz
    - Human can make sound between about 100 Hz and 6 kHz

The Physics of Sound

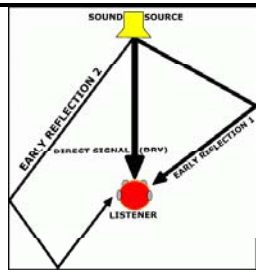
- Amplitude/Intensity
  - The pressure of sound waves in the atmosphere
  - Directly related to the acoustic energy or intensity of a sound
  - Amplitude, or height of a sound wave is a measure of the amount of energy in the wave
  - The decibel is a logarithmic unit of measure of sound amplitude, N = 20 log P1/P2
    - P1/P2: the two pressures to be compared
    - N: the number of decibel
  - Common Reference Level or threshold: 0.0002 dyn/cm2
    - The least amount of pressure necessary for an average human to hear a 1000 Hz tone.
  - Sound intensity level (in dB)
    - Light leaf rusting, calm breathing : 10 dB
    - Normal conversation at 1m: 40-60 dB
    - Jet at 100 m: 110-140 dB
    - Threshold of pain: 140dB

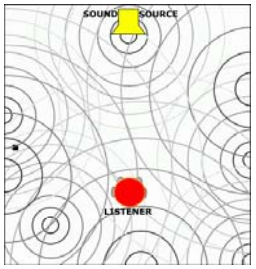
The Physics of Sound

- Pitch
  - The psychological perception of frequency
  - Pitch carries with it psychological and emotional effects
  - Lower pitch – more somber
- Loudness
  - Perceived from *both* amplitude ***and*** pitch
  - Low pitched sound tends to sound quieter
- Timbre
  - The tone quality or "color" of a sound
  - Allows you to distinguish two sounds of the same pitch and loudness (E.g.: a violin vs. flute playing the same note)
- Envelope
  - ADSR (Attack, Initial decay, Sustain, Release)
  - Percussion instrument vs. String instrument

The Psychology of Sound (Psychoacoustics)

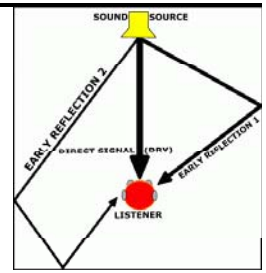
- Psychoacoustics deals with perceptions of sound and our subjective responses to sound
  - It is entirely connected to the environment and to our experiences
- Sound in the environment
  - Direct sound: reaches the listener first
  - Early sound: early reflections
  - Reverberant sound or "Reverb": the decay of early sounds

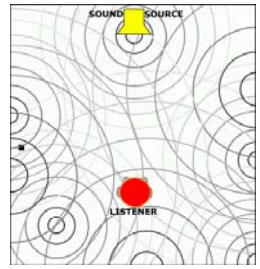




The Psychology of Sound (Psychoacoustics)

- Reverb vs. Echo
  - Reverb gives "space" to our perception of sound by bouncing off multiple walls (random delay of sounds at < 35ms)
  - Echo is the repeating of a sound (evenly delayed sounds at > 35 ms)
- Noise
  - Unwanted sound of any type
  - Unwanted random addition to a wanted signal in both analog and digital electronics
  - A pleasant sound has a regular wave pattern. The pattern is repeated over and over. But noise has the irregular wave pattern.





Complex Sound & Fourier Analysis

- Typical everyday sounds is said to be complex
  - Complex sound contains a range of frequency components (and their amplitudes)
  - Complex sound consists of
    - Fundamental tones: basic frequency
    - Harmonics: multiple of fundamental tones
    - E.g. : fundamental tone 440 Hz -> harmonics 880 Hz, 1320 Hz, 1760 Hz, etc
  - The relative intensity between fundamental tones and harmonics determines timbres.
    - E.g. : ViolinC4, Piano C4
- Fourier Analysis
  - A complex sound can be broken down into a series of simple sounds according to Fourier analysis.
  - Human ears performs a Fourier analysis in his acoustical law.

Sound Design

- Three areas in which you must make decisions about sounds: Voice/Speech, Sound Effects, Music
  - Voice/Speech : sound made by human using the vocal cords for talking, singing, laughing, crying, screaming, etc
  - Sound Effects : artificially created or enhanced sounds, or sound processes used to emphasize artistic or other content of films, TV, live performance, animation, video games, music or other media
  - Music : art form of sound that consists of pitch, rhythm, dynamics and sonic qualities of timbre and texture
- In audio signal processing, sound is divided into digital audio and MIDI (Musical Instrument Digital Interface)
  - Digital audio : uses digital signals for sound reproduction that includes analog-to-digital conversion, digital-to-analog, storage, and transmission
  - MIDI : industry-standard protocol that enables electronic musical instruments such as keyboard, computers, and other electronic equipment to communicate, control, and synchronize with each other

Speech

- The vocalized form of human communication
- It is divided into digitalized speech and synthesized speech
- Verbal vs. Nonverbal
  - Examples of Non-verbal speech : Laugh, sigh, moan, etc.
  - In 'The Sims' game, laughter, sighs, cry, etc
- Narration
  - Can deliver concrete information & Can replace text
- Dialogue
  - Between two people can convey anger, tension, friendliness (even if non-verbal – as in 'Knights of the Old Republic')

Sound Effects

- Sound effect is used to emphasize or augment the content to more realistically convey the information.
- In films and TV, a **sound effect** is a sound recorded and presented to make a specific storytelling or creative point without the use of **dialogue** or **music.** (e.g. door slams, weapons firing, background raining sound effects, etc)
- Sound effect is divided into natural sound effects and synthesized sound effects.

Sound Effects

- Contextual sound/Narrative sound
  - Sound interprets the visuals as it happens (E.g.: Jet engine during take off, normal flight, in trouble)
- Focusing attention (through sound spatialization)
- Defining space (open space vs. closed space)
- Establishing a place (sky vs. under the water)
- Creating environment (mood/atmosphere)
- Emphasizing/Intensifying action (explosion)
- Setting pace (car engine sound)
- Symbolizing meaning (correct vs. wrong answer)
- Unifying transitions (segues)
- Audio file – order of sound collage can produce a narrative

Music

- Music creates the right atmosphere in information deliver
- Music creates additional effects (mood, scene transition, emotional climax)
- Must be used in properly

Music

- Establishing place (music style)
- Creating environment/atmosphere
- Emphasizing action (tempii, dynamics)
- Intensifying action (rhythm, dynamics)
- Depicting identity (character main theme)
- Setting pace (tempii, rhythm)
- Setting time (musical style/period- e.g. classical)
- Unifying transitions (lead-ins, segues and overlaps)

Music

- Things to consider regarding use of music:
  - Spotting : The process of determining where music, sound effects, and ambiance are placed on a soundtrack or in a video game
  - Main titles : Sets the initial mood of the game – but be careful that the actual gameplay matches
  - Music painting : Background music
  - Leitmotifs : Repeating themes – e.g. music theme for a particular character
  - Weighted music : music that can literally draw you to tears in a film
  - Orchestration / instrumentation : kind of instrument to depict the musical piece (classical vs. techno)
    - Also, multiple musical pieces need to be seamless integrated
  - Cues : alerting you to a new event, like from exploration to combat mode
  - Unifying transitions (lead-ins, segues between scenes)

Sound Design

- Roles of **audio media** in **multimedia production**
  - Sound parallels picture : audio element combines with the visual element to create a mood or deliver information that is more potent than either element alone (sounds of battle with gunshots, cannon, and anguished screams complement the visual of a battle scene. The ferocity and destruction of war is conveyed by both media separately, but is intensified by both elements together)
  - Sound defines picture : sound is so distinctive that mental image is formed before visual (rainfall, bird calls)
  - Picture defines sound : sound is a literal translation of the picture (raging storm demands crashing wave sounds)
  - Sound counterpoints picture : both media elements contain unrelated information that creates an effect that is not conveyed by either media element alone (music score during battle in Gladiator)

Silence

- Silence "is deafening"
- Every moment of a multimedia presentation does not need to be filled with sound. Silence can be used to set a mood or to provide a moment for reflection.
- The audio channel is much more capable of maintaining attention if it is used as an interjection on the visual channel rather than being continuously parallel with the visual.
- Human brain is expert at detecting CHANGE.
  - E.g. If your eyeballs aren't always moving you won't see anything
- E.g. Asteroid sequence in Attack of the Clones

Sound Production

- Sound Sources
  - Production libraries
  - Internet
  - Recording (Foley/collecting)
  - Sampling (copyright)
- Audio Programming APIs
  - DirectSound & DirectMusic
  - OpenAL
  - Third party audio-libraries (requires license)
    - Miles Sound System (www.radgametools.com)
    - FMOD (www.fmod.org)
    - Sensaura (for PC, Xbox, PS2, Nintendo, ...)

Sound Production

- Game audio programming API:
  - Low-Level API
    - 2D sound functions (play, pause, stop, pan, volume, ...)
    - 3D sound functions (position, velocity, ...)
    - 2D musical segment functions (play, pause, stop, ...)
    - Hardware functions
  - Mid-Level API
    - PCM wave file loading
    - Massive audio file streaming
    - Sequential music segments queuing
    - Sound Resource Management
    - Sound Load/Unload management
    - One-function loading and playback on demand
    - Handle multiple compressed audio formats, (including Ogg Vorbis, MP3, WMA, ...)
    - Sound effects (E.g.: environmental reverberation)
  - High-Level API
    - Script parsing and loading sounds
    - Creating a high-level soundscape system
    - Creating an advanced dynamic musical cueing and transition system based on the audio scripting system

Sound Production

- Mixing
  - Never forget the goal format! (e.g. tape vs. cd vs. cheap speakers vs. surround sound)
  - Localize/spatialize (Reverberant sound)
  - Layer (use various sounds)
  - Levels (Try different "loudnesses" in mixing)
  - Audition
  - Audition with game – i.e., test it with the game & test it with multiple sounds together

In Your Game

- Find a well rated commercial game that is the closest to your game.
- Close your eyes.
- Have someone else play the game and just listen.
- Now play through the game again, and try and answer these questions:
  - Why are the sounds/music/speech there?
  - What are they doing to the visuals?
  - How are they doing something to the visuals?
  - Is there anything unusual about them? What special method was used to create them?
- In the documentation on your web site try to articulate, using concepts presented here, the answers to the above questions for your game!
- When you show your demo, play it first without any of the visuals.

Audition

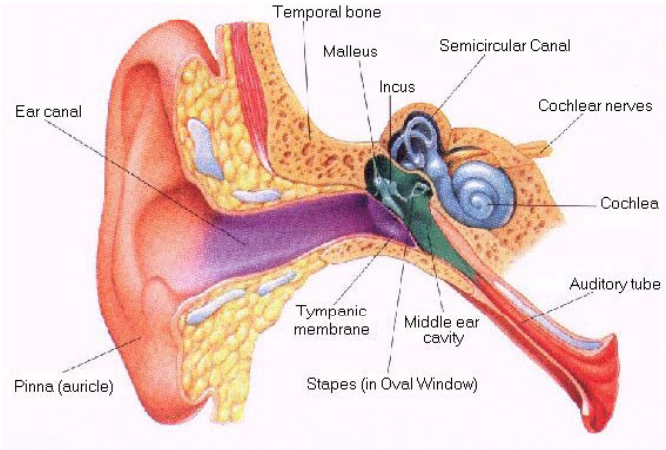


Image from http://www.infj.ulst.ac.uk/~pnic/HumanEar/Andy's%20Stuff/MScProject/workingcode_Local/EarChapter.html

Auditory System

- Ear
  - The outer/external ear focuses and directs sound waves into the middle ear.
  - In the middle ear, the acoustic energy from compression waves in air is translated into mechanical vibrations of the middle ear's bone structure.
  - The cochlea of the inner ear propagates these mechanical signals as waves in fluid and membranes, and finally transduces them to nerve impluses which are transmitted to the brain.
- Brain & Auditory sense
  - Auditory pathway
  - Auditory cortex

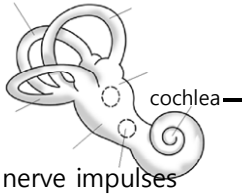
Auditory System

- External ear
  - It gathers sound energy and focuses it on the eardrum – i.e., capturing, focusing, filtering sound
  - Pinna
    - Collect and focus sound waves
  - Ear canal
    - From the pinna the sound pressure waves move into the ear canal
    - Amplifies frequencies in the range 3 kNz to 12 kHz.
  - Tympanic membrance (or Eardrum)
    - Its function is to transmit sound from the air to the ossicles inside the middle ear.

Auditory System

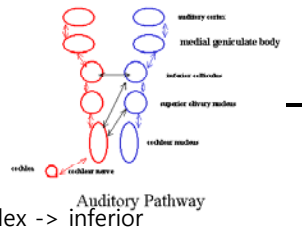
- Middle ear
  - Transfer acoustic energy from compression waves in air to fluid-membrance waves within the cochlea
  - Contains three ossicles, which couple vibration of the eardrum into waves in the fluid and membranes of the inner ear
  - Ossicles : Malleus, Incus, Stapes
    - Sound amplification : Eardrum -> hammer -> anvil -> stirrup -> Oval window of cochlea
    - Sound pressure is concentrated, leading to a pressure gain of at least 22.
    - The auditory ossicles can also reduce sound pressure by uncoupling each other through particular muscles.
  - Movement of ossicles may be stiffened by two muscles:
    - Tensor tympani
    - Stapedius

Auditory System



- Inner ear
  - Finally transduces mechanical sound signals to nerve impulses which are transmitted to the brain.
  - Cochlea
    - The sensory organ of hearing, which is distributed along the partition separating fluid chambers in the coiled tapered tube of the cochlea
    - Cochlea is filled with a watery liquid which moves in response to the vibrations coming from the middle ear via the oval window.
    - As the fluid movies, thousands of hair cells are set in motion, and convert that motion to electrical signals that are communicated via neurotransmitters to many thousands of nerve cells.
    - Displacement of the basilar membrane peaks at the cochlea base for high frequencies and at the apex for low frequencies
      - High pitches : at the base of the cochlear (near the oval widow)
      - Low pitches : at the apex

Auditory System



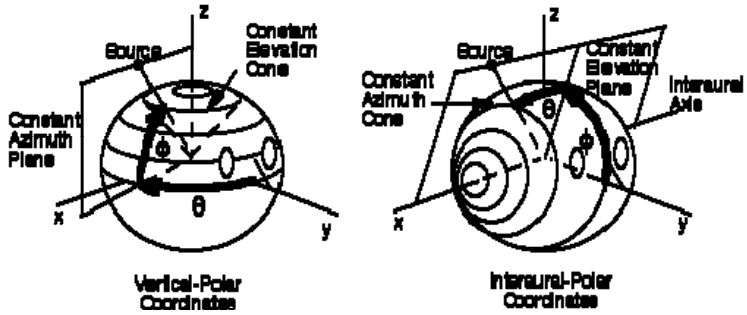
Auditory Pathway

- Auditory pathway
  - Cochlea
    - Cochlear nucleus -> Superior olivary complex -> inferior colliculus ->auditory cortex
  - Cochlear nerve (also auditory nerve or acoustic nerve)
    - Sensory nerve in the head that carries signals from the cochlea of the inner ear to the brain.
    - The temporal lobe is involved in auditory processing and is home to the primary auditory cortex.
  - Inferior colliculus
    - The principal midbrain nucleus of the auditory pathway and receives input from several more peripheral brainstem nuclei in the auditory pathway as well as inputs from the auditory cortex
  - Deafness
    - Conduction deafness, Sensorineual deafness, Central deafness

3D Sound Localization

- Spatialization
- Works well in plane of ears
- Based on:
  - Interaural time differences (ITD)
  - Interaural intensity differences (IID)
- Head-Related Transfer Function (HRTF)

Vertical-Polar Coordinate System



q: azimuth (angle between the nose and a plane containing the source and the vertical axis z)
f: elevation (angle between the horizontal plane by a line passing through the source and the center of the head)
r: range (distance to the source measured along this line)

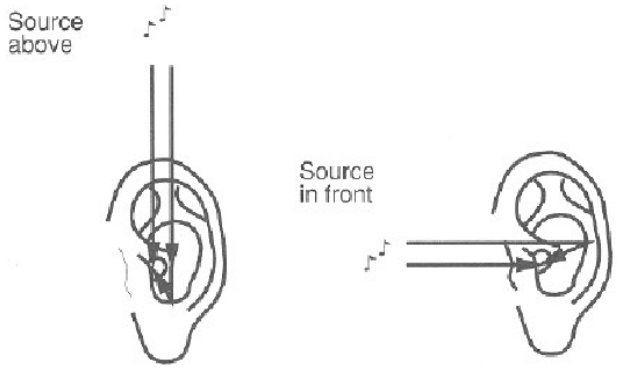
Azimuth Cues

- Interaural time difference (ITD)
  - Difference in the arrival time of the sound at the two ears
  - ITD is zero when the azimuth angle is 0 degree – i.e., the source is directly in front of or directly behind the head
  - ITD = (a/c)(q + sin q)
    - a: the head radius
    - c: the speed of sound (~343 m/s)
    - q: source azimuth
- Interaural intensity differences (IID)
  - Shadow by head
  - Difference in the intensity of sound reaching the ears
  - The closer ear hears a sound with higher intensity
  - Detectable for sounds with high frequencies (>1.5kHz)
  - For low frequency, ITD dominates

Elevation Cues

- Different reflections
  - Due to the asymmetry of the outer ear, especially the pinna
  - The path difference between the direct and pinna-reflected sound changes with the elevation angle
  - Sound coming from a source located above the user's head has quite a different reflection path than sound coming from a source in front of the user
- Different amplification (and attenuation)
  - By interference between reflected sounds
  - Some frequencies are amplified and others are attenuated
- Pinna provides the primary cue for source elevation
  - User's face and shoulders geometry also influences the way the sound is reflected towards the outer ear

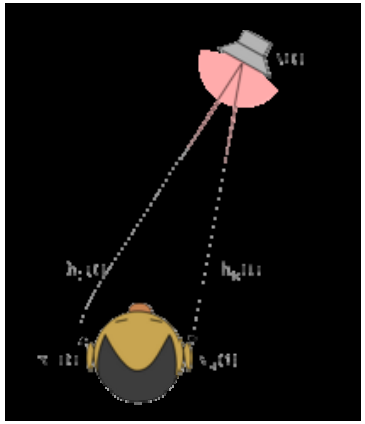
Elevation Cues



Range Cues

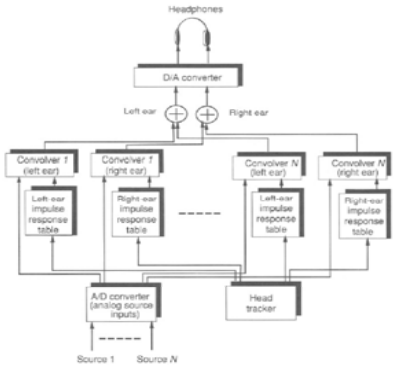
- Perceived loudness
  - Prior knowledge of a given sound source
  - Faint siren (which is a normally high-energy sound source) is perceived as being distant
  - Clear whisper (which is a normally faint sound source) is perceived as cing close
- Motion parallax
  - Change in sound source azimuth when the user is moving
  - Large motion parallax indicates a source nearby
- Ratio between direct and reflected sound
  - Energy of the direct sound drops off with the square of the source range
  - Energy of the reflected sound does not change much with range

Head-Related Transfer Functions



- HRTF captures all of the physical cues to sound localization.
- Experimental measurement of transfer function.
  - Sounds from speakers at different locations
  - Tiny microphones in the ears
  - Analysis of recordings from both ears
  - Head-Related Impulse Responses (HRIRs)
  - Head-Related Transfer Functions (HRTFs)
- Each individual has his/her HRTF signature, also called ear print.

Convolvotron



- Crystal River Engineering Inc Beachtron, Convolvotron, Acoustictron Audio
- HRTF-based spatial audio system
- The system can be customized for a particular individual by measuring and using that person's HRTF
- Echoes and room reverberation can be added by including a room simulation model
- Head motion can be accounted for by combining the absolute location of the source with the outputs of a head tracker to select the appropriate HRTFs.

References

- http://www.evl.uic.edu/spiff/class/cs426/Notes/Leigh_Sound2005.ppt
- http://ajou.ac.kr/~tetross/teaching/Sens.Percpt.08.Sound.Hearing.pdf
- Chiptune music http://en.wikipedia.org/wiki/Chiptune