



IMGD 3xxx - HCI for Real,
Virtual, and Teleoperated
Environments:
Human Hearing and Audio
Display Technologies

by
Robert W. Lindeman
gogo@wpi.edu

Motivation

- Most of the focus in gaming is on the visual feel
 - GPUs (Nvidia & ATI) continue to drive the field
 - Gamers want more
 - More realism
 - More complexity
 - More speed
- Sound can significantly enhance realism
 - Example: Mood music in horror games

Audio Displays

- Spatialization vs. Localization
- *Spatialization* is the processing of sound signals to make them emanate from a point in space
 - This is a *technical* topic
- *Localization* is the ability of people to identify the source position of a sound
 - This is a *human* topic, i.e., some people are better at it than others.

Audio Display Properties

Presentation Properties

- Number of channels
- Sound stage
- Localization
- Masking
- Amplification

Logistical Properties

- Noise pollution
- User mobility
- Interface with tracking
- Environmental requirements
- Integration
- Portability
- Throughput
- Cumber
- Safety
- Cost

Channels & Masking

- Number of channels
 - Stereo vs. mono vs. quadrophonic
 - 2.1, 5.1, 7.1
- Two kinds of masking
 - Louder sounds mask softer ones
 - We have too many things vying for our audio attention these days!
 - Physical objects mask sound signals
 - Happens with speakers, but not with headphones

Audio Displays: Head-worn



Ear Buds



On Ear



Open Back



Closed



Bone
Conduction

Audio Displays: Room Mounted

- Stereo, 5.1, 7.1
- What is the ".1"?
- Sound cube

Types of Sound

□ Music

- Opening/Closing
- Area-based music
- Function-based music
- Character-based music
- Story-line-based music

□ Speech

- NPC speech
- Your thoughts

□ Non-speech audio

Music in Games

- Opening/closing music
 - Can help set the stage for a game
 - Can be "forever linked" to the game
 - You must remember some...




- Area-based music
 - Each level (or scene) of a game has different music
 - Country vs. city
 - Indoor vs. outdoor

Music in Games (cont.)

- Function-based music
 - Music changes based on what you are doing
 - Fighting
 - Walking around

- This can be a very good cue that someone is attacking
 - If they are behind you, for example

Music in Games (cont.)

- Character-based music
 - Each playable character has his/her own "theme" music
 - Many RPGs use this
 - Film uses this too 

- Story-line-based music
 - As in film
 - Music contains a recurring theme
 - Used for continuity
 - Used to build suspense

Speech

- Player
 - Used to communicate with others
 - Used to hear your own thoughts
- Non-player characters
 - Used to convey information to you/others
- More and more "voice talent" being used
 - Big money
 - Return of radio?
- Often accompanied by subtitles

Non-Speech Audio

- Used to enhance the story
- Similar to Foley artists in film
 - The art of recreating incidental sound effects (such as footsteps) in synchronization with the visual component of a movie. Named after early practitioner **Jack Foley**, foley artists sometimes use bizarre objects and methods to achieve sound effects, *e.g.*, snapping celery to mimic bones being broken. The sounds are often exaggerated for extra effect - fight sequences are almost always accompanied by loud foley-added thuds and slaps.
(Source: www.imdb.com)
- Typically used to mimic (hyper-)reality

Non-Speech Audio (cont.)

□ Some examples:

■ Footsteps

- Vary depending on flooring, shoe type, or gait

■ Explosions:

- Vary depending on what is exploding

■ Bumping into things

- Walls, bushes, *etc.*

■ Objects in the scene





- Vehicles, weapon loading/firing, machinery

■ Animals

■ Anything that works!

Non-Speech Audio (cont.)

□ Real examples

-  ■ The screech of a TIE Fighter is a drastically altered elephant bellow, a woman screaming, and more
-  ■ Wookiee sounds are constructed out of walrus and other animal sounds
-  ■ Laser blasts are taken from the sound of a hammer on an antenna tower guide wire
-  ■ Light saber hum taken from a TV set and an old 35 mm projector to create the hum

<http://www.filmsound.org/starwars/#burtt>

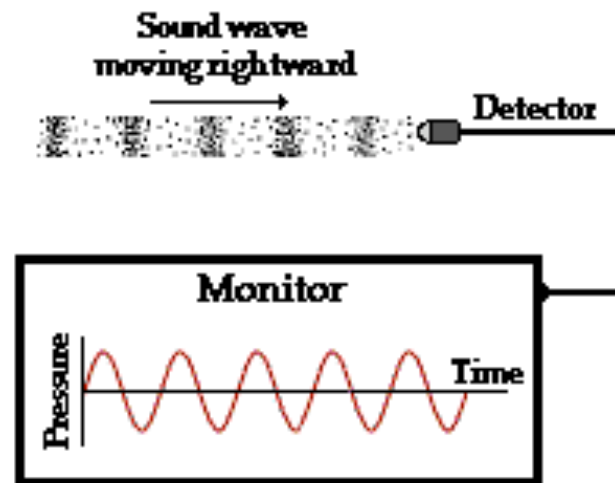
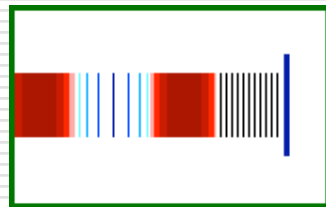
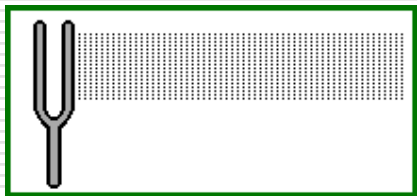


Non-Speech Audio (cont.)

- State of the character
 - Breathing, heartbeat
- Synchronized spatialized video and audio can increase immersion
- Confirmation of user action
 - Reload
 - Menu-item “ping”
 - Unlock a door

Structure of Sound

- Made up of pressure waves in the air

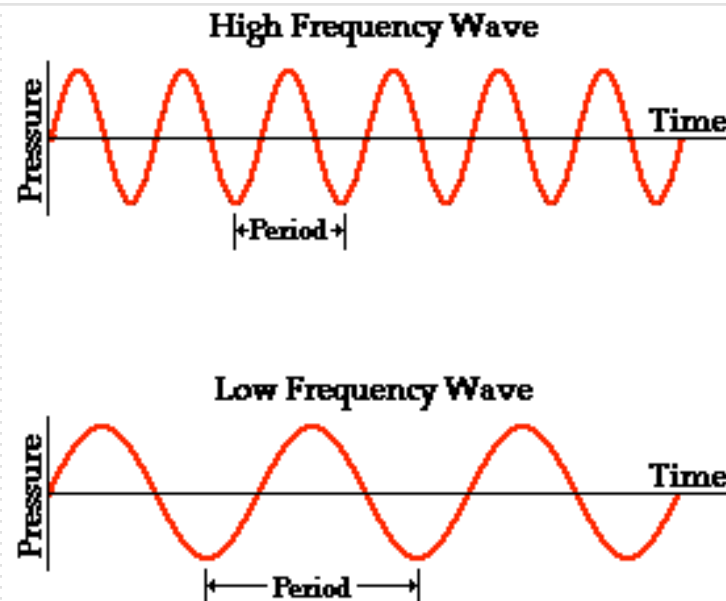


- Sound is a **longitudinal wave**
 - Vibration is in the same direction (or opposite) of travel

(<http://www.glenbrook.k12.il.us/GBSSCI/PHYS/CLASS/sound/soundtoc.html>)

Frequency and Amplitude

- Frequency determines the *pitch* of the sound
- Amplitude relates to intensity of the sound
 - Loudness is a subjective measure of intensity
- High frequency = short period
- Low frequency = long period

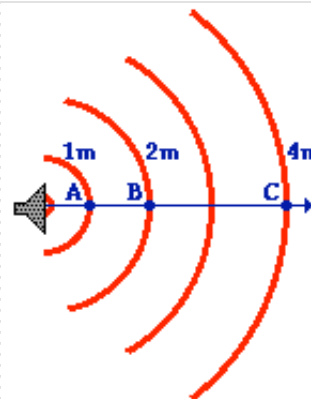


Distance to Listener

- Relationship between sound intensity and distance to the listener

Inverse-square law

- The intensity varies inversely with the square of the distance from the source. So if the distance from the source is doubled (increased by a factor of 2), then the intensity is quartered (decreased by a factor of 4).



Audio Processing

- Audio is made up of a *source* and a *listener*
- Music is typically source-less
 - May be 5.1 surround sound, *etc.*
- Sound undergoes changes as it travels from source to listener
 - Reflects off of objects
 - Absorbed by objects
 - Occluded by objects
- Does this sound familiar?

Audio Processing (cont.)

- Just like light, different materials affect different parts of a sound signal
 - Low frequencies vs. high frequencies
- We can trace the path of sound from source to listener just like we trace light
 - But, we are less tolerant of discontinuities in sound
 - It is more expensive to process "correctly"
- So, we cheat (as always ;-)

Source of Sounds

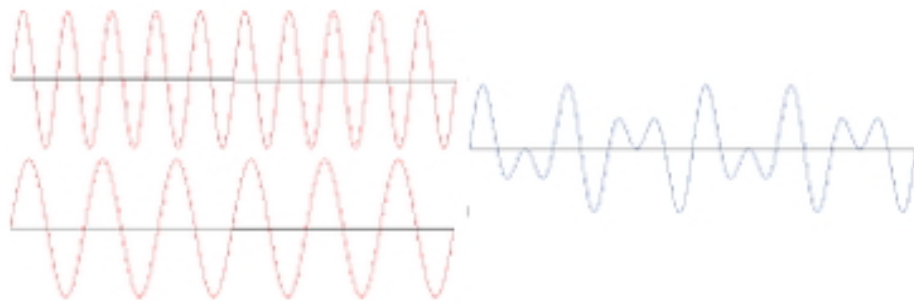
- Like textures, sounds can be captured from nature (*sampled*) or synthesized computationally
 - High-quality sampled sounds are
 - Cheap to play
 - Easy to create realism
 - Expensive to store and load
 - Difficult to manipulate for expressiveness
 - Synthetic sounds are
 - Cheap to store and load
 - Easy to manipulate
 - Expensive to compute before playing
 - Difficult to create realism

Synthetic Sounds

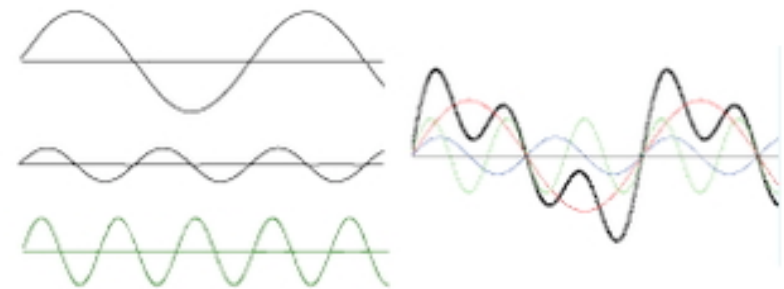
- Complex sounds are built from simple waveforms (*e.g.*, sawtooth, sine) and combined using operators
- Waveform parameters (frequency, amplitude) could be taken from motion data, such as object velocity
- Can combine wave forms in various ways
 - This is what classic synthesizers do
- Works well for many non-speech sounds

Combining Wave Forms

- Adding up waves creates new waves



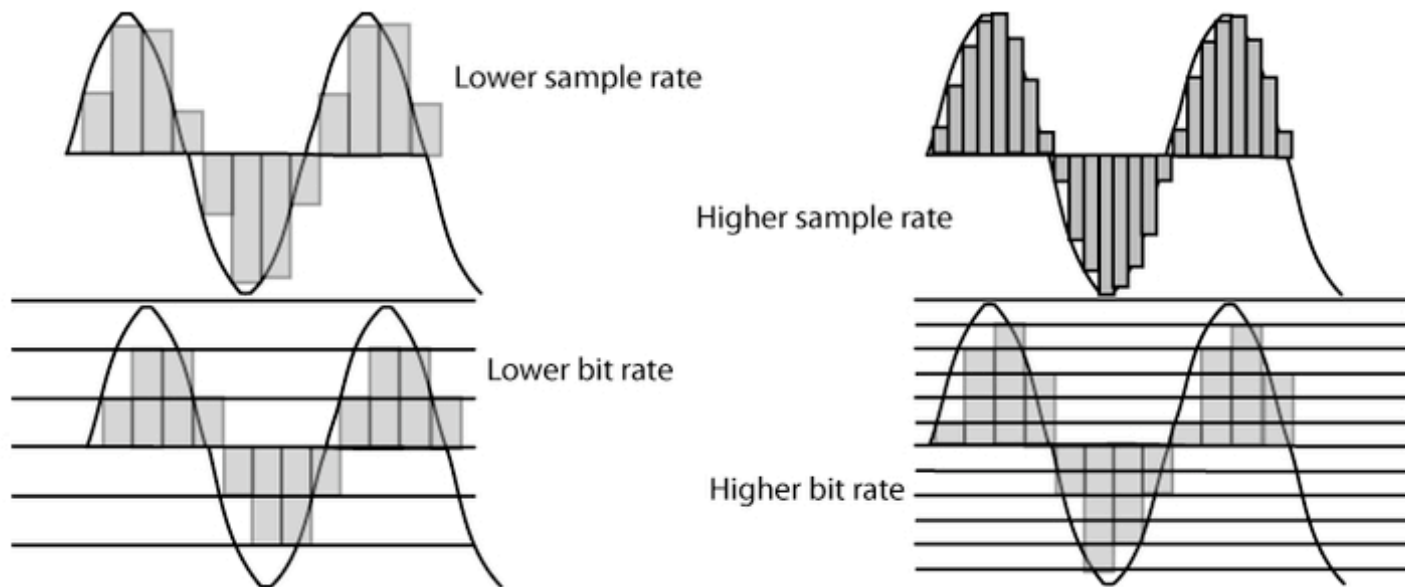
Two tones combined



Three tones combined

Sampling Rates and Bit Rates

- Analog signals need to be translated into digital ones
 - Actually, analog is better in terms of quality!
 - Digital is easier to handle (manipulate)



Spatialized Audio Effects

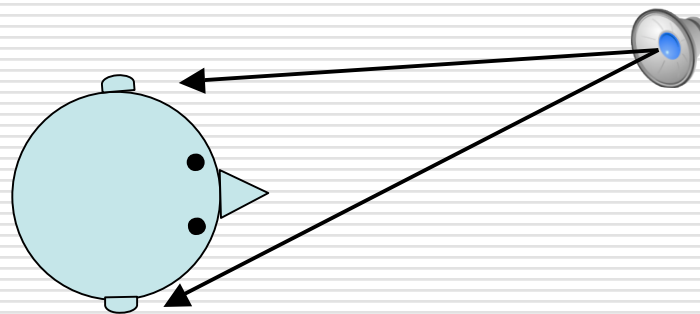
- Naïve approach
 - Simple left/right shift for lateral position
 - Amplitude adjustment for distance
- Easy to produce using commodity hardware/software
- Does not give us "true" realism in sound
 - No up/down or front/back cues
- We can use multiple speakers for this
 - Surround the user with speakers
 - Send different sound signals to each one

Spatialized Audio Effects (cont.)

- What is Dolby 5.1 surround sound?
- We hear with two ears
 - So, why is 5.1 (or 7.1) sound needed?!?!
- If we can correctly model how sound reaches our ears, we should be able to reproduce sounds from arbitrary locations in space
- Much work was done in 1990s on this

Head-Related Transfer Functions

- A.k.a. HRTFs
- A set of functions that model how sound from a source at a known location reaches the eardrum



Constructing HRTFs

- ❑ Small microphones placed into ear canals
- ❑ Subject sits in an anechoic chamber
 - Can use a mannequin's head instead
- ❑ Sounds played from a large number of known locations around the chamber
- ❑ Functions are constructed for this data
- ❑ Sound signal is filtered through inverse functions to place the sound at the desired source

More About HRTFs

- Functions take into account, for example,
 - Individual ear shape
 - Slope of shoulders
 - Head shape
- So, each person has his/her own HRTF!
 - Need to have a parameterizable HRTFs
- Some sound cards/APIs allow you to specify an HRTF to use
- Check Wikipedia or Google for more info!

Environmental Effects

- Sound is also influenced by objects in the environment
 - Can reverberate off of reflective objects
 - Can be absorbed by objects
 - Can be occluded by objects

- Doppler shift

The Tough Part

- All of this takes a lot of processing
- Need to keep track of
 - Multiple (possibly moving) sound sources
 - Path of sounds through a dynamic environment
 - Position and orientation of listener(s)
- Most sound cards only support a limited number of spatialized sound channels
- Increasingly complex geometry increases load on audio system as well as visuals
 - That's why we fake it ;-)
- GPUs might change this too!