Auditory Perception & Virtual Environments

COSC 6002 - Directed Readings in Virtual Reality: Technology & Perception

Bill Kapralos

January 27, 2003

UNIVERSITÉ YORK UNIVERSITY

Overview (1):

- **Introduction**
  - Importance of sound localization
  - Sound in a virtual environment (VE)
  - What exactly is sound?
  - Sound localization & auditory distance perception
    - Interaural cues (ITD, ILD), HRTF, reverberation
- **Recording Techniques (History of 3D Sound)**
  - Two channel stereo
  - Surround sound

Overview (2):

- **Simulating Audio Localization Cues in a VE**
  - Modeling ITD
  - Binaural audio, HRTF measurement and synthesis
  - Reverberation and modeling of room acoustics
  - Auditory distance simulation
- **Sound Output: Headphones vs. Loudspeakers**
  - Headphone listening
  - Transaural audio and crosstalk cancellation
- **Conclusions & Comments**

Introduction

Importance of Sound Localization:

- **Hearing Provides info About our Environment**
  - Spatial sounds give detailed info of our surroundings
    - Determine direction and distance to objects
    - Warn of approaching dangers e.g. predators
  - Unlike vision, hearing is omni -directional
    - Can hear in complete darkness!
  - Can guide the more "finely tuned" visual attention
    - Eases the burden of the visual system

Spatial Sound in a VE (1):

- **Importance of Spatial Audio in a VE**
  - Conveys basic info. to the the users
    - e.g. footsteps in small room vs. outside (large field)
  - Allows users to orient themselves
  - Increases situational awareness
  - Maintains a sense of environmental realism
    - Helps increase immersion and hence presence
  - Can enhance perception of video quality
  - Can provide a sense of ambience - mood and emotion

Spatial Sound in a VE (2):

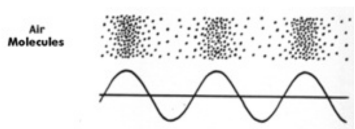
- **Spatial Audio Often Ignored in a VE**
  - When present, typically:
    - Cues are poor and don't always reflect natural spatial cues
    - "Far field" acoustical model assumed – source at infinity, plane waves
  - Emphasis typically placed on visual senses
    - Graphics
    - Stereo vision etc…

Implementing Audio Cues in a VE:

- **Imitate Human Sound Localization Cues:**
  - ITD ? Interaural Time Delay
  - ILD (IID) ? Interaural Level (Intensity) Difference
  - HRTF ? Head Related Transfer Function
  - Reverberation
  - Vision (e.g. easily localize a sound source we can see)
- **3D Audio by Simulating Human Cues**
  - ITD & ILD alone are simple but limited – ambiguous
  - HRTFs improve localization and reduce ambiguities

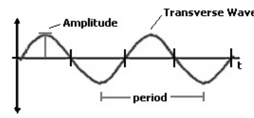
Physics of Sound:

- **What Exactly is Sound?**
  - Variations in air pressure caused by vibrating object
    - Guitar string, tuning fork, vocal chords, etc…
  - Alternating regions of compression and rarefaction of the air (medium) molecules



A Simple Sound Wave:

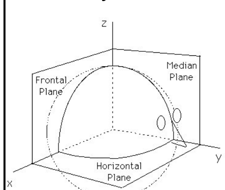
- **Sinusoidal (Sine) Wave**
  - Known as a tone or pure tone
  - Simple response
    - Simple analysis



- **Not Typically Encountered in Normal Listening**
  - Complex tones instead
- **Fourier Analysis**
  - Complex tone ? superposition of sinusoids!

Coordinate System (1):

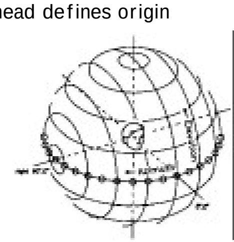
- **"Head Centered" Rectangular System**
  - Center of the head defines origin
  - x, y, z axis



- **Three Planes of Interest:**
  - Median: right/left separation
  - Frontal: front/back separation
  - Horizontal: up/down separation
- **Interaural Axis:**
  - "Line" passing through ears

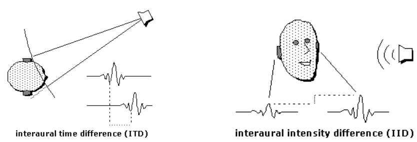
Coordinate System (2):

- **Spherical Coordinate System:**
  - Azimuth (?), elevation (?) and range (r) to specify coordinates
  - Center of head defines origin



Sound Localization (1):

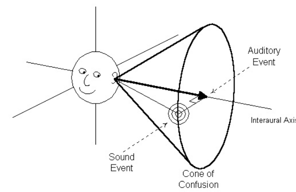
- **Lord Raleigh's Duplex Theory (1907):**
  - Assumes a spherical head with no pinnae
  - ITD: difference in time between arrival of sound at each ear - Low frequencies: < 1500Hz
  - ILD: intensity difference between sound at each ear - High frequencies: > 1500Hz



13

Sound Localization (2):

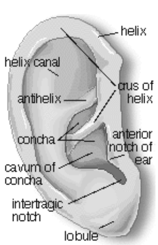
- **Shortcomings of the Duplex Theory**
  - Front-back Ambiguities
  - Cone of confusion



14

Sound Localization (3):

- **Head Related Transfer Function (HRTF)**
  - Filtering of sound spectrum by interactions of sound with head, torso and particularly pinna

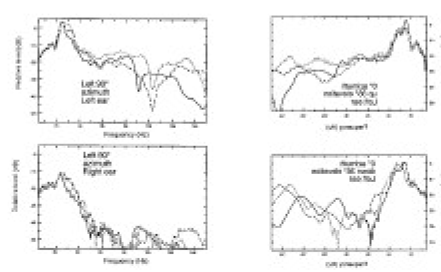


- **Pinna:**
  - Series of grooves and notches which accentuate or suppress mid & high frequency components in a position dependant manner
  - Each person's pinna differs ? filtering effects differ

15

Sound Localization (4):

- **Sample HRTFs:**

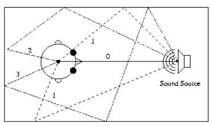


? = 90°, ? = 0° ? = 0°, ? = 36°

16

Sound Localization (5):

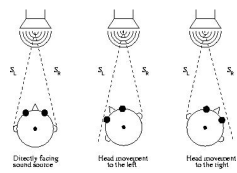
- **Reverberation**
  - Typical environments are rarely anechoic!
  - Sound waves interact with objects in the environment
    - Portion of sound waves are reflected and absorbed
    - Portion absorbed by the medium itself (e.g. air)
  - Collection of reflected waves (possibly 1000s) is known as reverberation

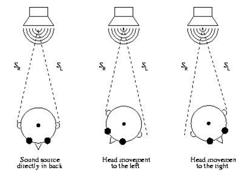


Sound Localization (6):

- **Dynamic Cues**
  - Head movements to resolve front-back ambiguities

Source directly in front





Source directly in back

Auditory Distance Perception:

- **Relative & Absolute Distance Cues**
- **Main Auditory Distance Cues:**
  - Sound level (intensity): inverse square law
  - Reverberation: ratio of direct to reverberant energy
  - Sound source frequency spectrum: attenuation of high frequency components as a function of distance
  - Binaural: distance dependence of ILD for sources in the "near field"
  - Source characteristics: spectrum composition etc.

Recording Techniques:

A Brief History of "3D" Sound

Listener "Sweet Spot":

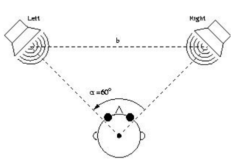
- **Auditory Effect Restricted to Small Region**
  - Listener must be placed in specific location relative to the loudspeakers
  - Even small movements can seriously degrade effect!
  - Common to all loudspeaker based systems
    - "True" and "non-true" 3D sound systems
  - Dependent on:
    - Technique used
    - Number of loudspeakers
    - Characteristics of loudspeakers

Monaural Techniques:

- **Single Microphone to Record Sound Event**
  - Recording conveyed using single loudspeaker
  - First method used to convey sound in films
- **Shortcomings:**
  - Cannot capture any binaural cues
  - Difficult to convey any ambience
  - Differentiating noise from signal of interest
- **Still Relevant Today**
  - Telephone – from 1876 to present

Stereophonic (Stereo) Techniques (1):

- **Synonymous with Two Channel Audio**
  - Actually refers to construction of believable, solid, stable sound "images" regardless # of channels
    - Stereo and "surround sound" refer to same thing!

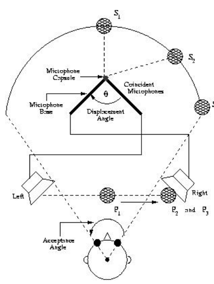


- **Stereo Setup:**
  - Listener and loudspeakers form equilateral triangle
  - Virtual source positioned between loudspeakers

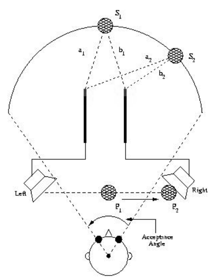
Stereophonic (Stereo) Techniques (2):

- **Artificial Techniques:**
  - Artificial adjustment of time and/or intensity delays
- **Coincident Microphone Techniques:**
  - Two microphones placed as physically close as possible to eliminate timing differences
  - Intensity differences position source
- **Spaced Microphone Techniques:**
  - Two microphones spaced some distance apart
  - Timing differences position source

Stereophonic (Stereo) Techniques (3):



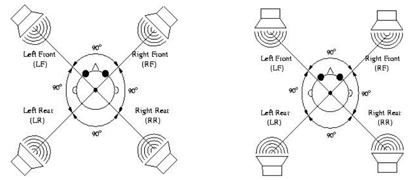
Coincident



Spaced

Surround Sound (1):

- **Quadraphonic ("Quad") - Four Channel:**
  - Two in front of listener (as in stereo)
  - Two in back of listener
  - Intended to allow positioning of source anywhere (360^o) on plane which loudspeakers are placed



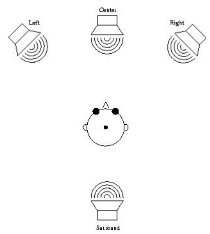
Surround Sound (2):

- Quad microphone:
  - Four microphone elements to capture sound event
  - One microphone per loudspeaker
- Matrixing:
  - Quad recordings consisted of four channels but two channel stereo dominated!
  - Encode four channels into two to use existing stereo transmission medium
  - Decode back into four channels for quad playback

Surround Sound (3):

- **Ambisonics**
  - Conceived as a system capable of accurate 3D sound
  - Capable of encoding sounds in any azimuth direction and vertically
  - "Special" microphone to record sound event
  - Flexible loudspeaker placement
    - Variable number of loudspeakers
    - Only requires length vs. width ratio of 2:1
    - No one-to-one mapping between mics and loudspeakers ? bigger sweet spot!

Surround Sound (4):

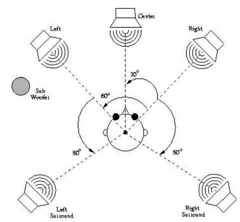


- **Dolby Stereo**
  - Four loudspeakers arranged in diamond shape
    - Front-left, center, front-right & surround
  - Encoding/decoding for two-channel compatibility

29

Surround Sound (5):

- **Dolby Digital**
  - Digital format
  - Three frontal channels, two or more independent surround channels & low frequency effects channel



- Perceptual coding
  - Auditory masking
- Dolby 5.1
  - Very popular – has become synonymous with surround sound!

30

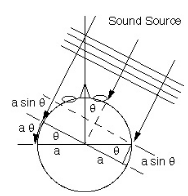
Simulating Human Sound Localization in a Virtual Environment

31

Modeling the ITD (1):

- **Woodworth's Model:**
  - ITD for sound located on azimuth plane
  - Assumes spherical head model, no pinnae

$$ITD = \frac{a}{c}(\theta + \sin\theta) \quad , \; -90^\circ \leq \theta \leq +90^\circ$$



- Valid for high frequency signals
  - Angular frequencies > a/c

32

Modeling the ITD (2):

- **Kuhn's Model**
  - ITD for sound located on azimuth plane
  - Based on ITD values obtained from "dummy" head
  - Separate expression for high and low frequencies
- **Larcher & Jot's Model**
  - Spherical head model to predict ITD values
  - Accounts for source elevation
- **Duda's Ellipsoidal Head Model**
  - 5 parameters measured from listener's head

Binaural Audio (1):

- **Recreate a Particular Sound Event**
  - Reproduce signals at each ear as they would be in original environment
- **Binaural Recordings**
  - Small mics inserted in ear canal of person or "dummy" head to record sound event
    - Captures audio cues – ITD, ILD, reverb., HRTF…
    - Headphone play back ? recreate original event
  - Specific to environment which they were made, including source and listener positions

Binaural Audio (2):

- **Binaural Synthesis**
  - Imitate binaural recordings
    - "Process" monaural sound with pair of measured HRTFs corresponding to desired source position
    - Specific to one particular source/listener position
    - Can also add any environmental effects (reverberation etc.)
    - HRTF measurement has its share of problems!!!

HRTF Measurement (1):

- **Assume HRTF can be Modeled by LTI system**
  - To measure HRTF for position **p** relative to listener:
    - Anechoic environment
    - Probe mics inserted in each ear of listener
    - Output "impulse" signal from speaker placed at **p**
    - Measure response at left and right ear
- **Problems with HRTF Measurement Process**
  - Each position in 3D space ? unique HRTF
    - Can only sample at discrete locations
    - Long, tedious process, specialized equipment

HRTF Measurement (2):

- **Generic (Non-individualized) HRTFs**
  - HRTF for position **p** differs for each individual but impractical to measure HRTF for each user
  - Instead of "individualized" HRTFs, use "non-individualized" HRTFs measured from:
    - Anthropomorphic dummy (e.g. KEMAR)
    - Above average listener
    - Average response of several listeners
  - Several HRTF libraries are available
    - MIT: KEMAR measurements
    - CIPIC: averaged – individuals + KEMAR

HRTF Synthesis:

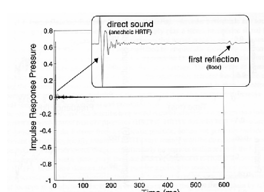
- **Measured HRTFs Form FIR Filter Coefficients**
  - Signal delivered to left & right ear:
    - Filter monaural sound with corresponding response
    - When presented to listener, impression of sound at position **p** is obtained
  - Can add reverberation cues
- **Performance Will Suffer**
  - Non-individualized HRTFs,
  - Handling non-sampled positions - interpolation

Reverberation:

- **Two Techniques to Add Reverberation:**
  - Artificial
    - Present listener with delayed & attenuated versions of sound source
    - Don't necessarily reflect physical properties
  - Auralization
    - Recreate a particular listening environment
    - Determine "exact" reflection patterns of sound waves using physical or mathematical modeling

Auralization (1):

- **Binaural Room Impulse Response (BRIR)**
  - Response of a particular room (environment)
  - Measured in similar manner to HRTF or modeled
  - Filtering sound with left & right BRIR response recreates environment



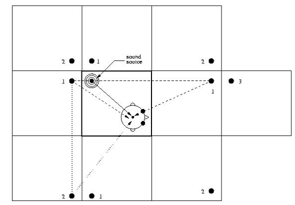
Actual BRIR of "standard classroom"

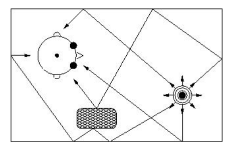
Auralization (2):

- **Modeling of the BRIR:**
  - Acoustic scale modeling
    - 3D scaled models
  - Computer & mathematical modeling
    - Wave based: solve wave equation (Helmoholtz equation) to recreate some soundfield
    - Ray based: Trace the paths followed by sound waves emitted by the source

Auralization (3):

Image Source Method

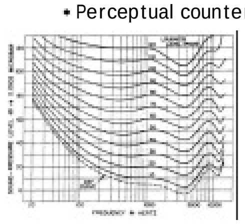




Ray Tracing Method

Distance Simulation (1):

- **Sound Level (Intensity) Scaling**
  - Scale sound source level following inverse square law
    - Simple, intuitive
    - Inverse square assumes spherical head, no pinna
    - Perceptual counterpart of intensity ? loudness!



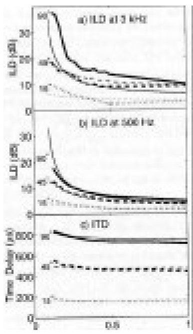
- Equal loudness contours
  - Loudness is frequency dependent
  - Lower frequency tones not as loud as higher frequency tones

Distance Simulation (2):

- **Reverberation**
  - Ratio of direct to indirect sound level
    - Automatically given when proper reverberation model (simulation) in place
- **Binaural Cues**
  - May not be required
    - Effect of binaural cues on source distance still open issue!
    - ILD can change significantly at very close distances

Distance Simulation (3):

- ILD and ITD as a function of distance

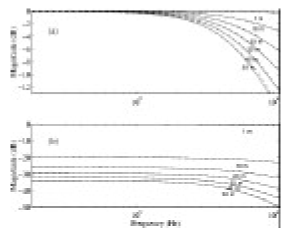


- ILD highly dependent on distance for source close by (within 0.5m) especially away from median plane
- ITD dependence is much smaller

45

Distance Simulation (4):

- **Source Spectral Content**
  - Relative cue to source distance
  - High frequency attenuation for large distance (> 15m)
  - Rather "weak" cue (see Begault)



- Analytical expressions to predict absorption as function of humidity, temp., frequency & distance

46

Sound Output:

Headphones vs. Loudspeakers

47

Headphone Based Displays (1):

- **Advantages:**
  - High channel separation ? minimal (if any) crosstalk
  - Isolate listener from external sounds & reverb.
    - Room acoustics or listener's position don't affect listener's perception
  - Sometimes, only means available for delivering audio
    - Aircraft cockpits
    - Multiple user environments
    - Where loudspeakers are impractical

48

Headphone Based Displays (2):

- **Disadvantages:**
  - Inside-the-head Localization (IHL)
    - Sounds appear as if they originate inside the head
    - Lack of externalization
  - Comfortability
    - Can be cumbersome
    - Impractical at times & can limit immersion
  - Greater rate of localization ambiguities
    - Front-back confusions
    - May move while listener is wearing them

49

Headphone Based Displays (3):

- **Overcoming IHL & Ambiguities:**
  - Provide listener with "realistic spectral profile of the sound at each ear" ? HRTF, reverberation
  - Head movements – dynamic cues
  - These techniques have their share of problems & may offset any improvements

50

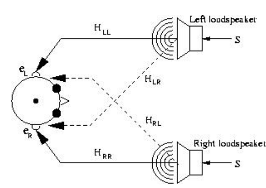
Transaural Audio (1):

- **Original Binaural Audio Meant for Headphones**
  - Ensured isolation between left/right signal
- **Transaural Audio:**
  - Binaural audio over loudspeakers
  - Left/right binaural signals to left/right loudspeakers
  - Overcome limitations inherent to headphone displays
    - IHL
  - Introduce new major problem
    - Crosstalk

51

Transaural Audio (2):

- **Crosstalk**
  - Non-isolation of left/right signals delivered to ears
    - Left/right signal heard by right/left ear
    - Delayed & attenuated version of ipsilateral signal arriving at contralateral ear



52

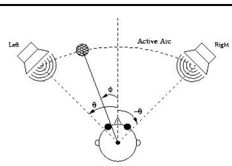
Transaural Audio (3):

- **Crosstalk Cancellation to Remove Crosstalk**
  - Add delayed and inverted version of crosstalk signal to opposite loudspeaker output
  - Matrix (frequency domain) solution
    - Single listener
    - Can generalize to ***N*** listeners (at least in theory)!
- **Problems with Crosstalk Cancellation**
  - In theory – great! But in practice:
    - Rely on HRTFs – HRTFs have their own problems!
    - Small sweet spot ? listener must be tracked

Amplitude Panning (1):

- **Make use of ILD**
  - Adjust amplitude gain of each loudspeaker signal in a manner to simulate directional properties of ILD
    - Listener perceives virtual image emanating from some position dependant on gain factors
  - Various amplitude panning techniques available
    - 2D & 3D loudspeaker configurations

Amplitude Panning (2):



- **2D Amplitude Panning**
  - Two channel stereo
    - Simplest configuration
    - Position sound source between two loudspeakers
    - Various trigonometric methods to compute gain factors
    - > 2 loudspeakers, all positioned on same plane
    - Pair-wise amplitude panning ? sound applied to 2 loudspeakers only

Amplitude Panning (3):

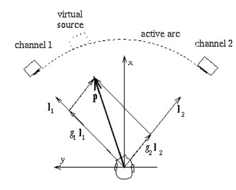
- **3D Amplitude Panning**
  - > 2 loudspeakers – non-coplanar
  - Loudspeakers typically equidistant to listener
  - Similar to pair-wise panning
    - Sound applied to a subset of loudspeakers only
  - No general trigonometric solution for arbitrary 3D loudspeaker configurations
    - Gain factor calculation very configuration dependant

Amplitude Panning (4):

- **Vector Base Amplitude Panning (VBAP)**
  - Arbitrary number of loudspeakers - 2D & 3D
    - Sound presented to 1, 2 or 3 loudspeakers only
    - Remaining loudspeakers can provide reflections or diffuse sound (e.g. reverberation)
  - Loudspeakers can be placed in any position
    - Almost equidistant to listener
  - Listening room must not be too reverberant

Amplitude Panning (5):

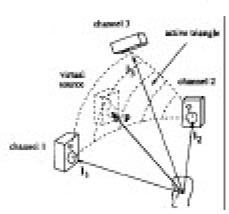
- 2D VBAP
  - Two channel setup treated as 2D vector base
  - Two unit vectors l_1, l_2 pointing to each loudspeaker
  - Unit vector pointing to virtual source is sum of weighted loudspeaker vectors: $p = g_1 l_1 + g_2 l_2$
  - Can then solve for weights g_1, g_2



- Virtual source positioned anywhere on “active arc” between the two loudspeakers

Amplitude Panning (6):

- 3D VBAP
  - Unit vectors to 3 loudspeakers (triangle)
  - Unit vector pointing to source is linear combination of these 3 unit vectors
  - Solve for weights in a similar manner as 2D VBAP



- Virtual source positioned anywhere on “active triangle” between the two loudspeakers

Conclusions

Summary (1):

- **Spatial Audio Important in any VE!**
  - Increases environmental awareness
  - Increases immersion and therefore presence
  - Can enhance perception of video quality
  - At the very least, adds ambience (more "lively")
- **Creating Spatial Audio Displays is Difficult!**
  - Potentially:
    - Complex
    - Computationally intensive

Summary (2):

- **For "True" 3D Sound Simulate Human Cues**
  - Binaural Cues
    - Simple to implement
    - Limited to horizontal plane localization
    - Ambiguous: cone of confusion, front-back

Summary (3):

- HRTFs
  - Overcome limitations inherent with binaural cues
  - Externalize a sound source when using headphones
  - "True" 3D sound capability
  - Large variation amongst individuals
  - Difficult to measure: time consuming, tedious, specialized equipment (e.g. anechoic chamber)
  - Potentially large storage capacity
  - Interpolation required for non-sampled positions

Comments:

- **Finally, How Much "Reality" is Required?**
  - Depending on intended application, realism may not be most important consideration
    - Reverb provides more accurate distance judgments but reverb may reduce directional accuracy
    - In entertainment applications complete room acoustics modeling is not necessary – simple artificial reverb techniques usually suffice and save $$$
    - Headphone vs. loudspeaker based systems

Finally... The End!

65