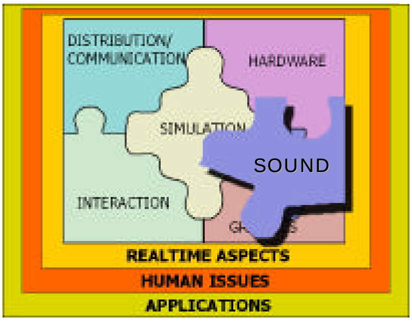
VR - Sound



The need of audio in VR

- In Virtual Reality the display has to produce the immersion in a 3D computer generated world. In addition to the visual and tactile displays, audio is needed. In VR:
  - Sound enhances the presence of the simulation.
  - It enhances the display of spatial information (particularly space beyond the field of view). This includes acoustical properties of the environment and the geometrical relationships of the source and listener
  - Data-driven sound can convey simulated properties of the constituents of the environment. For instance mass, force of impact...
  - Audio signals provide a higher degree of temporal resolution than visual display.
  - Audio is also another channel of communication between the user and the environment.
  - Useful in designing systems where a user monitor several communication channels at once ('cocktail-party effect')

Auralization or audio in 3D.

- Principles.
  - **Auralization** is the process of rendering audio data by digital means to achieve a 3D sound space.
  - Principle: binaural human hearing. (Binaural = hear using two ears.)
  - From the two signals we extract information about the location of sound sources.
  - The research in psycho-acoustic provides the necessary findings that today's 3D audio systems are based on.
  - Multidimensional audio reproduction is possible with a few speakers.
- The types of displaying audio differ in :
  - size of the listening area (sweet spot),
  - amount of spreading of stereo images,
  - amount of coloring (tonal changes),
  - amount of stereo left/right information that is lost during processing,
  - ability to achieve effect on headphones as well as speakers.

Auralization or audio in 3D.

- 3D audio is not:
  - Extended Stereo, where an existing stereo (two channel) soundtrack add spaciousness and make it appear to originate from outside the left/right speaker locations.
  - Surround sound, where another dimension front/back is added to the stereo.
- What we need is: 3D interactive sound system
  - Sounds are projected in all three dimensions (left/right,up/down,front/back) in real-time with the less coloring (tonal changes) introduced by processing .

The Basics of *Real* Acoustics

- The acoustics of a *real* world has three components:
  - **The sound source**: Object that emits sound waves. These waves are created through a variety of mechanical processes.
  - **The acoustic environment:** In the medium sound waves are absorbed, reflected, refracted and diffracted in different ways depending on their frequency, and material and geometry of the environment.

    Reflections are multiple. Humans are capable of individually perceiving first order reflections; second and higher order reflections usually combine in reverberations.
  - **The listener:** Sound receiving object, typically a 'pair of ears'. From the arriving waves the listener extract information about the sound sources and the environment. In case of relative motion source-listener Doppler effect is present.

The Basics of *VR* Acoustics

- In the *virtual* world we can talk of an auditory world, which reproduced the characteristics of the real, and it is composed by:
  - **The auditory actor:** Entity emitting sounds through its interface.
  - **The auditory space:** The environment that has to be modeled. Sound is reflected, attenuated, refracted and also diffract around objects. An auditory space object models the geometry of the enclosures in the world.
  - **The listener:** They move around and orientate in the world, capable of perceiving the sounds.

The Basics of Human Hearing

- Our auditory system consists of two ears and a brain.
- The two primary localization cues are :
  - **Inearaural intensity difference (IID)** IID refers to the fact that a sound is louder at the ear that it is closer to, because the sound intensity at that ear will be higher than the intensity at the other ear (signal that has been shadowed by the listener head).
  - **Interaural time difference (ITD).** ITD means that a sound will arrive earlier at one ear than the other.

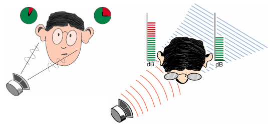


Fig1.ITD & IID representations.

The Basics of Human Hearing

- ITD and IID are not sufficient to precise location .
  - Before a sound wave gets to the ear drum, it passes through the outer ear structure, called the pinnae. The two pinnae act as variable filter that affects every sound that passes through them. By this information the brain knows how to figure out the exact location of a sound in space.
  - The pinnae are the key to accurately localizing sounds in space for wavelengths in the centimeter range or smaller.
  - In general the higher the frequency of a sound, the shorter its wavelength, and the better it can be localized.

3D Audio Reproduction

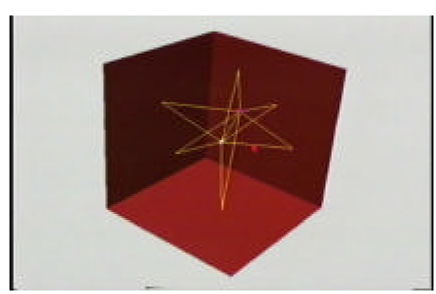
- A 3D audio system aims to digitally reproduce a realistic sound field. It has to recreate IID,ITD, outer ear effects...
- It is necessary to capture the listening cues by analyzing what happens to a single sound as it arrives at a listener from different angles: the parameters Head-Related Transfer Functions (**HRTF)** account for all these acoustical effects.
- For measuring the HRTF of an individual we place a speaker at a known location relative to the listener, play a known signal through that speaker, and record the microphone signals.
- Every individual has a unique set of HRTFs, also called an ear print (they are interchangeable in most applications). Then, with software this parameters are processed and filtered for reproducing signals.
- The listener with headphones cannot tell the difference between the sound from a source and the same sound played back by a computer and filtered by the HRTF corresponding to the original speaker location

Auditory Display

- There are several approaches to the reproduction of auralized audio data. Two main categories:
  - HEADPHONES
    - The HRTFs of a listener are simulated by convolving a filter with the original signal, producing the effect that they would have on a sound emanating from a particular direction. The summed signals of convolved audio streams are fed into the headphones.
      - Advantages: Stability of the listener relative to the headphone placement and the low costing environment.
      - Problems: Specific to every individual and instability of the measurements.
  - LOUDSPEAKERS (stereophonic arrays)
    - Placed around the user. The perceived position of the sound is controlled by the amplitude of the signals fed into the individual speakers. This produce 3D effect.
    - To display the sound some effects have to be included as the absorption in the medium, the reflections in the walls, reverberations, Doppler effect...

Environmental simulation

- As sound waves does not only move directly from the source to the listener, it gets colored and diffracted on its way from the source to the listener.
- We have second order reflections due to walls, obstacles, materials etc...
- Lets watch an real-time example of a simulation of just that:
  - A Beam Tracing Approach to Acoustic Modeling for Interactive Virtual Environments.
    - Nicolas Tsingos, Thomas Funkhouser, Addy Ngan, Ingrid Carlbom, Bell Laboratories, Princeton University



Sound generation

- Sound can be sampled and synthetic:
  - Sampled sounds are digital recordings of a sound that are temporally mapped and played back on demand.
  - Synthetic sounds are created from arbitrary parameters. Approaches used for generating sounds: Fourier synthesis, signal multiplication, filtering, frequency modulation, timbre trees (in which parameters of motion will determine the parameters of generation)…Generate sounds is a difficult task.

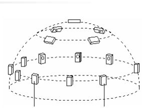
Real-time synthesis of sound-effects

Hardware and software

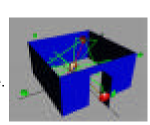
- Convolvotron (Hardware)
  - developed by Foster at the NASA Ames VIEW Lab, and improved at Crystal River Engineering. which later developed the Maxitron. Principal commercial product available for serious auralization work. It combines a control CPU and DSP convolution engine to process audio signals using a library of measured HRTFs.
- DIVA (Software)
  - created by the Helsinki University of Technology
- The Vanilla Sound Server (VSS) (Software)
  - It is a project of NCSA which allows in real-time data -driven sound production to be controlled by interactive applications. It provides an architecture for mapping, coordinating and synchronizing control signal for sounds produced by software synthesis or by peripheral digital audio systems such as MIDI-enabled devices.
- EAX (Software/Hardware)
  - Creative:s sound API.
  - Only positional simulation + added reverberation and sound effect
  - Now also supports some occlusion (in EAX3.0)
- JASS (Java Audio Synthesis System)
  - http://www.cs.ubc.ca/~kvdoel/jass/jass.html

Hardware and software

- Pioneer Sound Field Control System (PSFC) (Hardware)
  - It is a DSP-driven hemispherical 14-loudspeaker array, installed at the University of Aizu Multimedia Center. Co. It is collocated with a large screen rear-projecting stereographic display.



- Aureal (Hardware/software)
  - Vortex1, Vortex2 chips.
  - Manufactures an advanced low-price chip for real-time simulation of spatial environments.
    - Wavetracing
      - Compare to Ray tracing, but for sound.
      - Reflections, occlusions, damping
    - 48 hardware rendered sound-sources at once.
  - A3D - an API for accessing the chip.
  - Out of business, purchased by Creative, future unsure.
  - Most of the employees are working at NVidia, expect things happen there.
    - NForce hardware on Xbox
    - Capable of occlusion, reflection etc, but currently not in use.



Hardware and Software

- AuSim GoldMiner (Hardware/Software)
  - Built upon knowledge (and people) from Aureal.
  - 80 sound sources, 16 listeners
  - 64 digital in, 64 analog in
  - Runs as a separate PC with PCI-cards.
  - Currently only positional simulation, Environmental simulation in the future
- OpenAL
  - Similarly to OpenGL but for sound.
  - Open Source for simulating 3D sound in real-time.
  - Portable, Unix, windows.
  - Still very early in development
- OpenAL++
  - C++ library ontop of OpenAL
  - Will go OpenSource.
  - Eases up usage of OpenAL a lot.

Audio-Visual Synergy

- Together, synchronized 3D audio and 3D visual cues provide a very strong immersion experience.
- 3D audio removes a lot of requirements on the visual part, i.e.. The graphics don't have to be as nice!
- It is needed a single framework that integrates arbitrary sounds and motions for virtual environments.This is done when audio parameters are taken from motion parameters, and motion and audio are synchronized.
- Head Movement and Audio
  - Audio cues change dramatically when a listener tilts or rotates his head. The ears alert the brain about an event outside of the area that the eyes are currently focused on, and we automatically turn to redirect our attention.
  - Interactive audio is much better than prerecorded audio because it allows the listeners head motion be properly simulated.