Virtual Environments (COMP0013)

Spatial Audio for Immersive VEs

David Swapp
Department of Computer Science
University College London

Overview

PART 1: Perception of Spatial Audio

- What is spatial audio?
- Why simulate spatial audio?
- How can spatial audio be simulated?
- Physics of sound
- Psychoacoustics

Overview

PART 2: Synthesis of spatial audio

- Methods for synthesizing spatial audio
- Headphones vs. speaker array
- Hardware & Data requirements
- Environmental acoustic modelling

What Is Spatial Audio?

- “Spatial audio” describes any of a variety of techniques for simulating the 3D soundfield that occurs in a real environment.
- Presentation via headphones or speaker array
- Real environment may contain many audio sources, but the soundfield must be simulated only at 2 ears.
- As well as direct sound waves, reflections & diffractions of these waves from objects in the environment must be accounted for.

Origins of Scientific Study of Spatial Audio

- 1881: Clément Ader placed carbon microphones at either side of the stage of the Paris Grand Opera.
- Telephone lines connected these to listening rooms 3km away at the Paris Exhibition.
- You can try a more recent example of this technique here (use headphones!)

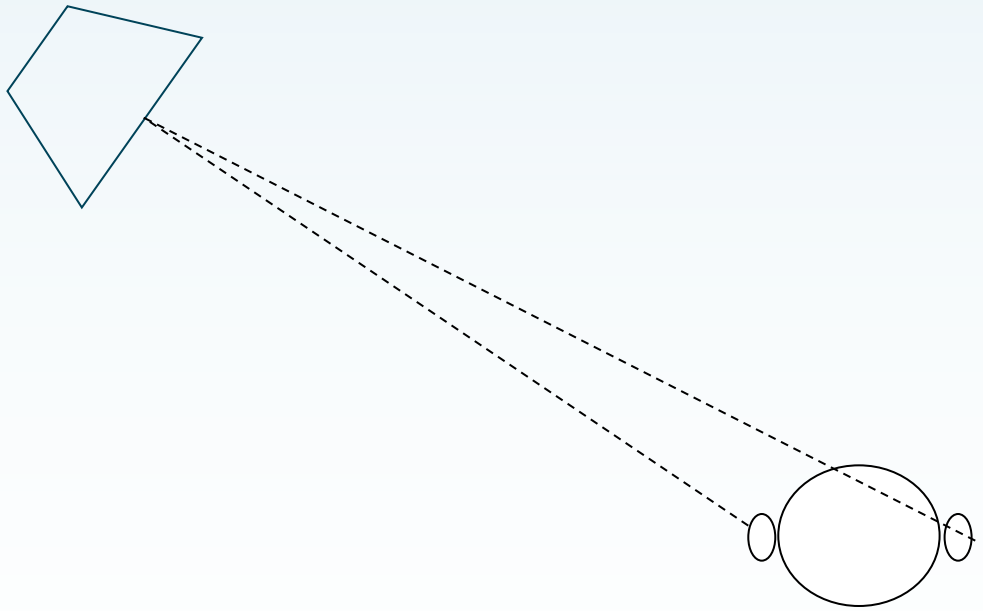
  https://www.youtube.com/watch?v=IUDTlvagjJA&feature=youtu.be

Théâtrophone.

Why Simulate Spatial Audio?

- Important for spatial attention
  - Visual attention often guided by acoustic cues
- Sense of presence
  - Presence reinforced by (coherent) sensory cues
- Applications
  - Architectural
  - Social / Collaborative VR – consider limited visual fov

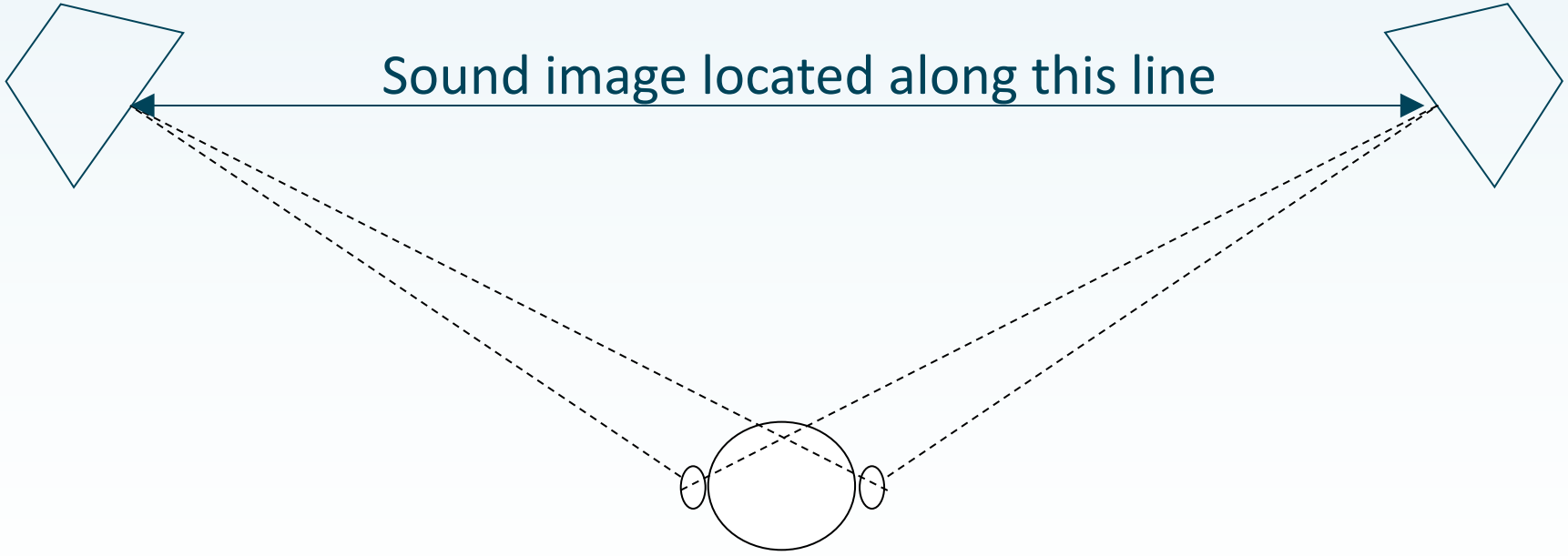
Audio simulation basics: monaural



- Sound comes from direction of loudspeaker
- Single channel
- Contains no directional information
- Distance cues can be simulated

Audio simulation: moving sound image

With two outputs, amplitude panning can produce a moving sound image:



Sound image can appear to move between the 2 speakers

N.B. still only requires a single channel

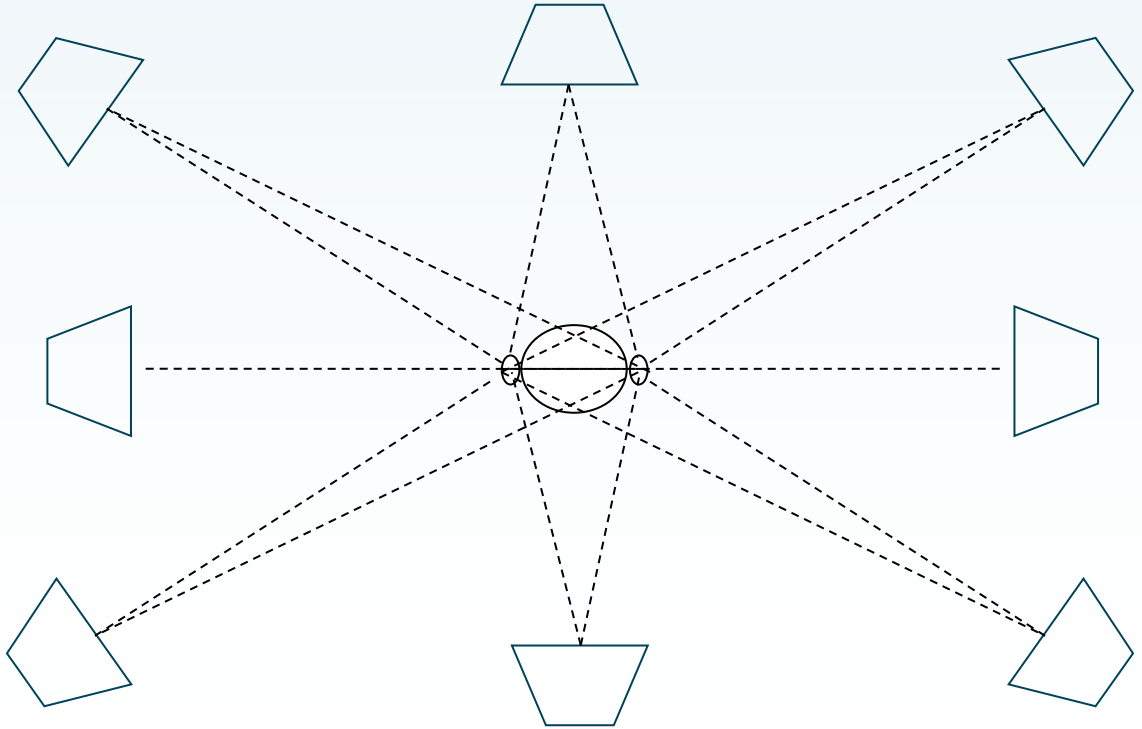
Audio simulation: stereo

2-channel stereo:

- 2 distinct audio channels
- Limited directional information in one spatial dimension via amplitude panning
- Amplitude panning can be achieved within the stereo mix (not just at the output)
- Consequently can simulate any number of sources

Audio simulation: multi-channel

Multi-channel audio: surround sound??



Audio simulation: multichannel

Multi-channel audio:

- Extension of stereo principle to multiple channels and multiple outputs
- Can allow front-back or elevation to be simulated.
- State of the art in computer games and movies can simulate up to 16 channels

Limitations of multi-channel audio

- Relies only on manipulation of audio level (panning) & environmental effects
- Rotating sound source seems to jump from one speaker to another
- Human audio system also uses other cues so should also try to simulate these:
  - **Phase** is used for localisation of lower-frequency sounds (<1.5kHz)
    - Human vocal range is ~60Hz to ~1.5kHz
  - **Level** is better for higher frequencies
  - **Filtering** directionally-dependent influence of our own body on the signal entering our inner ears

Beyond audio panning: reconstructing the soundfield

- Goal is to present sound waves at the listener's ears that would occur in a real environment
- Encode phase as well as amplitude information
- Technically difficult to achieve - no elegant algorithms
- Sound waves do not behave like light waves!
  - Interference at "human" scale
  - Phase matters

Physics of sound

- Comparison with light: some similarities but many differences
- Sound is caused by vibrations of particles in the form of a longitudinal waveform.
- Wave speed varies (~340ms^{-1} in air) – faster in denser media
- Audible frequency range is ~20Hz to 20kHz (17m to 17mm wavelength).
- Waves reflect specularly if surface is much larger than wavelength.
- Waves diffract if surface is approx. same size as the wavelength
  - Consider that most sound comprises multiple frequencies

Physics of sound (2)

- Waves refract if temperature of medium (e.g. air) changes
  - Temperature inversion effect

- Propagation of sound waves is in many ways analogous to propagation of light waves, but there are crucial differences:
  - Lower speed of sound allows for perceptually significant temporal effects e.g. reverberation and Doppler shifting
  - Coherent nature of sound waves means that interference is more prevalent

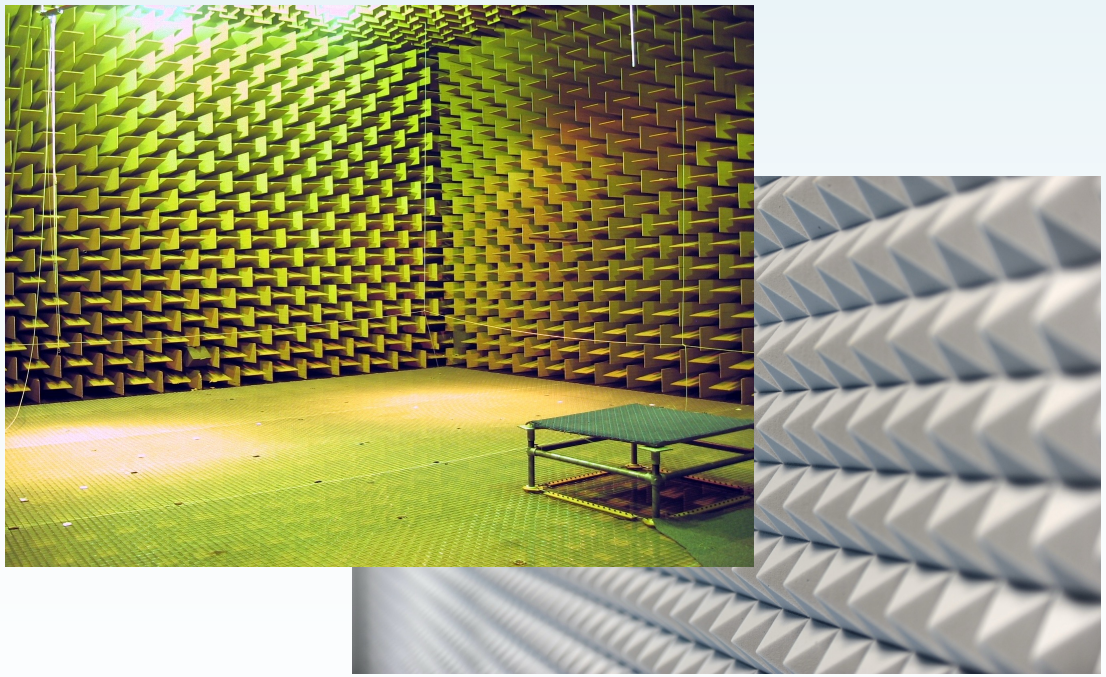
Environmental effects on Soundwaves

- **Absorption by air:** greater for humid or polluted environments
- **Collision:** material and shape/size of object affects amount of reflection, absorption and diffusion
- **Reflection:** solid surfaces will reflect more than soft furnishings which will absorb energy from the sound waves.
- **Diffusion:** If object is small or collision is near edge: collisions within a quarter wavelength from an edge are considered to diffuse (scatter) rather than reflect.

Reverberation

- Sound waves are attenuated (due to spread, absorption, scattering…) until they lose (almost) all their energy
- Reverberation time of a room is the length of time it takes for a sound signal to drop away to an insignificant level (typically 60dB drop).
- Large empty rooms (thus fewer reflections) → longer reverberation time
- Solid surfaces (more reflection, less absorption) → longer reverberation time

Reverberation



Arnaud Dessein, CC BY-SA 3.0,

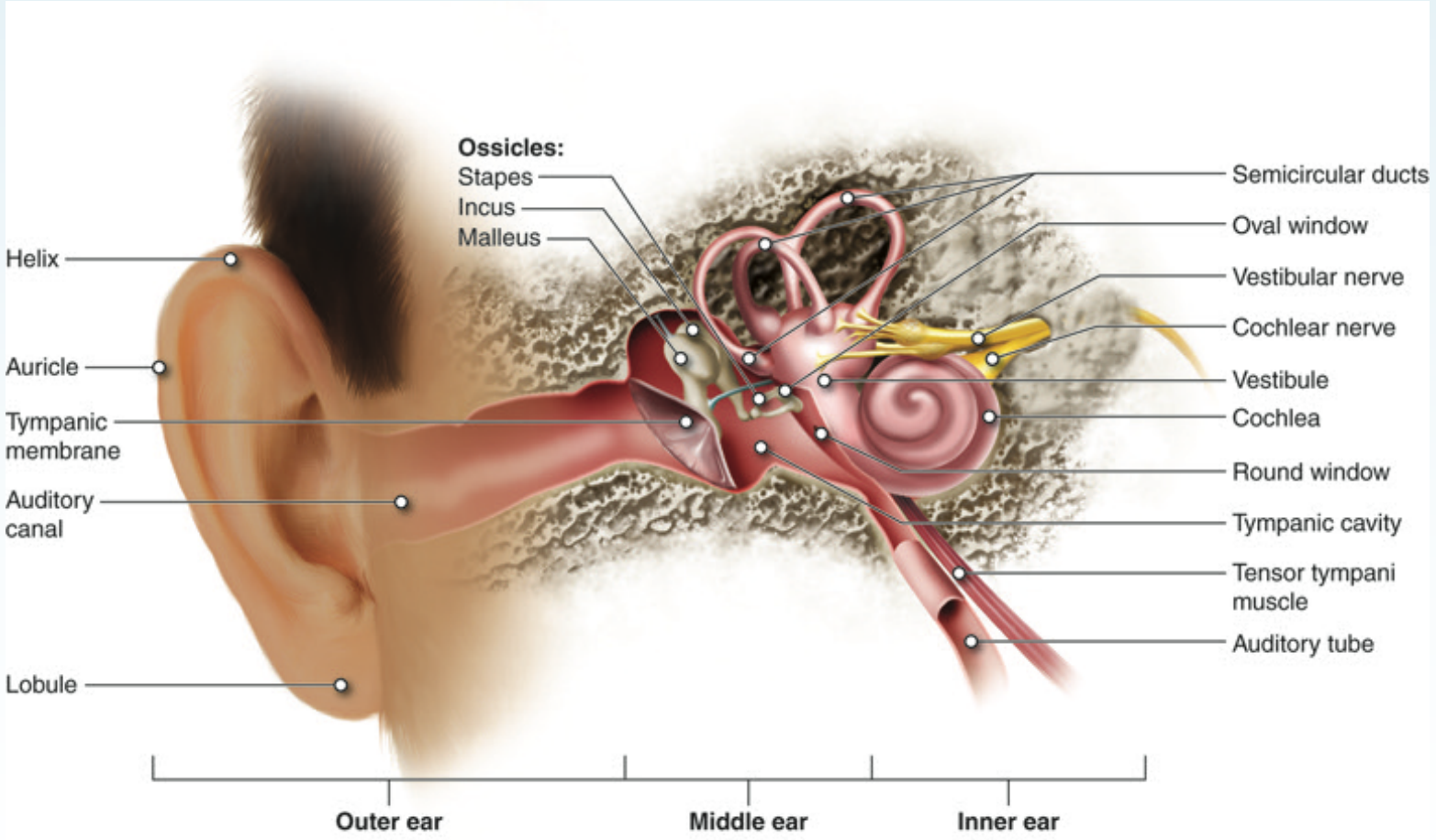
guillaumepaumier.com, CC-BY

Low reverberation

Jorge Lascar, cc-by-2.0

High reverberation

Human Audio System



Original artwork by Cenveo, licensed under Creative Commons Attribution 3.0. Available at: https://courses.lumenlearning.com/nemcc-ap/chapter/special-senses-hearing-audition-and-balance/ (accessed 22/10/20).

Psychoacoustics

Means of associating:

– measurable audio stimuli (e.g. frequency, power)

with:

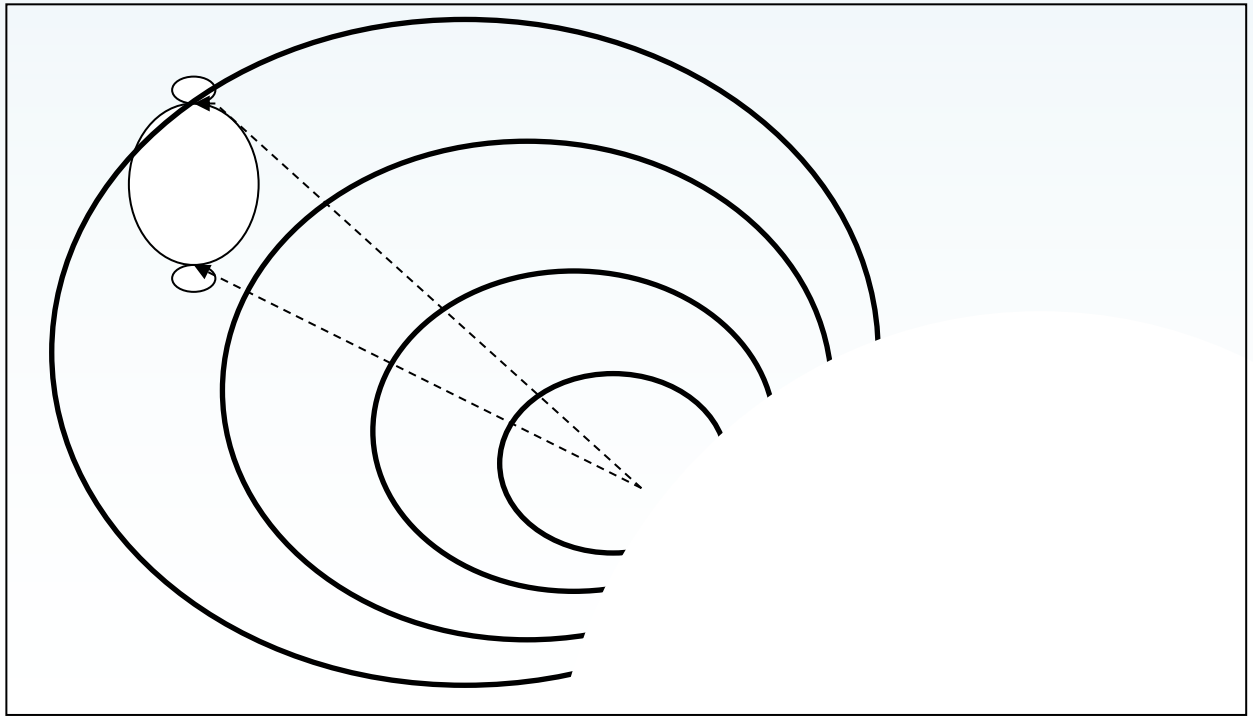
– subjective sensations (pitch, loudness)

Ranges of sensitivity:

- Frequency range approx. 20Hz to 20kHz (~10 octaves)
- Amplitude range hard to determine, but order of magnitude for ratio of loudest audible sound (at pain threshold) to quietest audible sound at 4kHz is one million.

Sound source location

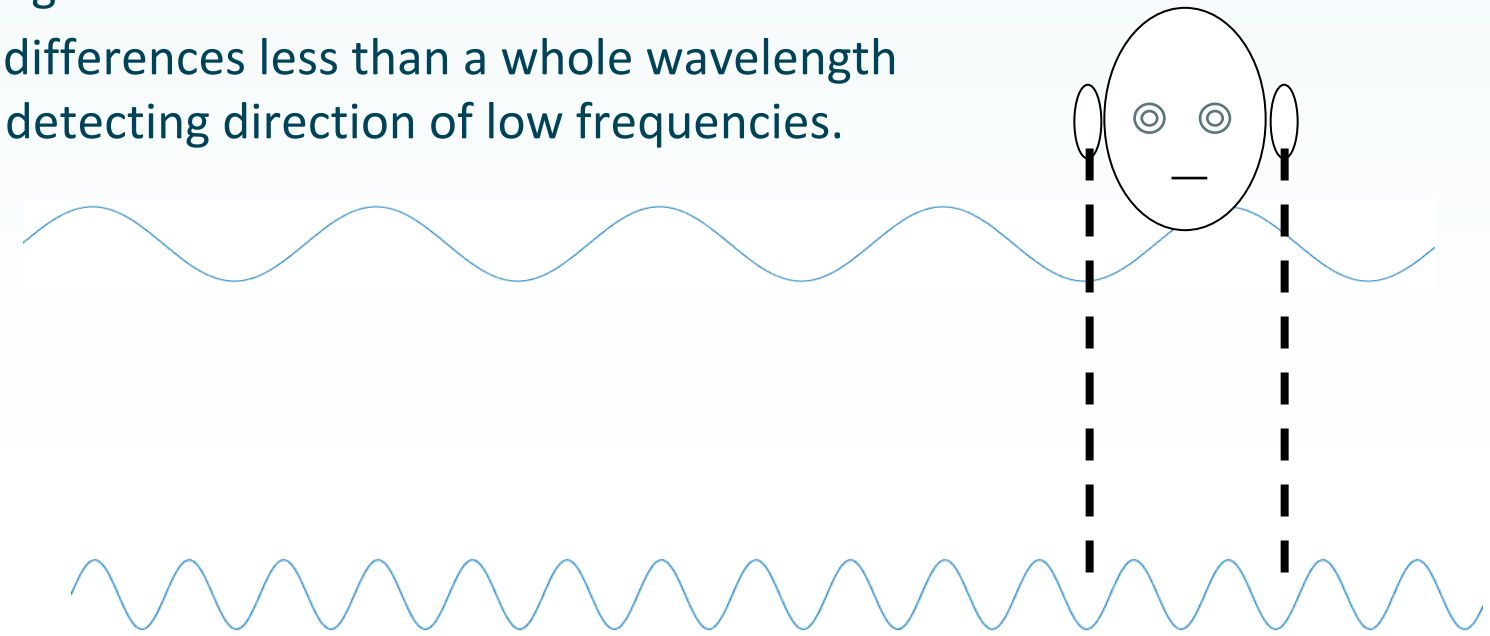
Direction and distance estimates are made on account of differences in the sounds that reach our left and right ears.



Source direction estimation

(1) Interaural time difference:

- Sound arrives at a fractionally different time (less than 1ms) at the left and right ears.
- Human audio system is sensitive to the phase difference between the left and right channels
- Only works for phase differences less than a whole wavelength thus works better for detecting direction of low frequencies.



Source direction estimation (2)

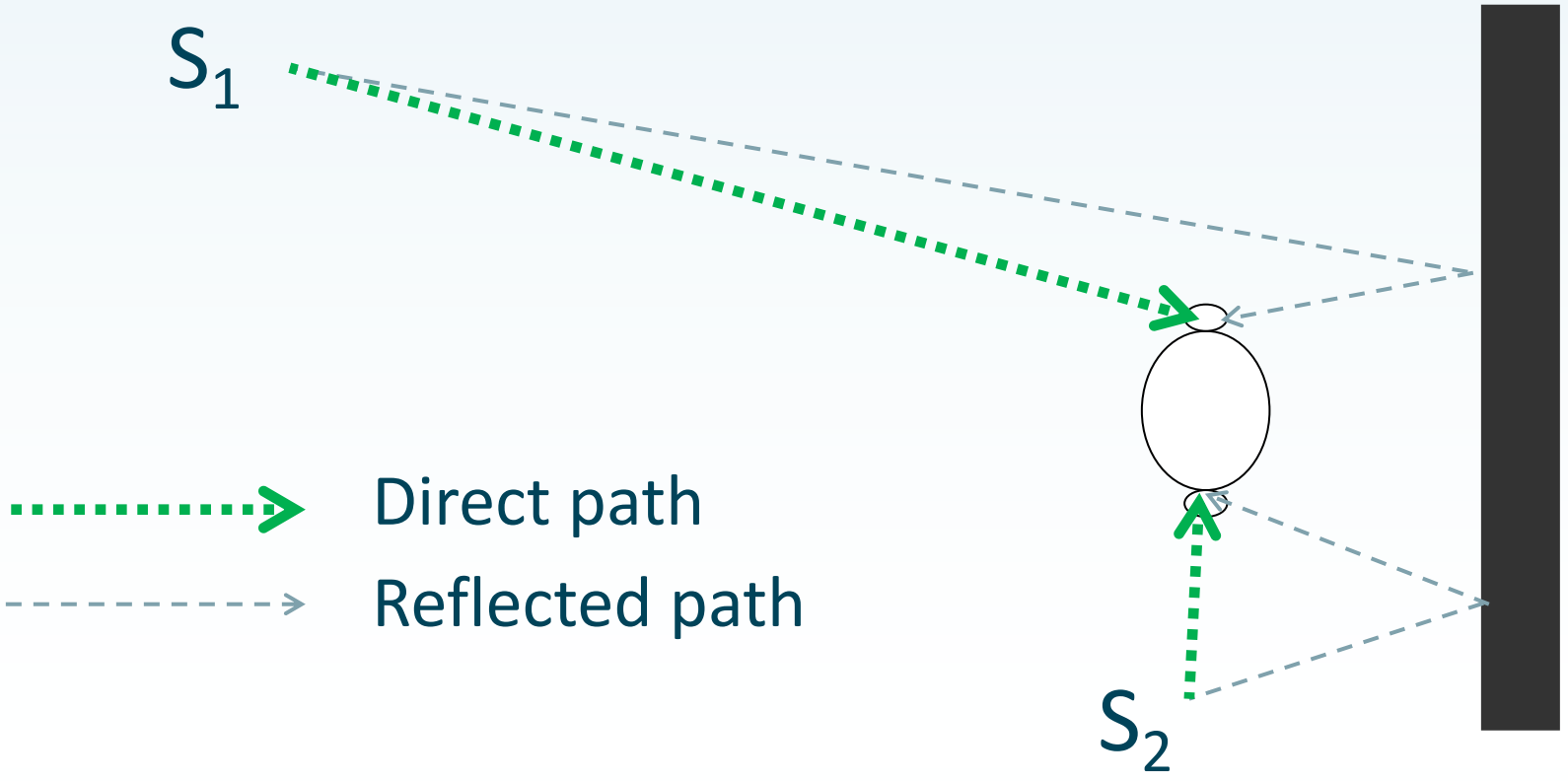
(2) Interaural intensity difference:

- Sound is scattered when it strikes the head and upper torso – an effect known as “shading”.
- Intensity of sound at ear further from the source is consequently reduced.
- Shading is more pronounced for high frequencies since these are scattered more.

Source distance estimation

- How to differentiate a loud sound far away and a quiet sound close by.
- We can detect differential attenuation across the frequency range as sound sources move further away from us.
- Difference in the amount of reflected sound: more reflected sound implies that sound source is further away.

Distance estimation from reflected:direct sound ratio



Head-related transfer function (HRTF)

- Model of filtering effect of our bodies
- Useful for headphone setups
- Composed of:
  - Filtering by outer ear flap (pinna) affects the propagation of different (especially high) frequencies.
  - Precise nature is determined by the ear shape.
  - Upper torso reflects frequencies (especially mid-range) to produce very short time-delayed echoes.
  - Length of time delay varies with the elevation of the sound source.

HRTF

- Measured by placing small mics at entrance to ear canal and taking many measurements:
  - Varying frequency
  - From all around the listener
- Time-consuming to measure so generic HRTFs often used

HRTF properties:

- Unique to individual
- Monaural cue (though we have 2 distinct HRTFs)
- Modifies both frequency spectrum **and** timing of incoming signal .
- Varies with direction of incoming signal
- Affected by changes in clothing, hairstyle etc.

Part 2 Overview

Synthesis of spatial audio

- Methods for synthesizing spatial audio
- Hardware & Data requirements
- Headphones vs speaker array
- Environmental acoustic modelling

Binaural Simulation

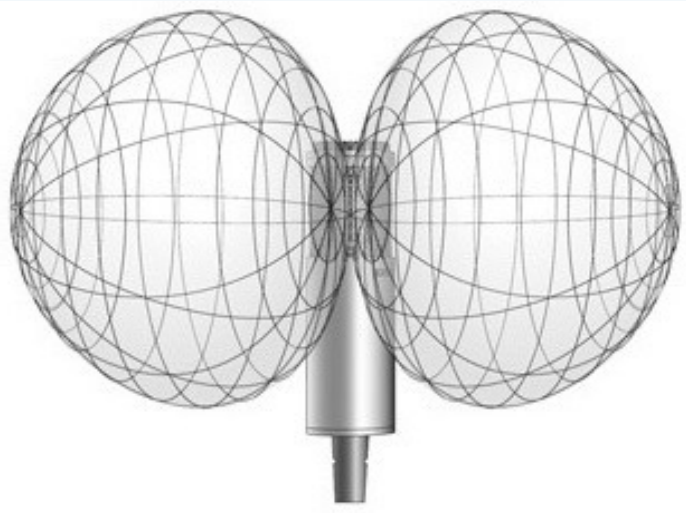
- 2 output channels
- Originally conceived as means of recording acoustics for evaluation of auditoria
- Takes account of listener's physical characteristics via HRTF
- HRTF highly effective though laborious to compute
- Generic HRTFs often used
- Requires headphones (or crosstalk cancellation) for listener

Soundfield Reconstruction without Headphones

- Requires many output channels:
  - >= 4 for horizontal simulation
  - >= 8 for 3D
- Must account for location of listener relative to the speaker array
- 2 predominant methodologies:
  - Ambisonics
  - Wavefield Synthesis

Ambisonics

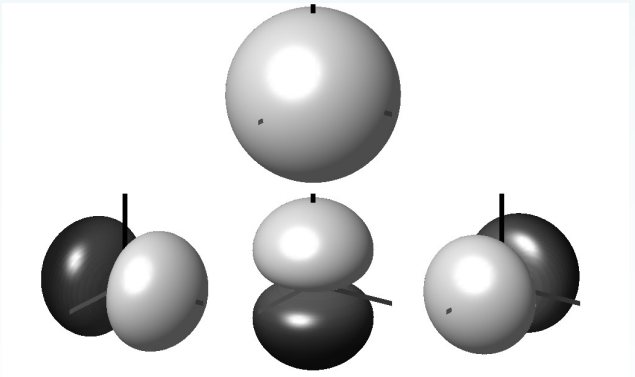
- Relies on directional recording of sound
- Figure-of-8 microphone is fundamental component for recording
- First order ambisonics requires 3 orthogonally arranged mics

Back  Front

- Sensitive to sounds from opposites side (e.g. front/back)
- Front and back have opposite polarity
- Sound from front hits front then back of mic
- Sounds from sides are cancelled out
- => No pickup from sides

Ambisonics

- First-order ambisonics encodes soundfield in 4 channels (3 orthogonal channels + omnidirectional)
- Combined output referred to as a B-format signal:
  - Audio pressure W
  - Front-back X
  - Left-right Y
  - Up-down Z

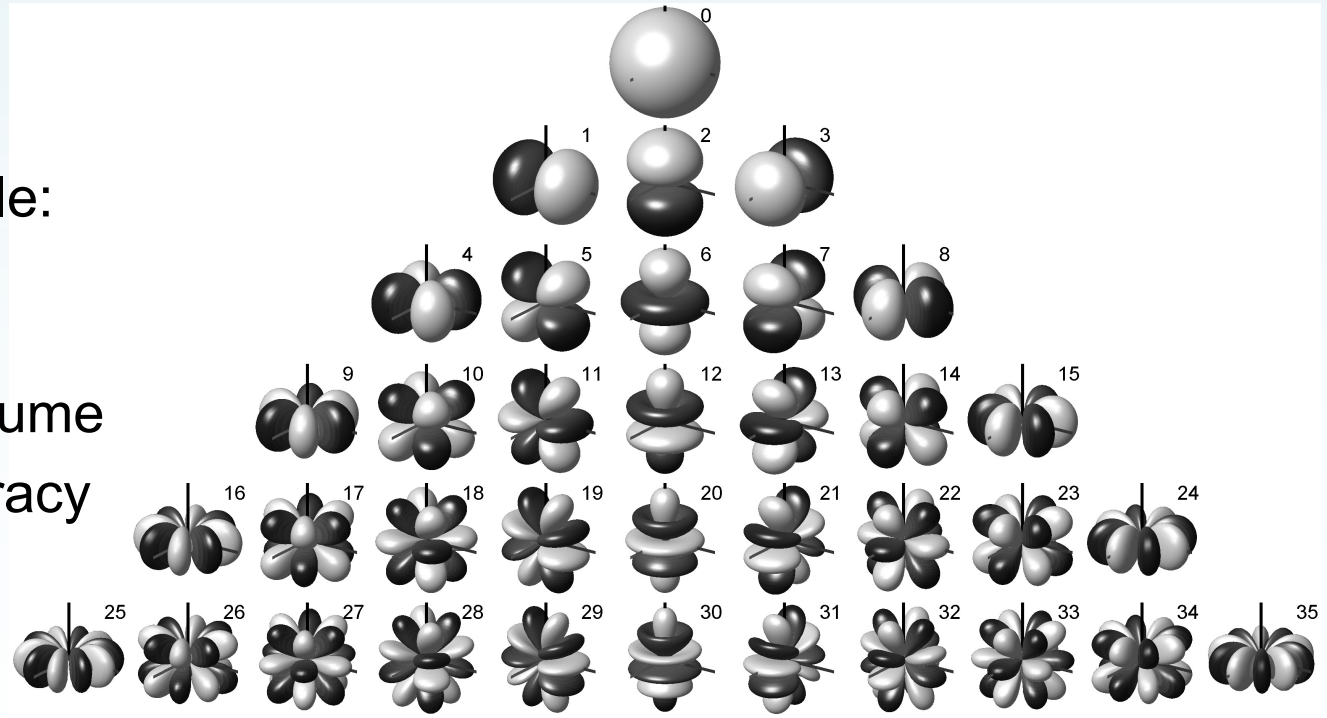


Franz Zotter, CC BY-SA 3.0

- Signals can also be synthesised (e.g. conversion of mono or stereo recordings)

Ambisonics

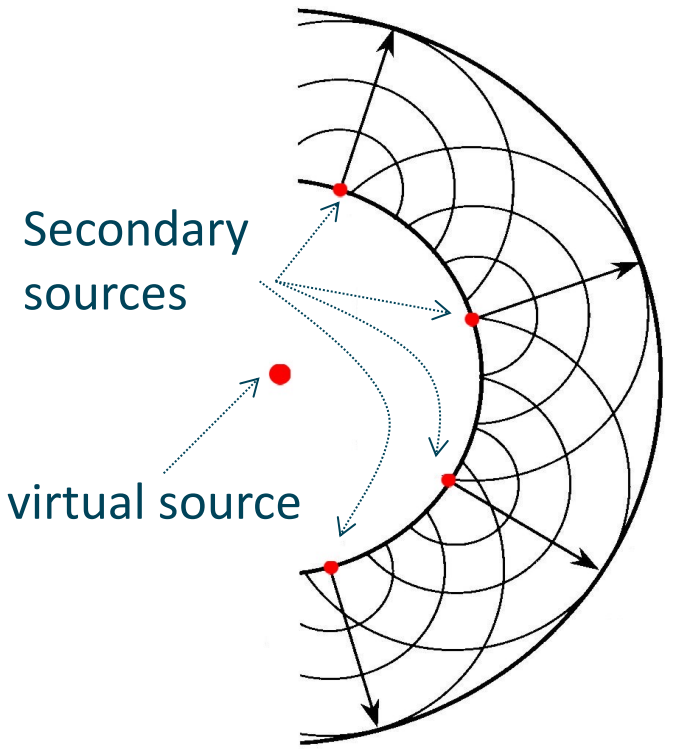
- Can be easily operated upon by spatial transformations (e.g. rotations)
- => Encoding and decoding are separable: signal can be decoded for any arbitrary speaker array (including headphones!)
- Very accurate but for limited listener volume
- Higher-order ambisonics increase accuracy and listener volume



Franz Zotter, CC BY-SA 3.0

Wavefield Synthesis

- Based on Huygens' Principle
- Soundfield of environment is approximated by array of loudspeakers acting as secondary sources
- Many loudspeakers in dense linear arrangement produces large stable soundfield populated by numerous virtual sources
- Complex algorithm
- Approximately uniform error over large area
- Can be experienced by many listeners



Processing Pipeline for Virtual Audio

Headphone (binaural) presentation

Speaker array presentation

Headphones vs Speaker Array

- Headphones avoid the need to account for the location of the listener relative to the speaker array
- Speaker array avoids the need for computation of HRTF
- Headphones provide acoustic isolation (useful in some circumstances)
- Headphones can be an encumbrance
- Speaker array better for loud, low frequency sound

Data requirements

- Sampling rate of audio source:

  To represent a signal of frequency ***f*** , it must be sampled at a rate of at least ***2f*** (Nyquist theorem);

- Dynamic range of audio source:

  Dynamic range is nonlinear and can be adequately represented by as few as 256 gradations (which conveniently fits into 8 bits of data) although high quality digital audio will be represented by 16, 20 or 24 bits

Environmental acoustic modeling

- Analogous in some ways to building a graphical model.
- Geometry of the environment described as 3D meshes
- Characteristics of surface materials:
  - **absorption** and **diffusion** co-efficients
  - These may vary for different frequencies.

Environmental acoustic modeling (2)

- Also important differences to computer graphics
- Acoustic models require much less geometric detail
- Propagation is modeled from the sound source until we have a sufficient representation at the listener (ear)
- Reverberation
- Can compute “impulse response”

Simulation of environmental effects

A model of the acoustic properties of an environment can be used as the basis for computing a simulation of the propagation of sound through that environment. There are various ways of achieving this:

- **Finite element methods**
- **Image source methods**
- **Ray tracing**
- **Beam tracing**

For a review of techniques see:

Funkhouser et al. (2003). *Survey of methods for modeling sound propagation in interactive virtual environment systems*.

Recap

- Spatialized audio is multifaceted concept with different engineered approaches
- Physics of sound waves
- Human audio system
- Modelling of level and phase differences of sound propagation to listener's ears
  - Panning
  - Ambisonics
  - Wavefield syntheis
- Modelling of the effect of the environment on acoustic propagation
- Trade-off between modelling for speaker array or headphones