

Spatial Audio

Matteo Luperto
Eleonora Chitti

matteo.luperto@unimi.it
eleonora.chitti@unimi.it

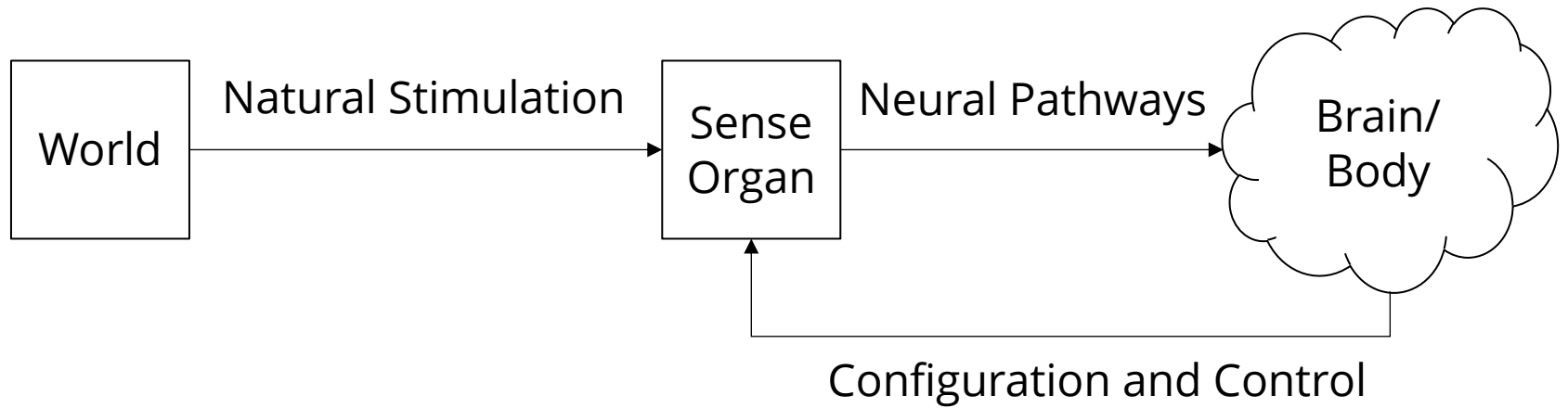


UNIVERSITÀ DEGLI STUDI
DI MILANO

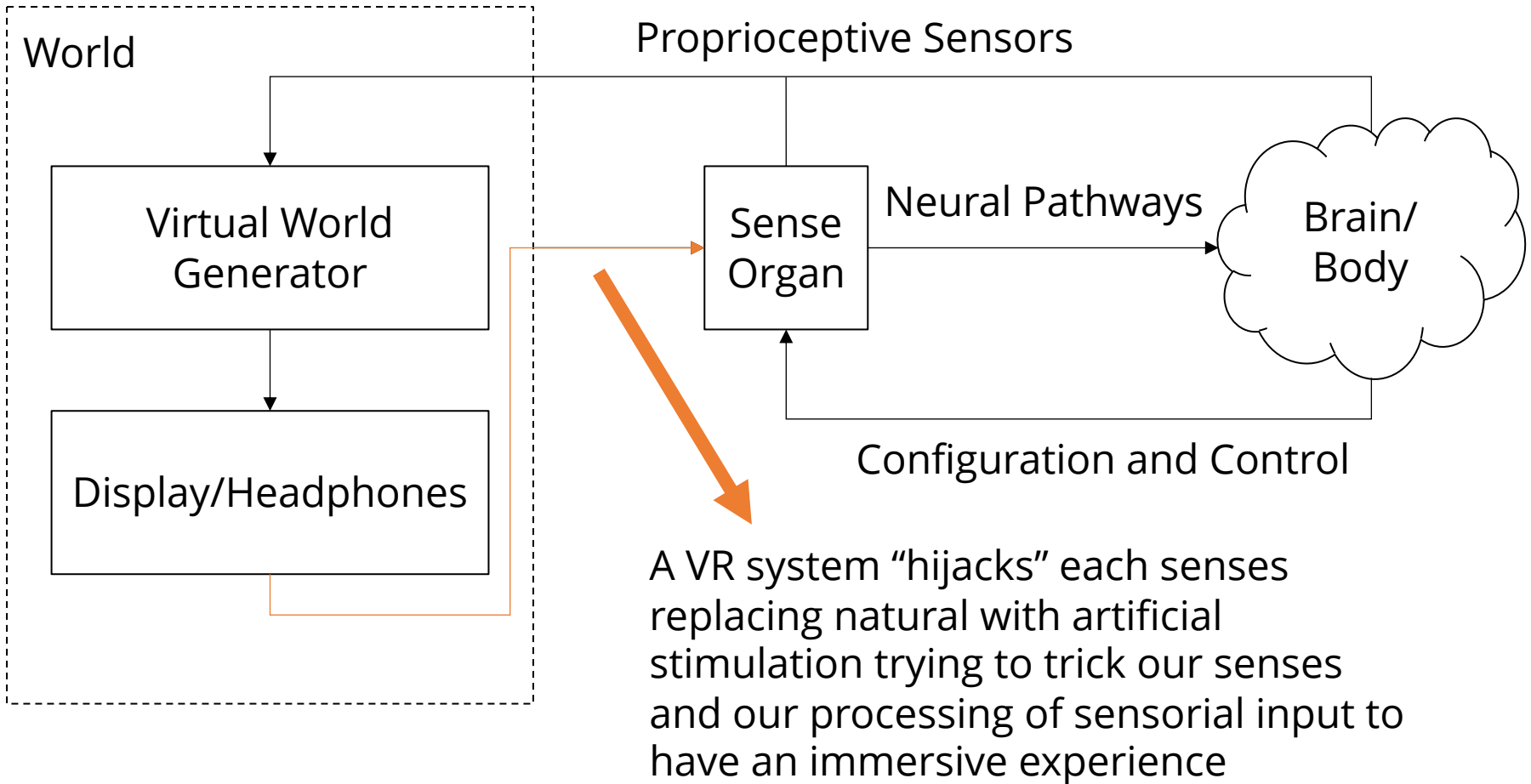
30/05/24

Realtà Virtuale 2023/2024

Perception



Senses and VR



Perception + VR

Up to now you have primarily seen how this is done with vision/sight as:

- Vision is the primarily sense that we need to “fool”
- The development in the last 15 years of portable VR kits and tracking algorithms/devices has de-facto enabled VR as a commodity

Perception + VR

Up to now you have primarily seen how this is done with vision/sight as:

- Vision is the primarily sense that we need to “fool”
- The development in the last 15 years of portable VR kits and tracking algorithms/devices has de-facto enabled VR as a commodity

However, for other senses that are “less complex” than vision, the technology suited for a VR experience was already there we already experimented how to “trick” those senses into perceiving something different than what is happening

Audio directionality



In music you try to have an immersive experience by recording/mastering tracks so to give the feeling to the listener of being “live” with musicians:

- Guitar on the left,
- Bass on the right,
- Drums behind,
- ...

What you do is give a “directionality” to each instrument/sound

Audio directionality



By focusing on a simpler setting (only 1 sense, hearing, fixed subject position) music tried for decades to recreate a “virtual reality” experience

Audio for VR

Spatialized audio and VR-related techniques are perhaps the oldest form of “virtual reality” experience, and the only commonly used and settled technology

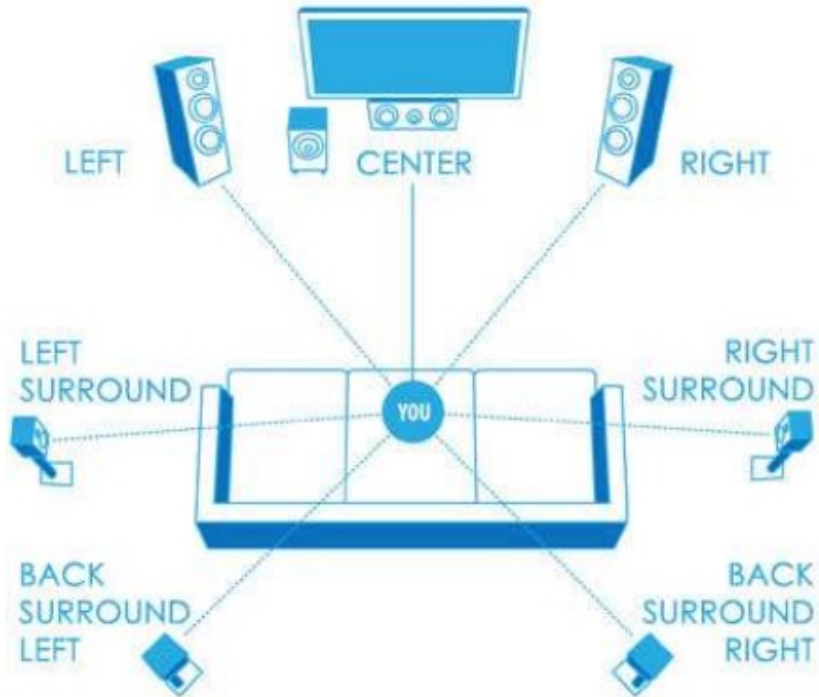
- Stereo audio started in 1930' – mainstream since 1960'

... but its use is mostly related to:

- Music
- Cinema

Also, audio-VR interfaces can be pretty cheap
(and probably you are using one right now)

World- vs user-fixed



audio 7.1



headphones

Headphones \approx Visors



Audio → Spatialized Audio



The main difference between (spatialized) audio in music/movies and VR is that the former assumes a fixed/passive position of the listener.

With VR we introduce a feedback involving the user's movements to dynamically adapt the sound.

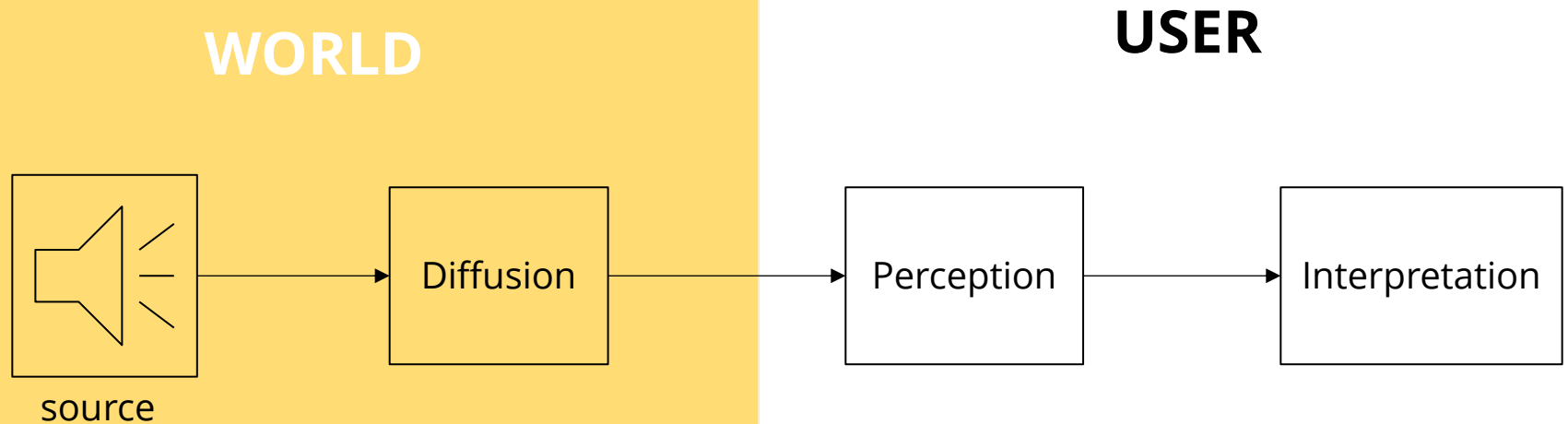
For that we rely on headphones.

However, we lose the immersivity of surround system due to "feeling" certain frequencies.

Today's outline

- Audio 101 – some basics about how sound works
- Perception
- Spatial audio modeling
- Available SDKs
- Design Tips

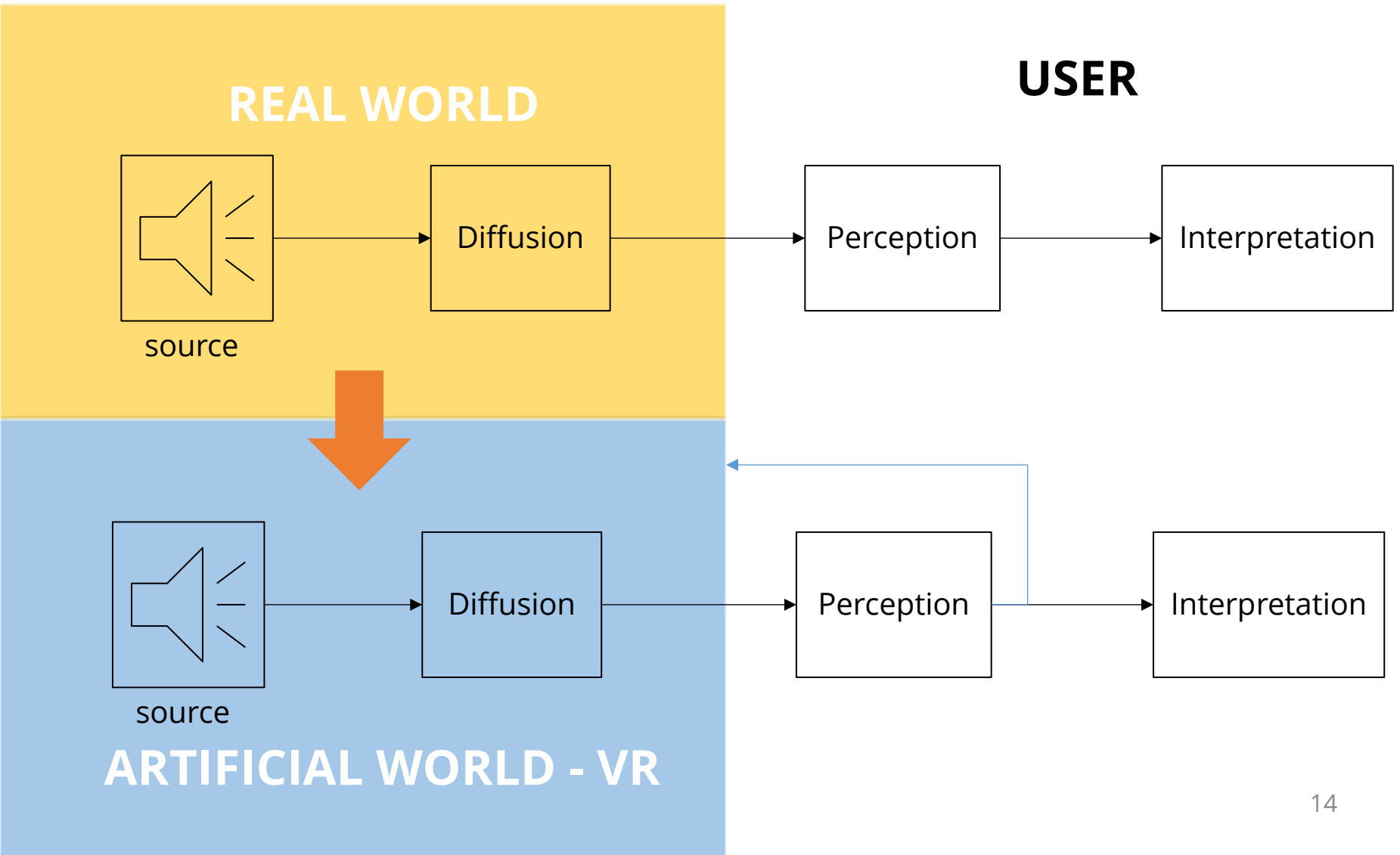
Audio modelling for VR



We need to understand how sound works in our world so to replicate its dynamics in VR.

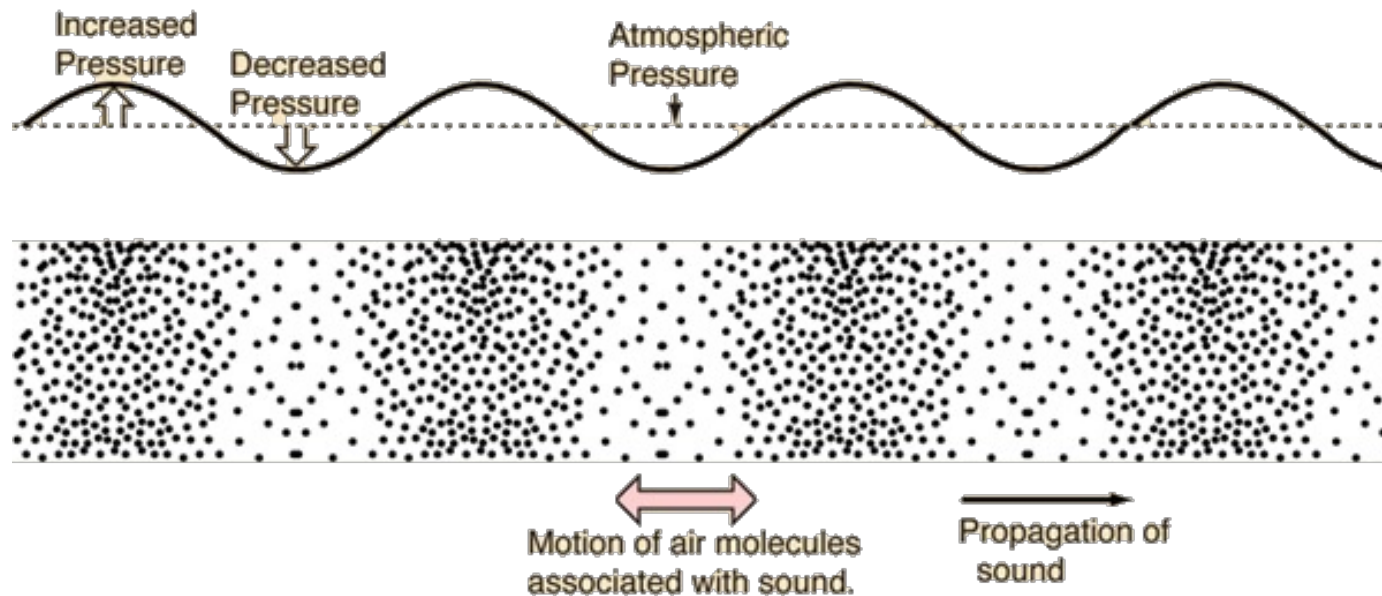
We need to understand how we perceive and process audio so to “fool” our senses

Audio modelling for VR



Audio waves

Sound is a longitudinal wave of compression and rarefaction of air molecules



Both sound and light are both propagated by waves, there are many similarities

Speed of sound

- Speed of sound (in air 20°) is 343 m/s
 - it is slow, compared with the speed of light
- We can detect delays / differences in speed and distortion
- Example: Doppler effect

$$f_r = \frac{s + v_r}{s + v_s} f_s$$

r : sender; s : receiver; v : velocity; s : media speed propagation; f : frequency;
If the source is moving at $v_s = -43.2 \text{ m/s}$ the receiver perceives it at 300m/s
Result: a frequency of 1000hz (source) is perceived as 877hz (receiver)

We can detect this cues and use them
(more on this later)

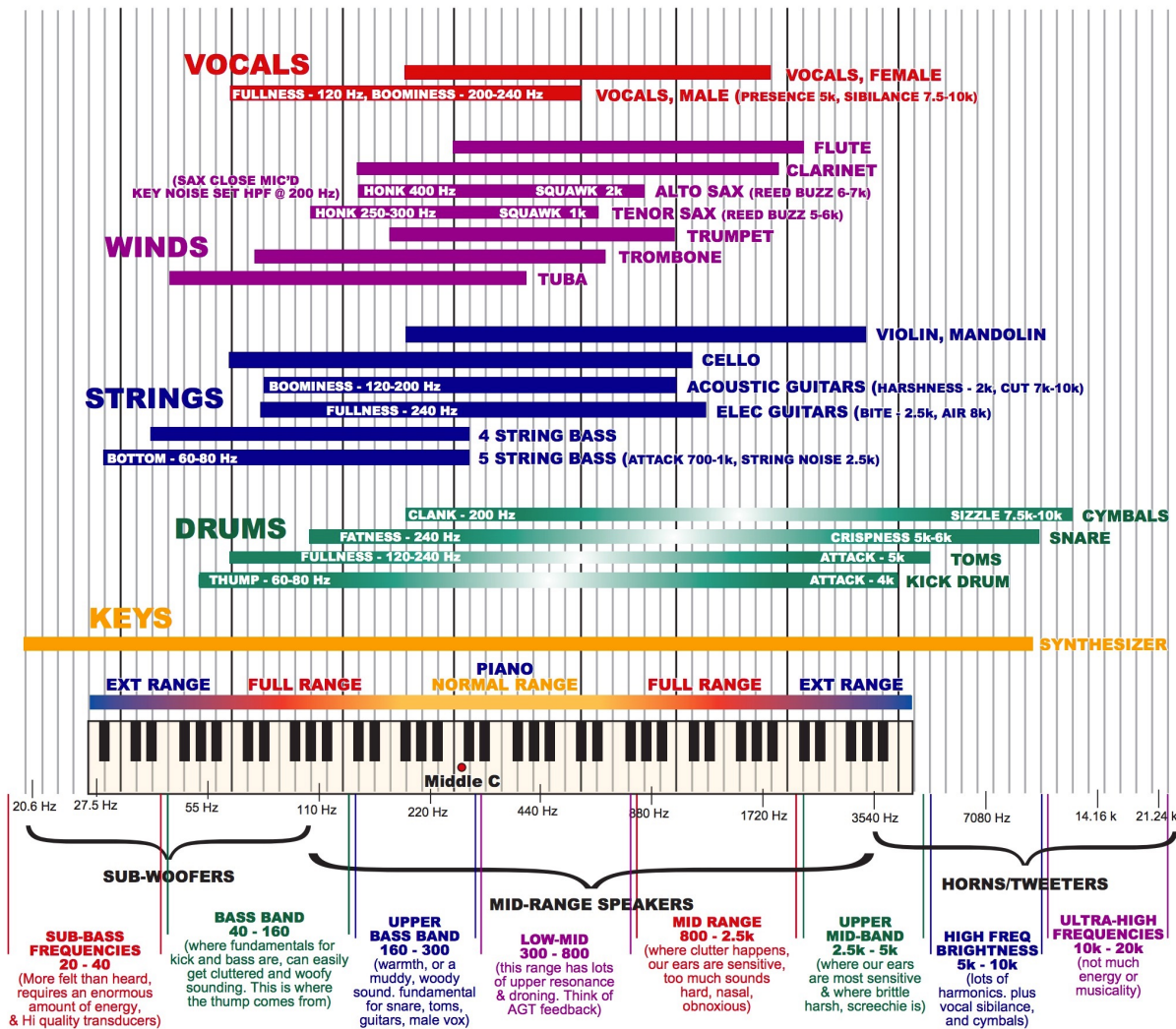
Audio waves

20 - 25k Hz

very broad spectrum

HF range decrease with age

$f > 14k$ Hz cannot be heard by most adult, but do not contain many info



pic from Reddit

Low and High Frequency

There is a big difference between Low Frequencies (**LF**) and High Frequencies (**HF**) and this affects both how they propagate and how we perceive them.

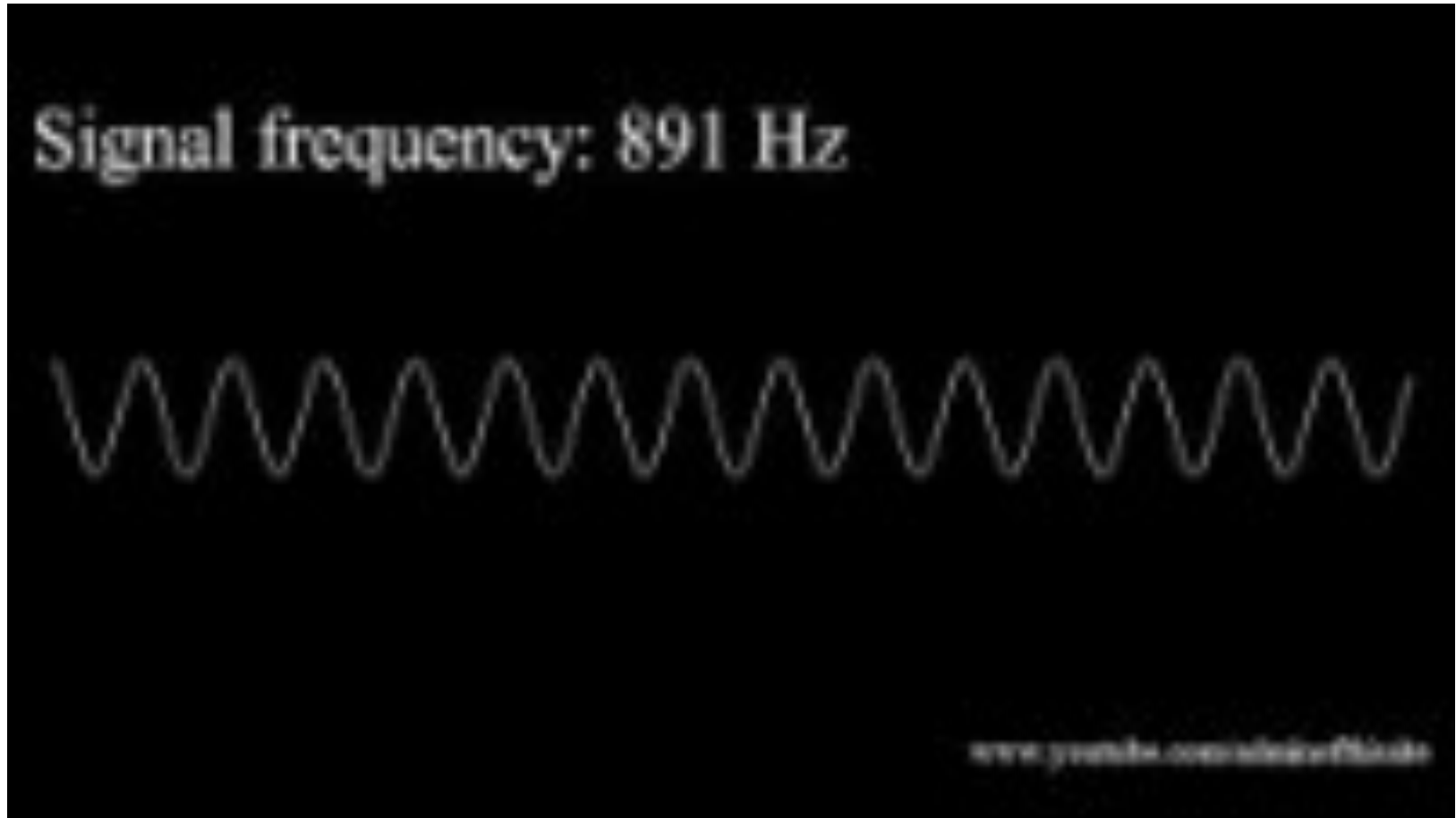
$$\text{wavelength} \rightarrow \lambda = \frac{v}{f}$$

v ← speed (340 m/s)
 f ← frequency

$$f = 20 \text{ Hz} \rightarrow \lambda = \underline{30 \text{ m}}$$

$$f = 25000 \text{ Hz} \rightarrow \lambda = 13.6 \text{ mm}$$

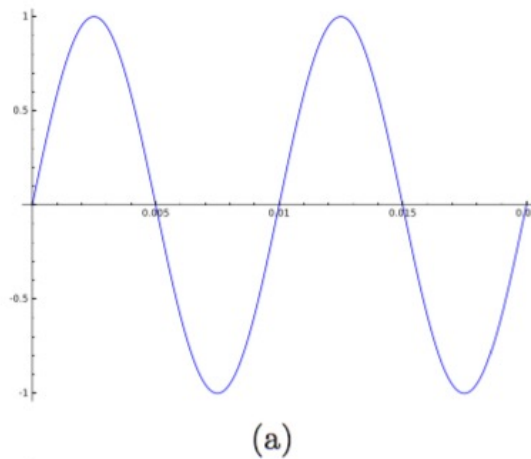
Audio spectrum - video



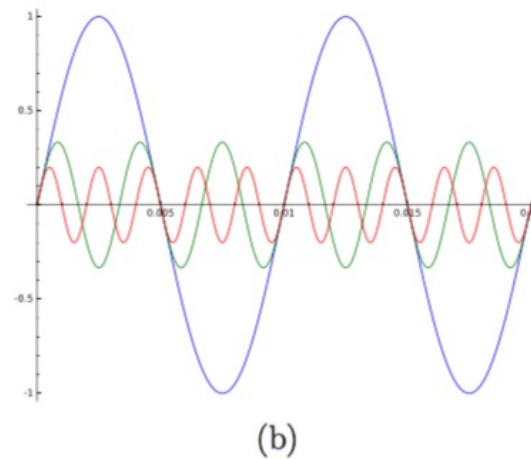
Audio Waves are Waves...



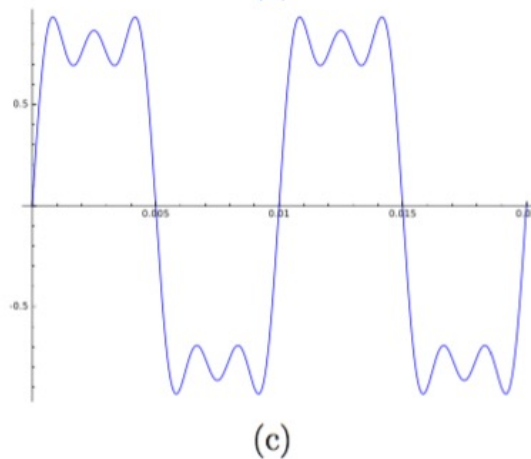
Spectral decomposition and Fourier analysis are important.



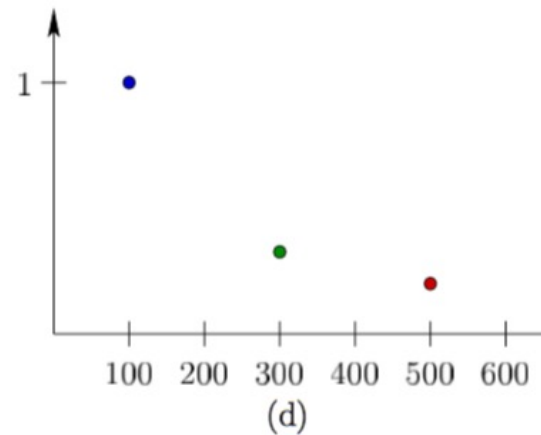
(a)



(b)



(c)



(d)

Sound sources

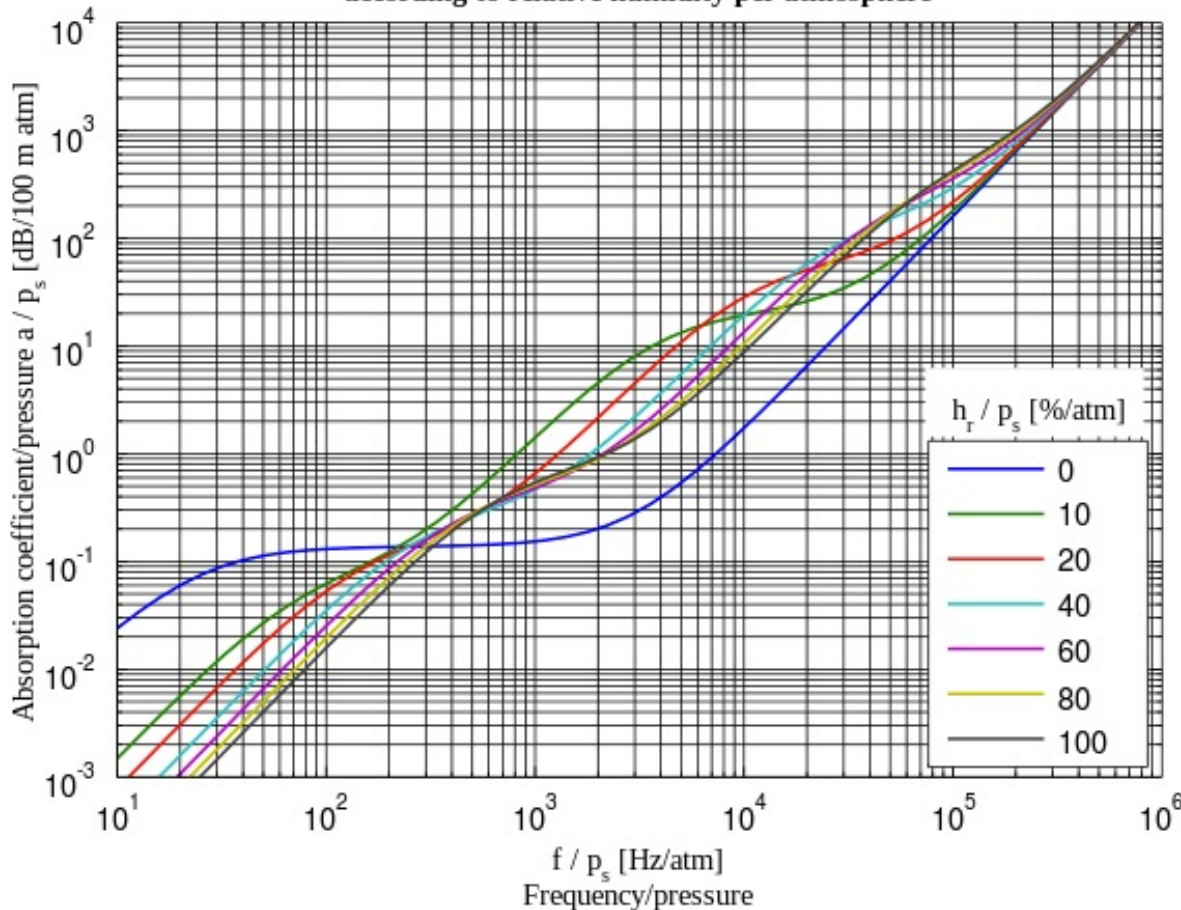
In *our* world audio sources are complex and diffused

In *virtual* worlds audio sources are single points in the space and this can cause modelling issues



Sound Absorption

Sound absorption coefficient per atmosphere for air at 20°C
according to relative humidity per atmosphere



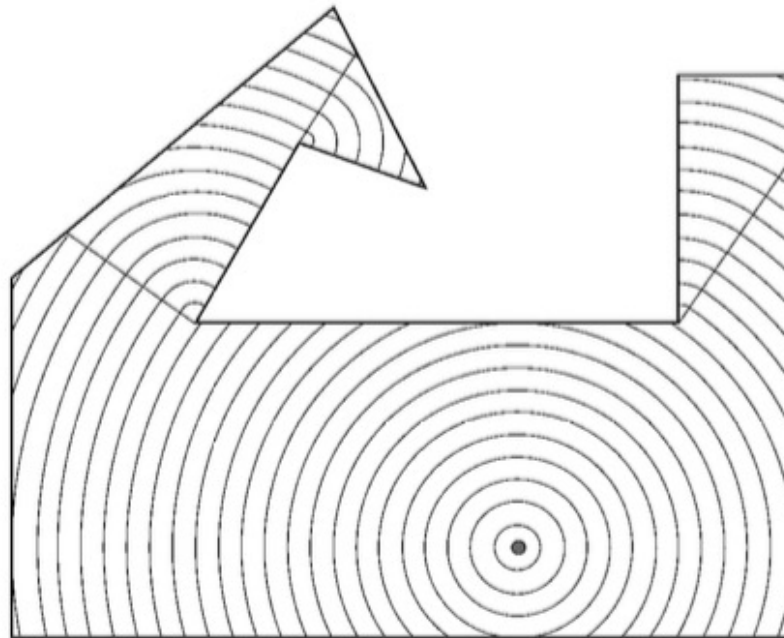
Environmental sound absorption changes a lot from LF to HF.

Consequently, sound changes with distance from the source as different spectral components are subject to different absorption level.

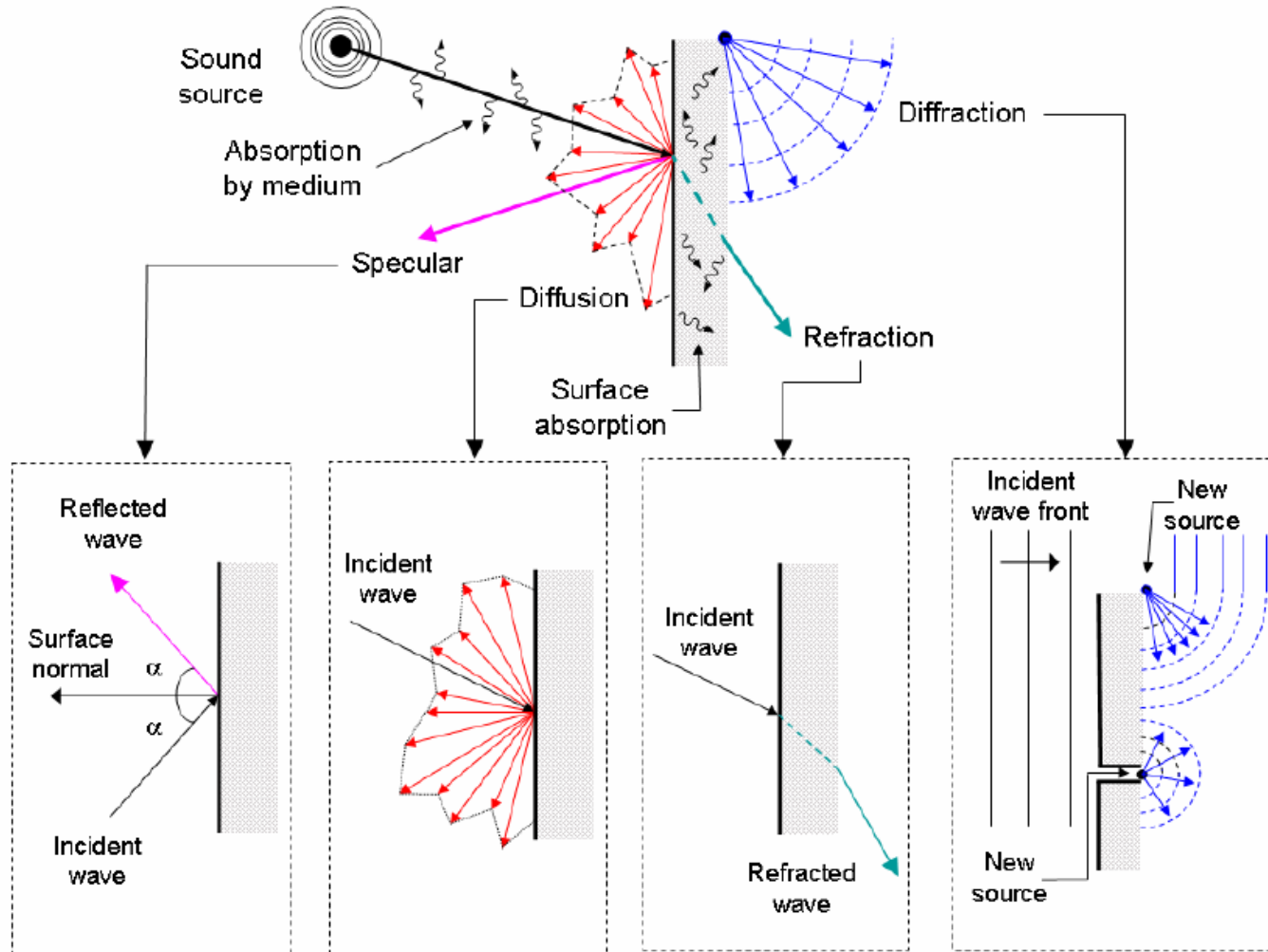
Reflection – Transmission – Diffraction

If an audio wave hits a wall:

- (most) of its energy will bounce – **reflection**
- (some) penetrate– and propagate faster than in air – **refraction**
- (some) trespass the wall and goes on after- **transmission**
- waves bend across corners / obstacles - **diffraction**
- **HF** and **LF** behave very differently



Reflection – Transmission – Diffraction

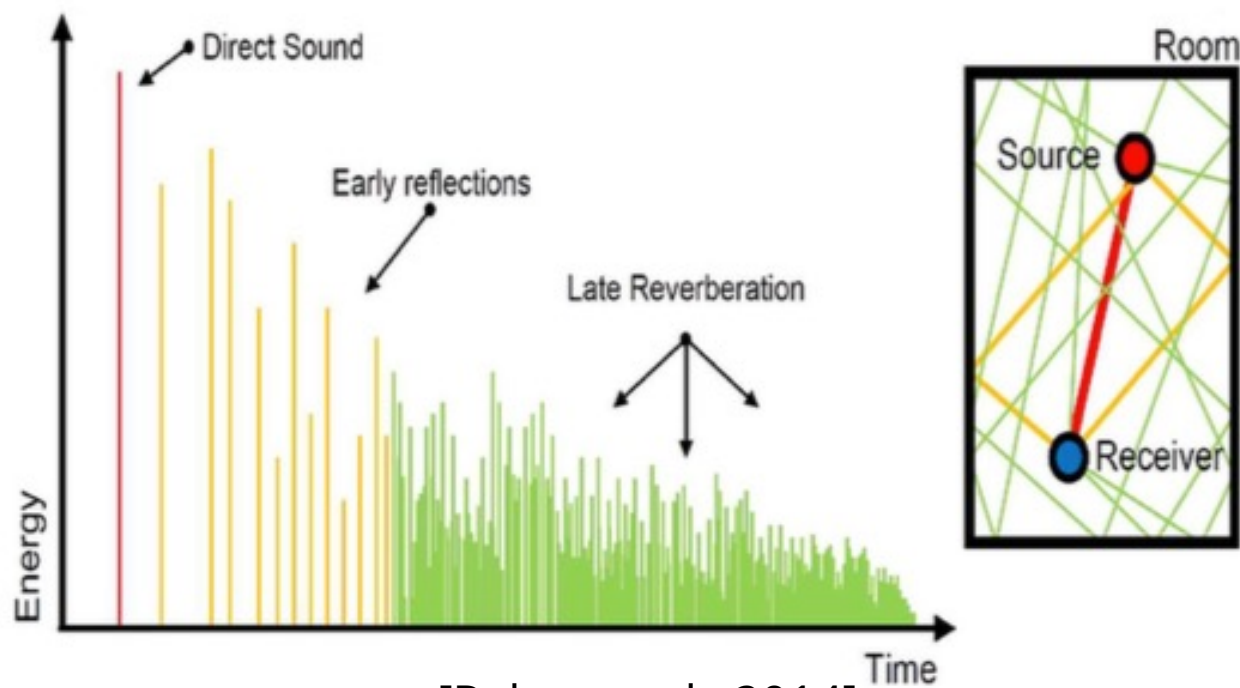


[Kapralos et al, 2008]

Audio Propagation

As a result a single sound is replicated into many (many) slightly modified copies, all of which are then perceived by the user.

Audio is a simpler information stream than video, but audio propagation is more complex than light propagation



[Pelzer et al., 2014]

Our *sensors*: eyes and ears

Eyes capture at (relatively) low frequency a complex data

- e.g. 1920*1080 (2M) pixel ~ 1k Hz (60 Hz) (fullHD)

Our *sensors*: eyes and ears

Eyes capture at (relatively) low frequency a complex data

- e.g. 1920*1080 (2M) pixel ~ 1k Hz (60 Hz) (fullHD)

Ears capture a 1-dimensional signal (each) at high frequency

- e.g. left + right @ 44100 Hz

Each ear can be seen as a 1-pixel high-frequency high-resolution camera
(but we still have only two-pixel resolution)

Ear physiology is complex

(and difficult to model – from a VR perspective)

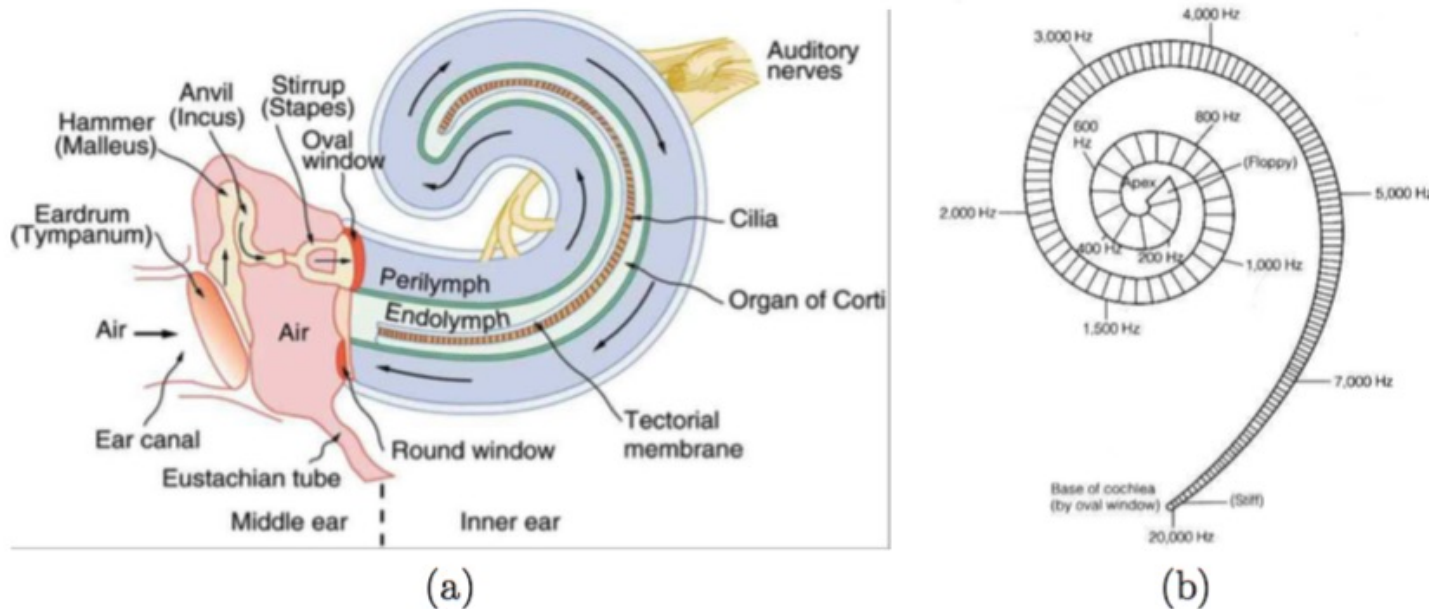


Figure 11.5: The operation of the cochlea: (a) The perilymph transmits waves that are forced by the oval window through a tube that extends the length of the cochlea and back again, to the round window. (b) Because of varying thickness and stiffness, the central spine (basilar membrane) is sensitive to particular frequencies of vibration; this causes the mechanoreceptors, and ultimately auditory perception, to be frequency sensitive.

Ear physiology is complex

(and difficult to model – from a VR perspective)

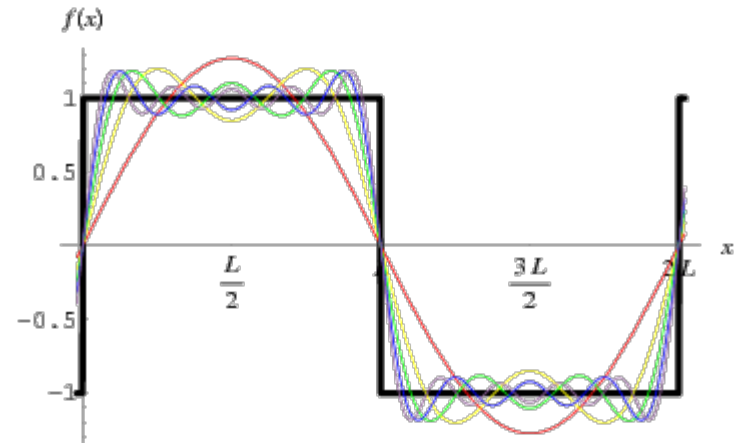
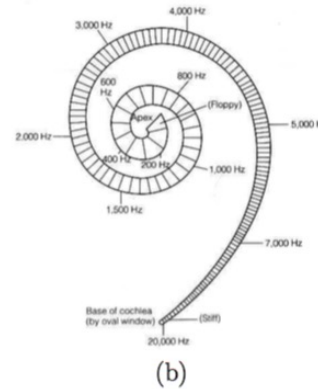
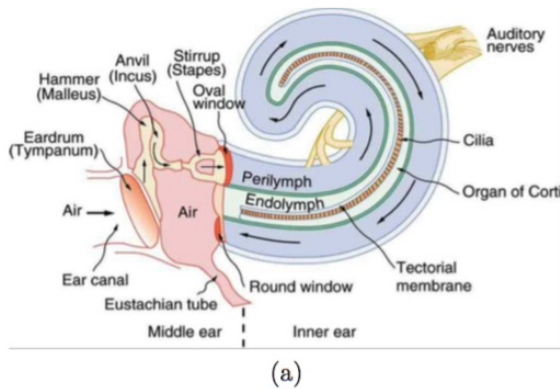


Figure 11.5: The operation of the cochlea: (a) The perilymph transmits waves that are forced by the oval window through a tube that extends the length of the cochlea and back again, to the round window. (b) Because of varying thickness and stiffness, the central spine (basilar membrane) is sensitive to particular frequencies of vibration; this causes the mechanoreceptors, and ultimately auditory perception, to be frequency sensitive.

However, what “roughly” ears do is to perform a spectral decomposition and frequency-based analysis of sound waves

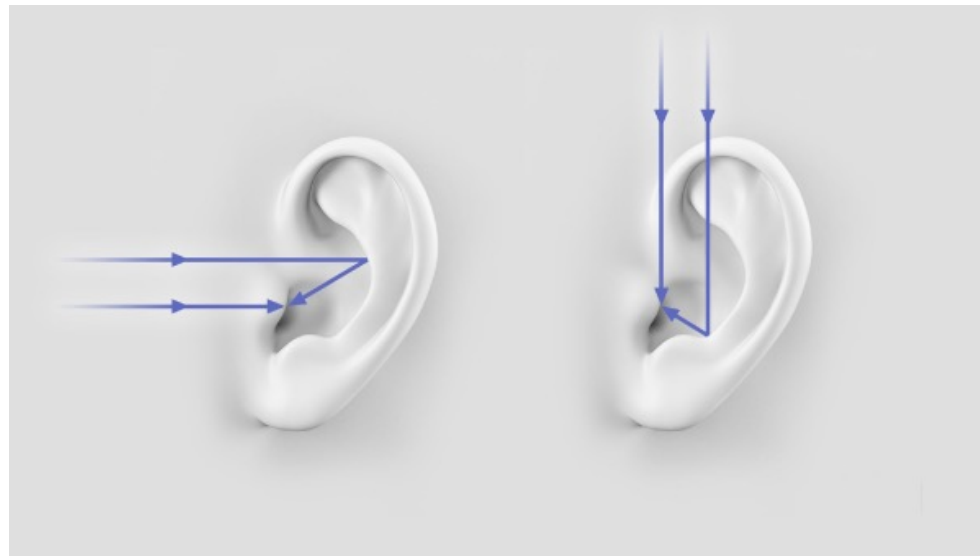
Auditory perception and interpretation

Audio perception involves a lot of “brain processing”, due to our the ear ability and to the different types and phenomena which affects sound waves

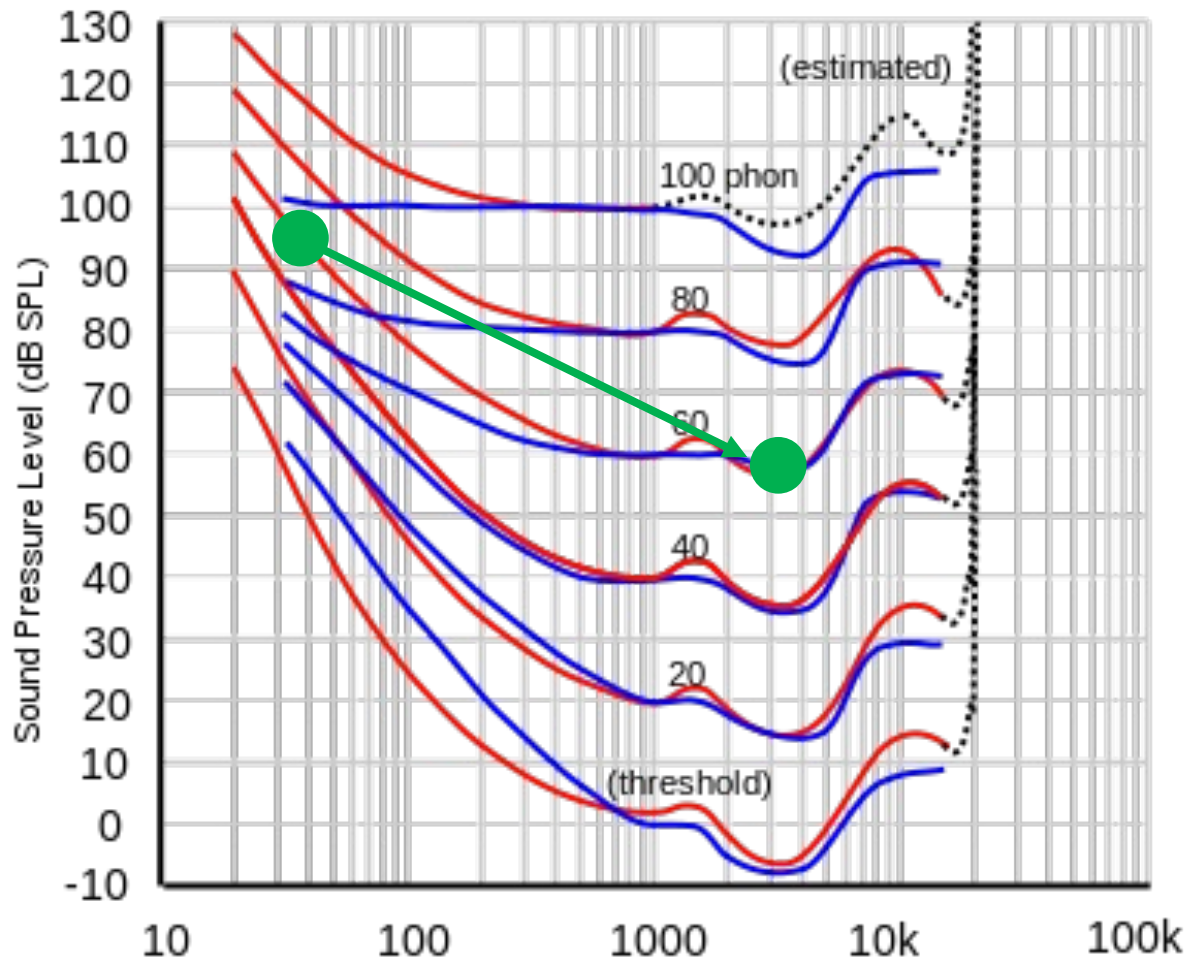
- different wavelengths results into different diffraction and reflections;
- also, HF and LF waves propagates differently and have different energies
- we have 2 ears listening to two versions of the same sound at once
- ears and perception copes with adaptation, missing data and assumption.

Auditory perception and interpretation

Auditory perception also involves some “mechanical” components; e.g., our pinna (outer ear) distort the sound in a controlled way that is used by the brain for processing



Equal loudness contour curves



As LF and HF behaves differently, we also perceive them differently.

A LF require a (far) higher volume to be perceived as loud as an HF one.

Equal-loudness contours (red) (from ISO 226:2003 revision)
Fletcher–Munson curves shown (blue) for comparison

Auditory perception + freqs

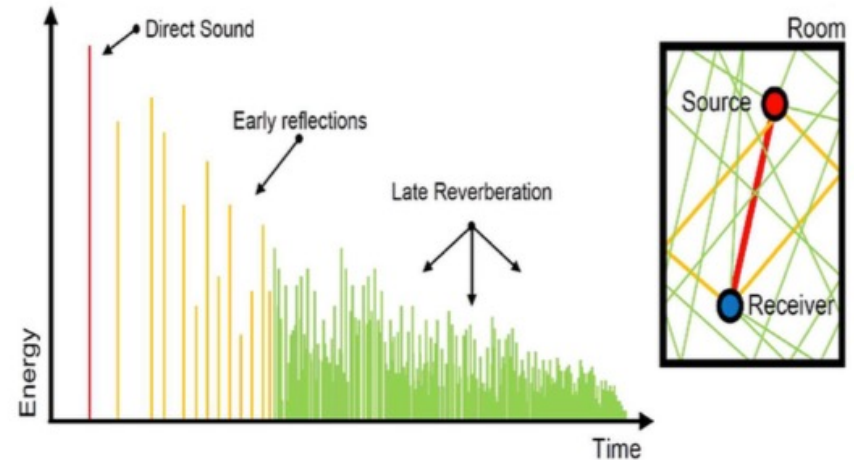
low frequencies

- high wavelength
- require high energy (db) to be perceived
- can go through objects / walls
- air absorption is low
- can be perceived far from their source

high frequencies

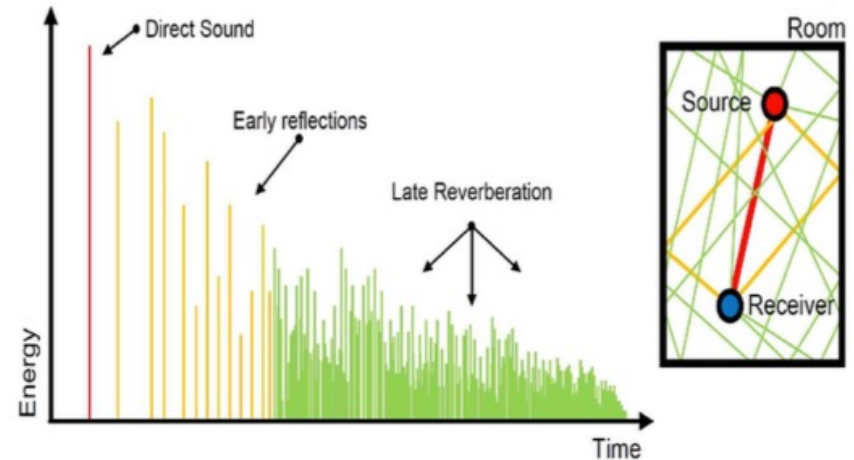
- low wavelength
- require low energy (db) to be perceived
- occlusion / objects change sound
- are absorbed by air
- can be perceived, easily, but close to the source

Auditory Perception and diffusion



- An audio perceived far from its source is subject to different types of distortion - different wavelength
- Also, different versions of the same sound are perceived together, with a **reverb**, because of reflection
- Small delay for directional sound between two ears

Auditory Perception and diffusion

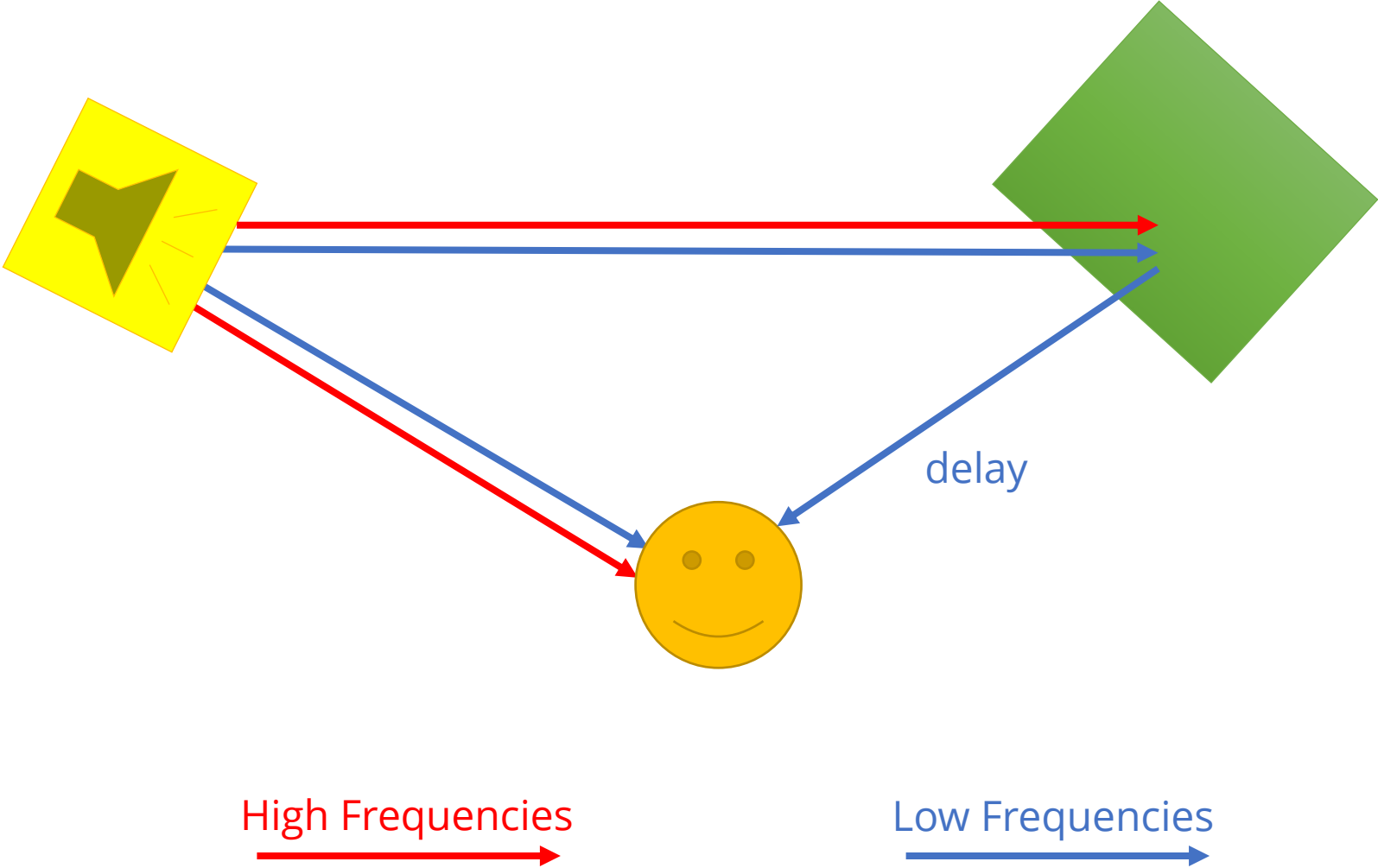


- An audio perceived far from its source is subject to different types of distortion - different wavelength
- Also, different versions of the same sound are perceived together, with a **reverb**, because of reflection
- Small delay for directional sound between two ears

All of these info (and others) are filtered out by our brain:

- we perceive just one version of the signal ...
- ... but we use other data to localize the position of the source

Sound perception + propagation

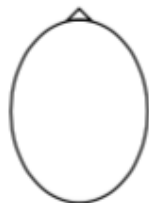


Precedence effect

- If two similar sounds arrive at two different times, only one is perceived
- rather than hearing a jumble, people perceive a single sound.
- based on the first arrival, usually has the largest amplitude.
- echo is perceived if the timing difference is larger than the echo threshold (approx 3 to 61 ms)



Left



Right

Auditory Illusions

The fact that perception involves signal processing that is done by our brain to process its input causes also several auditory illusions

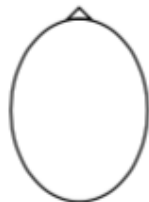
- Precedence effect is one of those
- Glissando Illusion
- Shepard Tone



Illusions have been used to give “effects” to games to play with the players



Left



Right

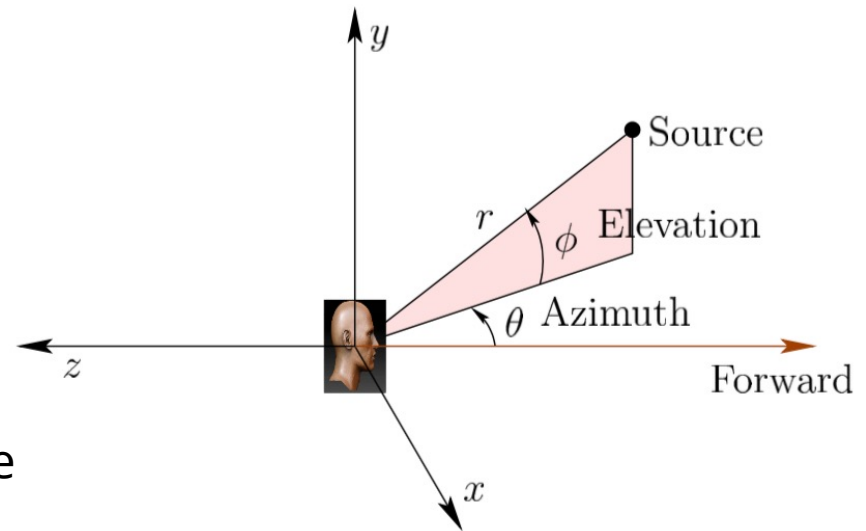
Auditory Illusions



Auditory Perception: Localization

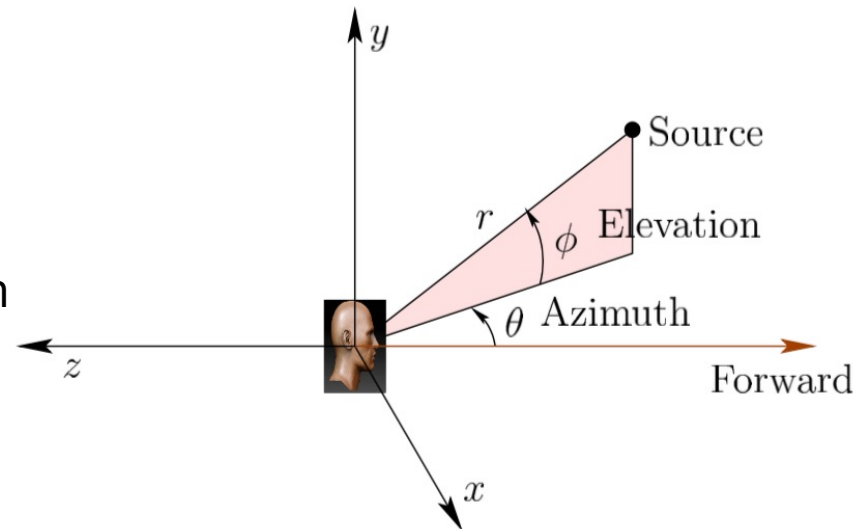
Localization is the estimation of the source of a sound by hearing it.

It is important in games but is particularly important in VR experience as it what personalizes the audio experience for the user allowing a more immersive and realistic perceptual experience.

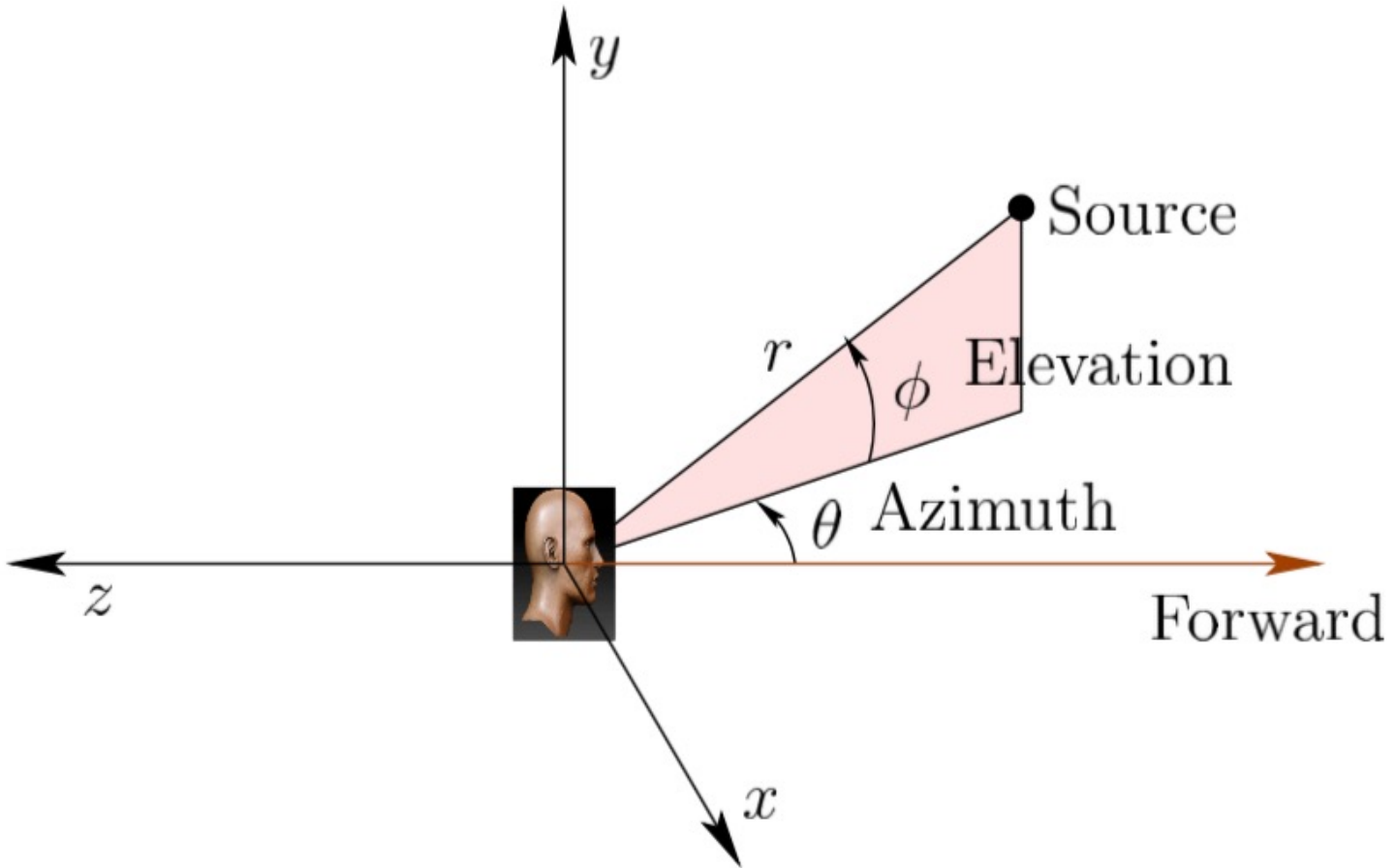


Auditory Perception: Localization

JND (Just Noticeable Differences) in localization (pitch, angle, distance, elevation) are not linear and depend on the several factors (mostly HF and LF)



Measuring distance from the source



minimum audible angle (MAA) depends mostly on frequency and elevation

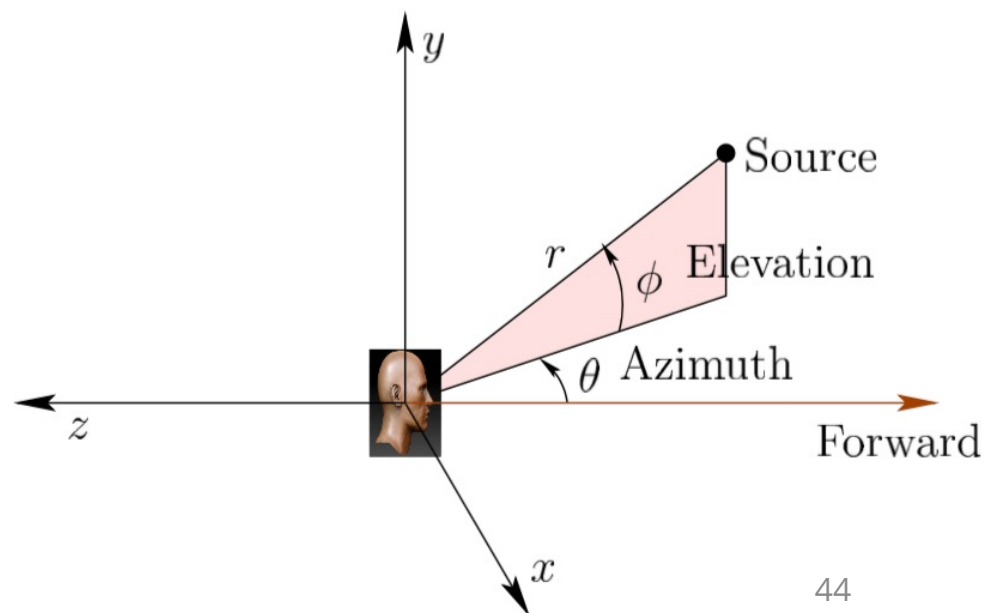
Localization and cues

To localize sound sources we use both:

- Monaural cues rely on sound reaching a single ear;
- Binaural cues based on difference in the same signal when perceived by the two ears jointly.

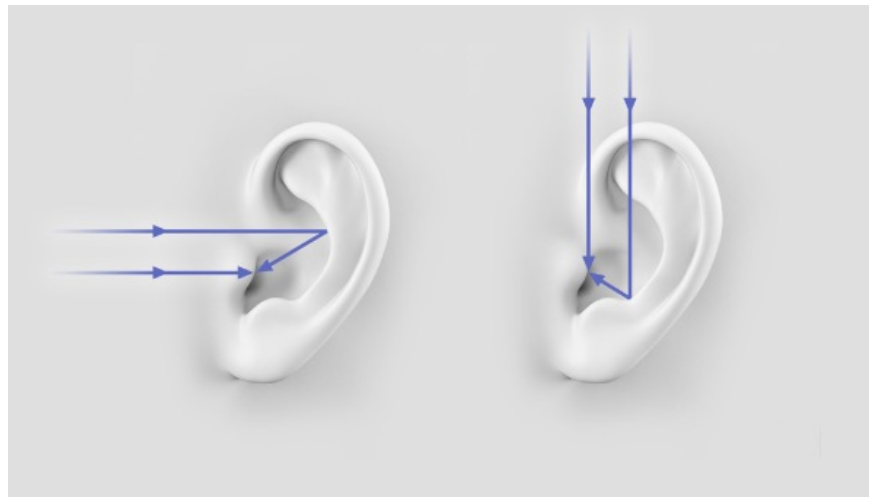
To improve localization we often (and often without noticing) perform movements with our head to improve localization (e.g., tilting moving slightly our head in one of our DOF)

The use of both monaural and binaural cues is similar to what we also do with vision to estimate the distance of an object.



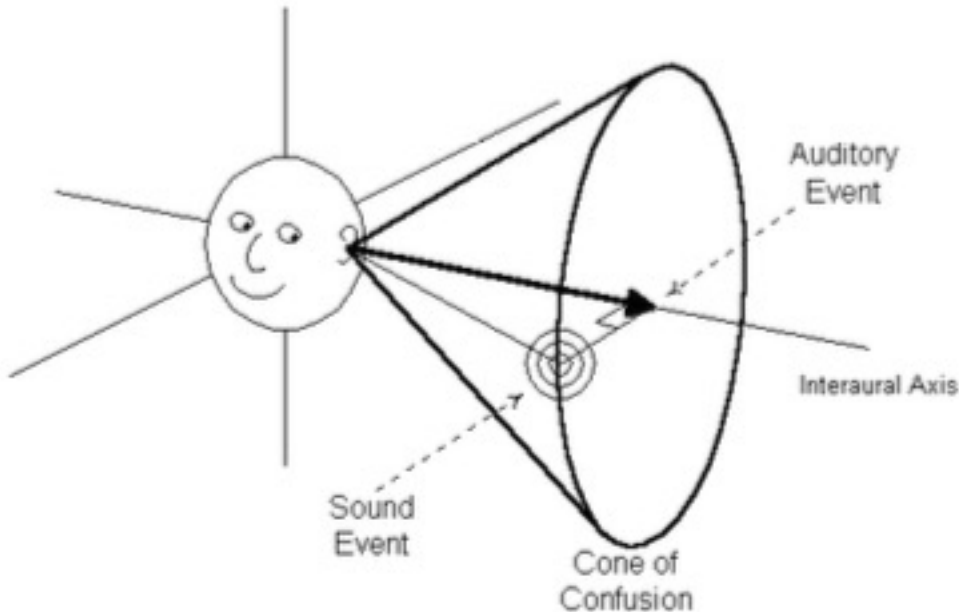
Monaural Cues and localization

- amplitude decrease quadratically w.r.t. distance;
- distant sounds are heavily distorted due to different wavelength attenuation (so, if we know the original sound we can estimate its distance)
- ears are asymmetric; the auricle/pinna distort the sound to identify its source (especially elevation). We do not perceive such distortion, only use it for localization
- reflections and repetition in both ears are particularly important (room / indoor). Indoor we perceive multiple version of the same sound, we filter them and use echoes and reverb for *echolocation*



Binaural Cues and localization

- *Interaural level difference (ILD)* – sound magnitude difference – can perceive *acoustic shadows* (attenuation due to occlusion);
- *Interaural time difference (ITD)* – delay perceived between two ears (~0.6ms between the ears) when listening the same source
- *Head motion* – Doppler effect - triangulation

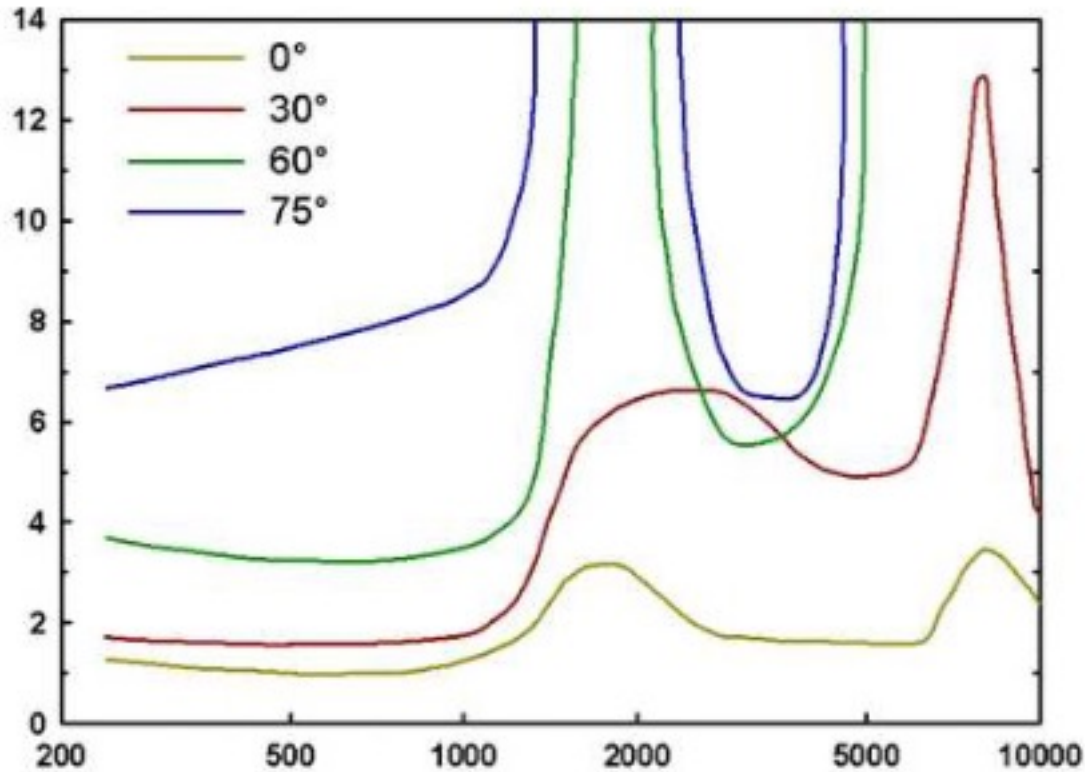


We use binaural cues to create a *cone of confusion* where the audio source may be. Slight movements of the head are used to triangulate the sound source and reduce uncertainty

Binaural Cues and localization



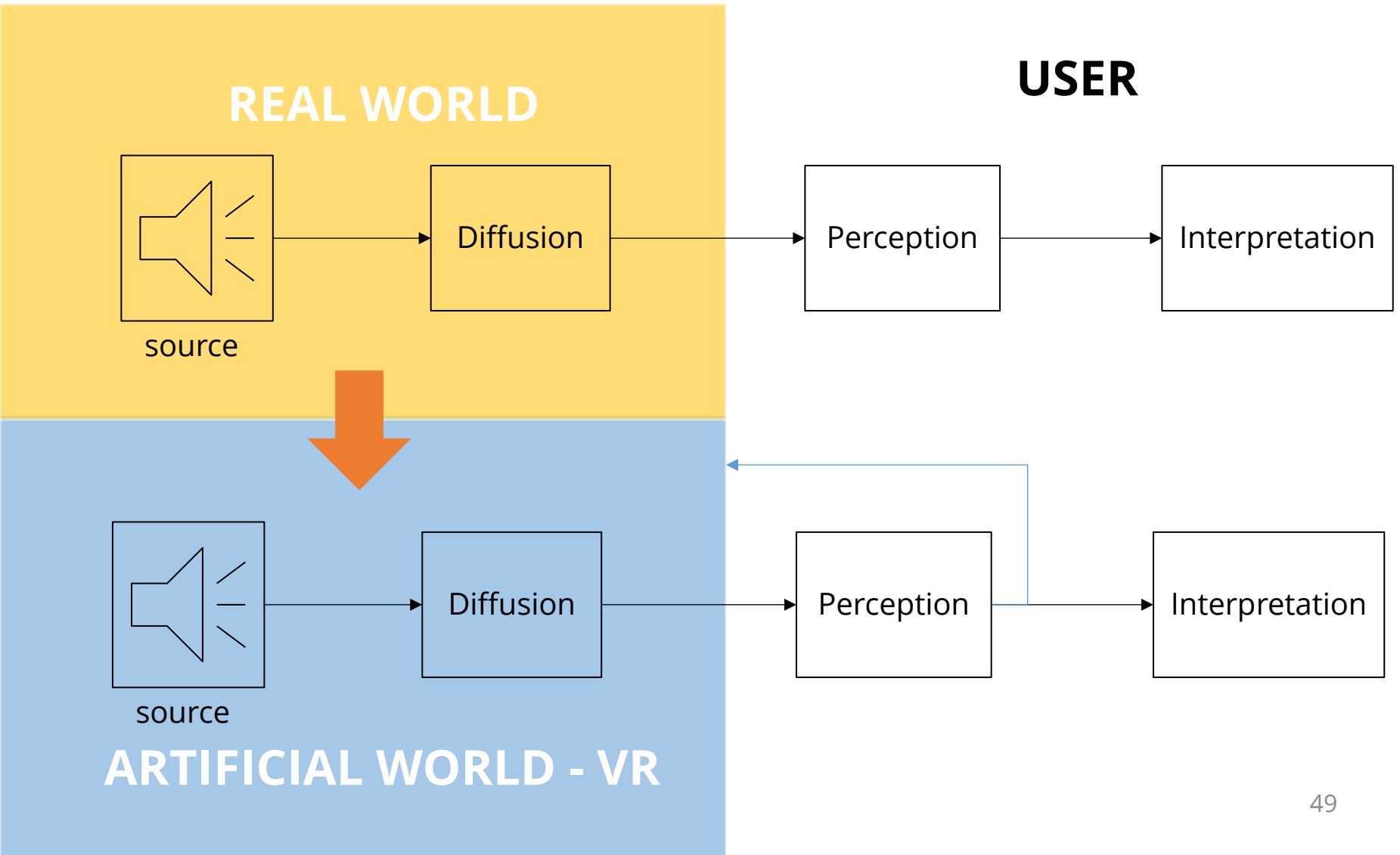
Minimum Audible Angle



Also here, we are better / worse in localizing HF/LF especially when we estimate Azimuth/Elevation

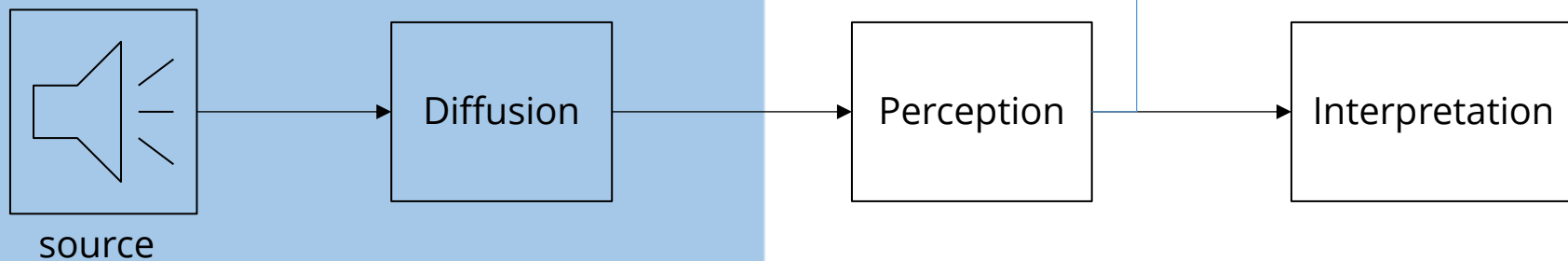
(colors indicate different Azimuth angles)

Audio modelling for VR



- What environment / world?
- What sounds? What samples?
 - How can we record good sounds for VR?
 - Which samples to use?
 - Where to put those?
- What propagation/diffusion model?

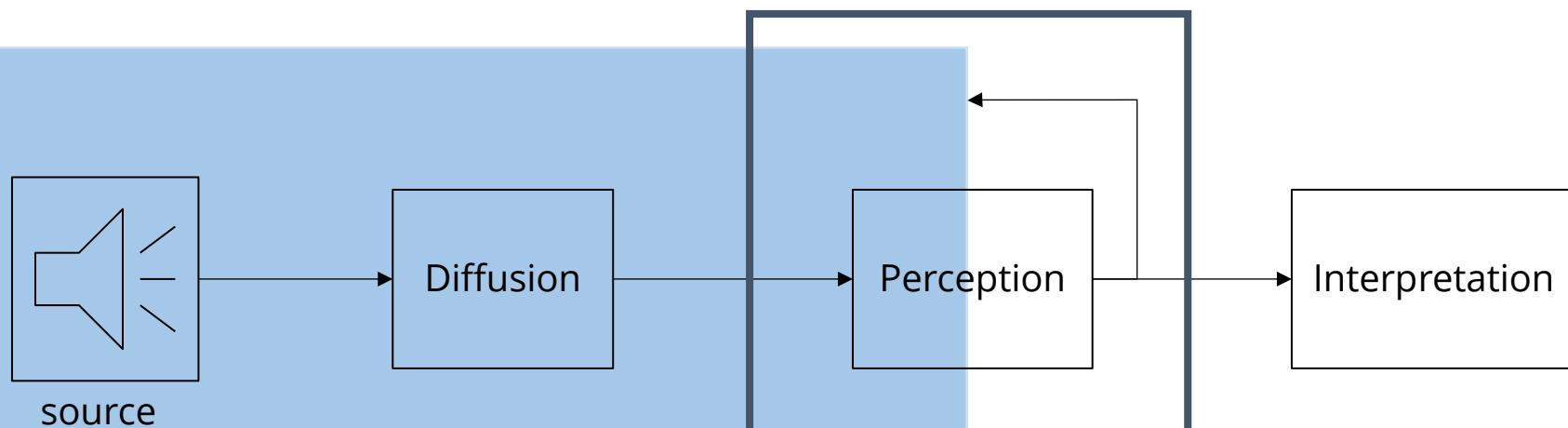
In order to have a proper spatialized audio – one that can be used in a AR or VR setting by a user for localizing the audio source / the movement of the audio source, we have to deal with all these questions.



ARTIFICIAL WORLD - VR

- What environment / world?
- What sounds? What samples?
 - How can we record good sounds for VR?
 - Which samples to use?
 - Where to put those?
- What propagation/diffusion model?
- **What perception model?**

A key component of spatialized audio is to model also some characteristics of perception (how our ears perceive sound) so to control this step by exploiting the fact that earphones bypass the outer pinna / auricle



ARTIFICIAL WORLD - VR

Auditory rendering

Virtual audio should be consistent with visual cues + past auditory experiences.

To do so we transform the signal according to the user movements

- Techniques: spectral decomposition – signal processing
 - Frequency domain
 - Fourier analysis
- *Filters* – transform and distort the signal
 - Sampling rate = 2*max frequency ~ 40 kHz = 44100 Hz

E.g.; Linear filters

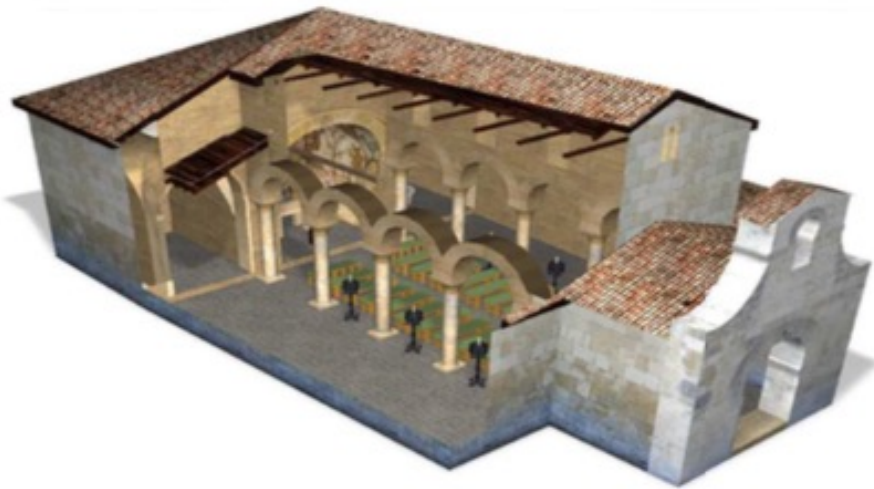
$$y[k] = c_0x[k] + c_1x[k - 1] + c_2x[k - 2] + \dots + c_nx[k - n]$$

$$y[k] = \frac{1}{2}x[k] + \frac{1}{4}x[k - 1] + \frac{1}{8}x[k - 2] + \frac{1}{16}x[k - 4]$$

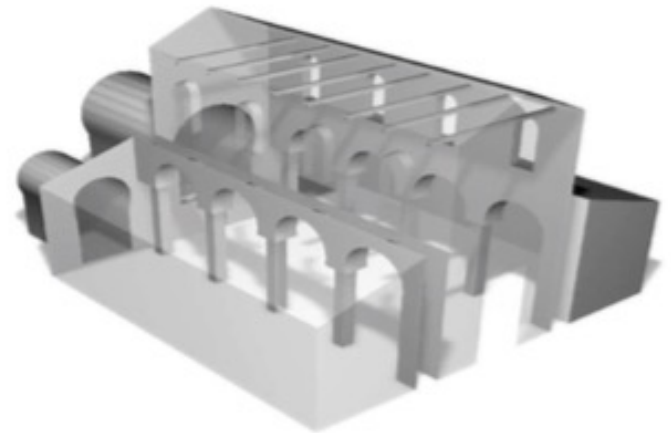
Example: exponential smoothing

Acoustic modelling

Room model for audio rendering can be much more simpler than those used for visual rendering



(a)



(b)

small objects are invisible to sound – spatial resolution of 0.5 m
We can focus only on large elements / materials

Acoustic modelling

Room model for audio rendering can be much more simpler than those used for visual rendering

However:

- different shapes reflect sound wave differently
- Sound waves propagates differently in materials
- Smaller objects / corrugated objects (e.g. bricks) can results in scattering

Audio propagation is difficult to simulate – for hi-fi performance, where user is static, the “best” solution is binaural recordings



Neumann KU1000

Acoustic modelling

Room model for audio rendering can be much more simpler than those used for visual rendering

Audio propagation is difficult– for hi-fi performance the “best” solution is binaural recordings.

However in binaural recordings the head is “fixed” – useful for replicate how the sound is perceived by a passive listener.

For VR a “360” audio is more useful – ambisonic sounds.

E.g. YouTube 360 supports First Order Ambisonic Sounds (L/R/F/B)



Ambisonic Sounds

Ambisonics sounds are full-sphere recordings, trying to record the sound at 360°.

Note that a 360° recording is very different from what we perceive as we “listen” to the sound *after* our brain processed it, while ambisonics capture the raw signal.

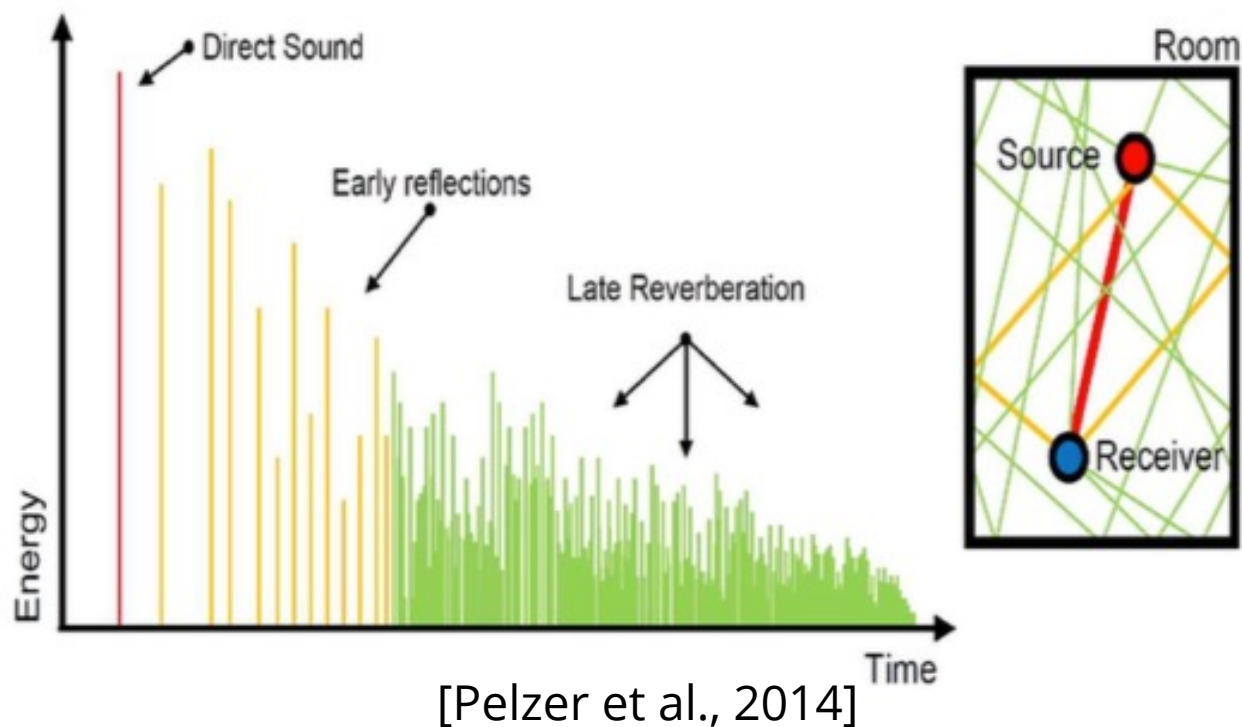
When using ambisonics is the artificial world/engine that processes the sound so that the users perceives it correctly



Acoustic modeling

Two components:

1. how sound propagates in the environment
2. how it's perceived



Acoustic modeling

Two components:

1. how sound propagates in the environment
2. how it's perceived

Numerical Sound Propagation:

Helmholtz wave equation: closed form solution often do not exist, computationally expensive;



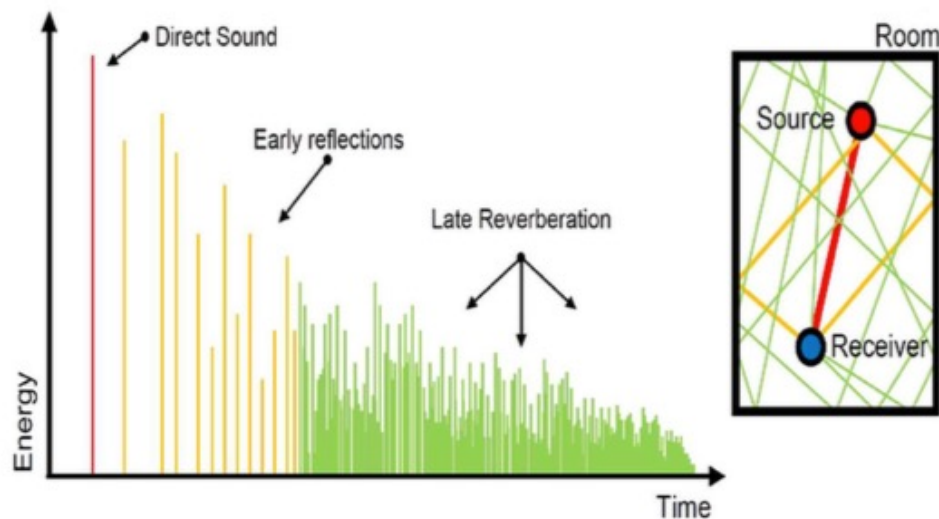
Acoustic modeling

Two components:

1. how sound propagates in the environment
2. how it's perceived

Visibility-Based Sound Propagation:

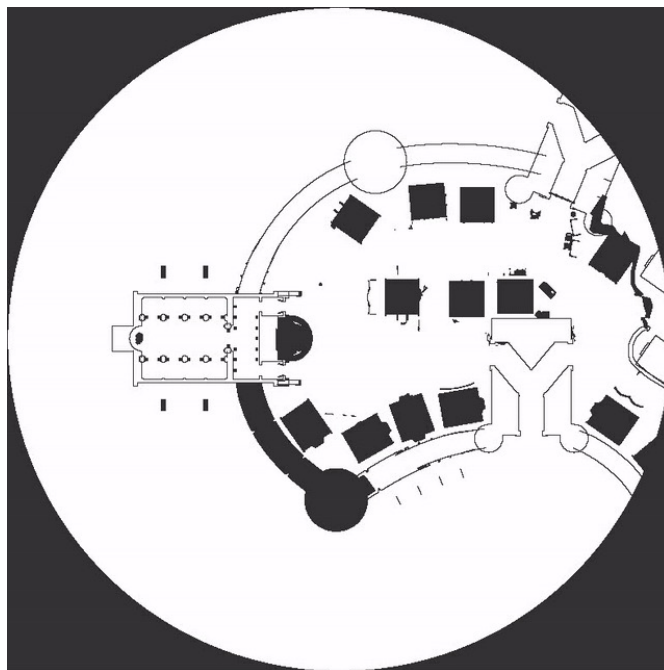
Works well, similar to ray tracing, done using Montecarlo Methods



Sound propagation

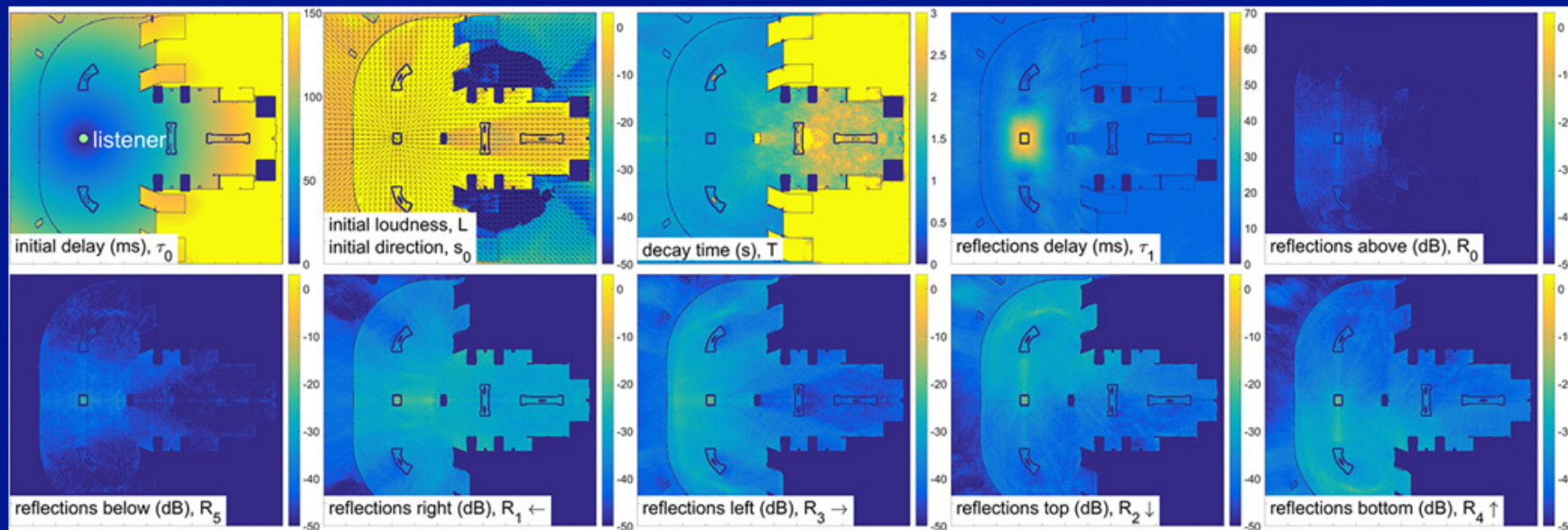
Microsoft Project Acoustic:

- Compute a propagation model for a given environment using a cloud-based solution on Azure
- Pre-computed model is used to process real-time wave diffusion



EX: Microsoft Project Acoustic

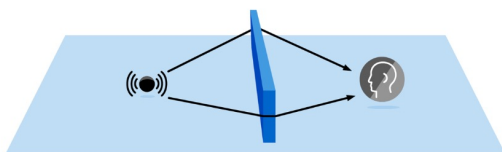
Available as a plugin for Unreal and Unity



EX: Microsoft Project Acoustic

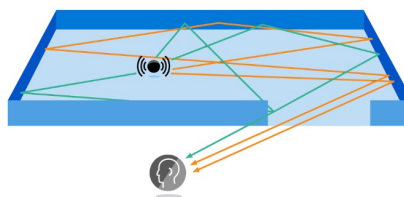
Available as a plugin for Unreal and Unity

Obstruction



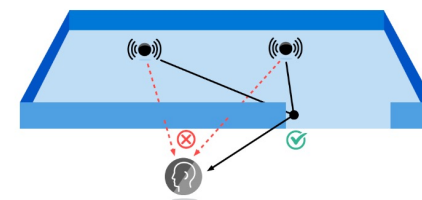
Sound is weakened when it diffracts around obstructions.

Occlusion



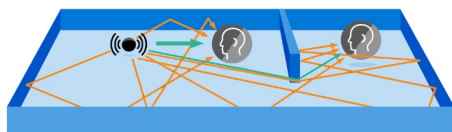
Occlusion is total reduction in loudness from geometry, involving complex propagation and diffraction (diffracted paths in green).

Portaling (sound redirection)



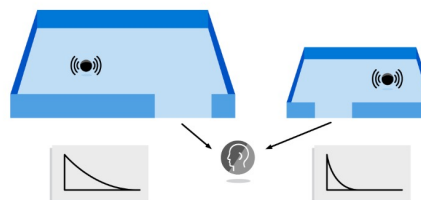
Sound heard "through the wall" (red) from source direction is incorrect. Doors redirect sounds, which is more believable.

Reverberance



Left: Close to the source, direct path (green) is loud compared to reflections (orange), resulting in high clarity and low reverberance. Right: Behind the partition, direct path is weakened from diffraction, causing low clarity and high reverberance, conveying that the source is in a different room.

Decay time



Larger rooms reverberate longer.

Head-Related Transfer Function (**HRTF**)

Two components:

1. how sound propagates in the environment
2. how it's perceived

Perception involves modelling distortion due to our own head and ears.

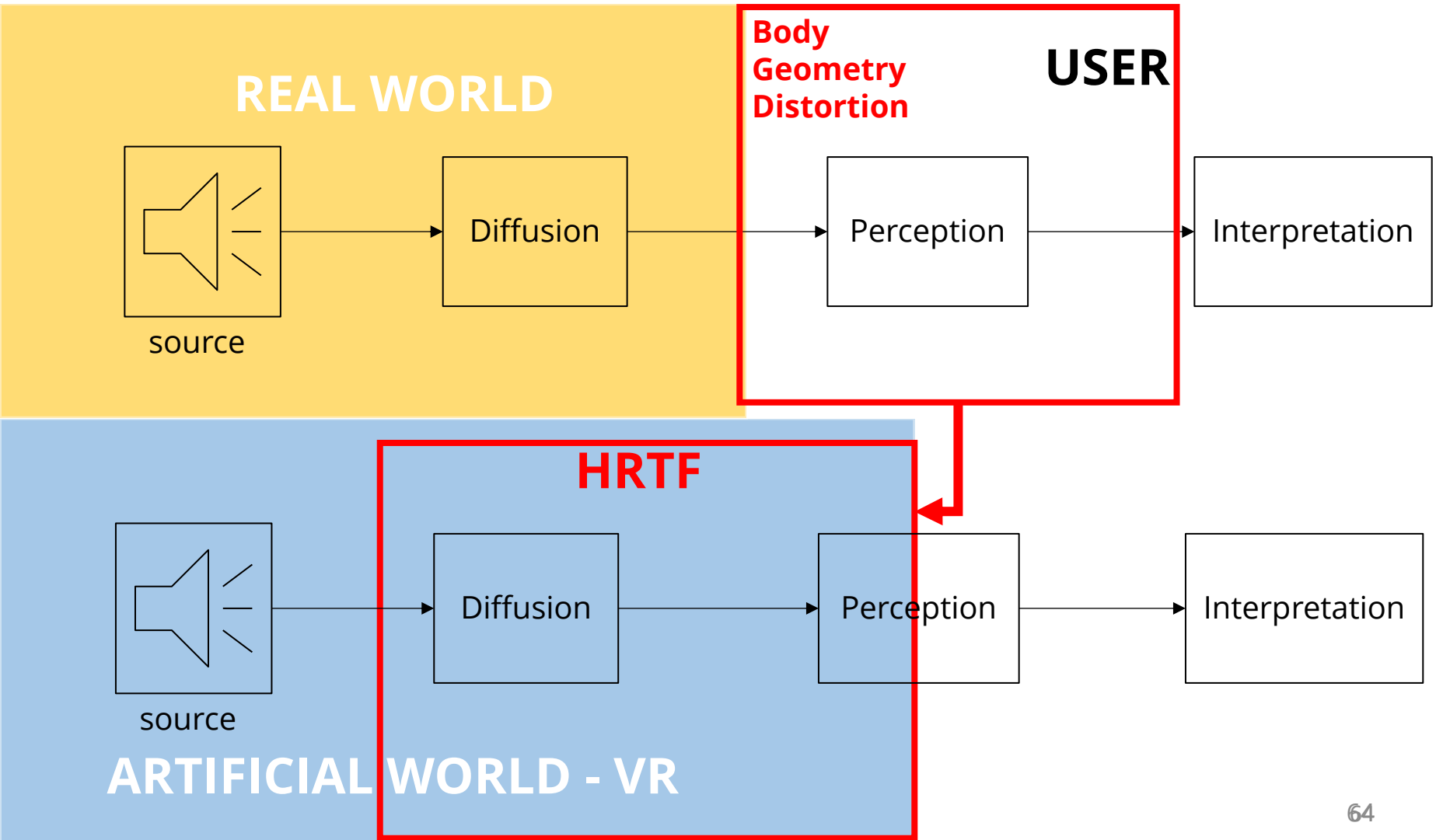
This is done using

Head-Related Transfer Functions (**HRTF**) are:

- Linear filter that distort sound by simulating how the sound is perceived source-head
- Model are approximate (HRTF “should” be tailored on each user measuring inner-ear components)



Head-Related Transfer Function (**HRTF**)



Head-Related Transfer Function

HRTFs provide influenced filtering of the sound applied to both ears:

- micro-delay between ears
- Modeling the directional filtering that ear-flaps, the head itself and the shoulders contribute to.

Adding HRTF filtering already immensely improves the sensation of direction over a conventional panning solution.

Direct HRTF is a part of the solution, as it models only the direct path of audio and not how it is transmitted in space (e.g., occlusion) – and sound reflection (similar to global illumination problem in graphics).

Unity provides an interface for Spatial Audio using HRTF
<https://docs.unity3d.com/Manual/AudioSpatializerSDK.html>

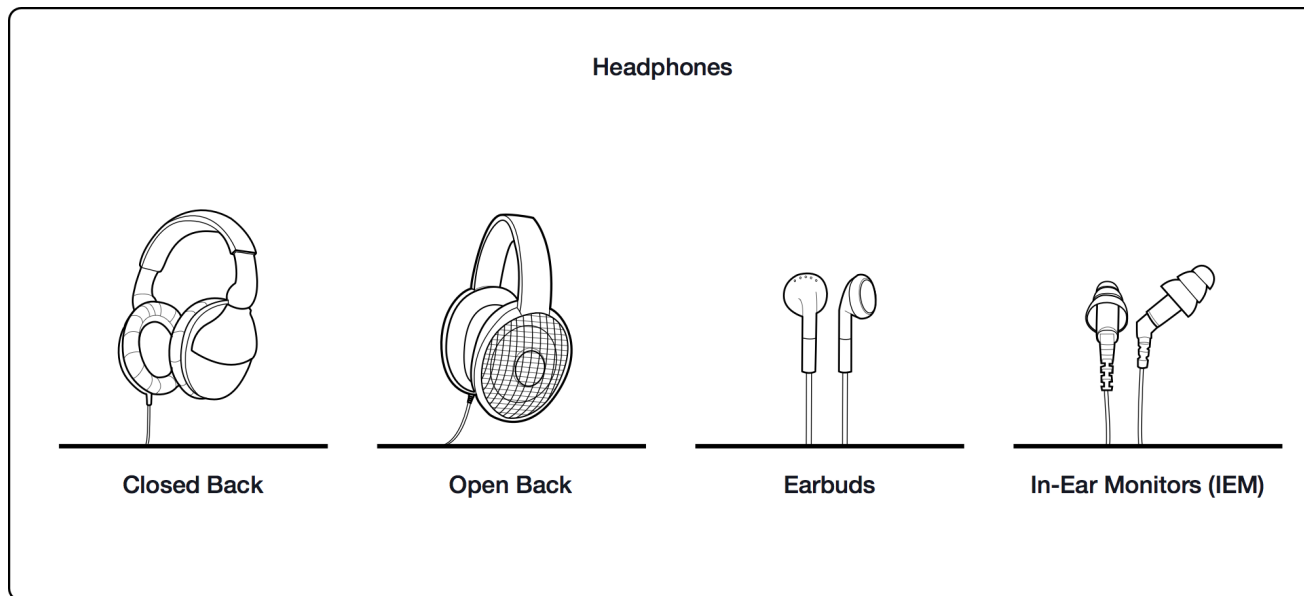
Head-Related Transfer Function

- HRTF are a highly-parametrized transfer function which model the head-related distortion *for each user*
- HRTF computes how audio is perceived from inside the user head
- Using headphones you can bypass the distortion of your own ears and use the simulated one instead

However, the estimation of such parameters is complex

- Solution: «*borrow*» the head of an average listener using an average HRTF of someone else

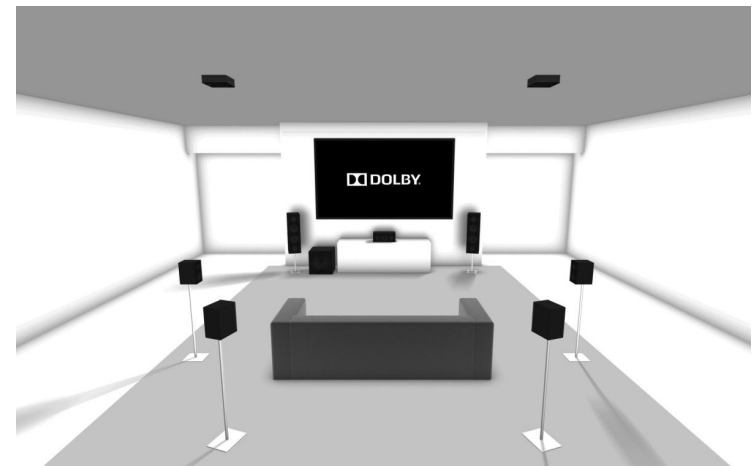
Audio devices for VR



Audio devices for VR

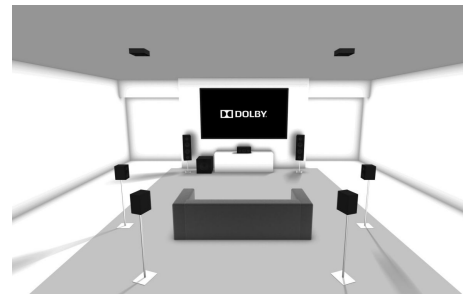
Several type of output devices:

- Stereo Speakers
- Surround Systems
- Headphones



Bluetooth introduce latency (up to 0.5s) which can be annoying for VR, so it should be avoided

External Speakers



External speaker are world-fixed, so not the best solution for VR

- Imprecise imaging due to panning over large portions of the listening area.
- No elevation cues. Sounds only appear in a 360 degree circle around the listener.
- Assumption of immobile listener; in particular, no head tracking.
- (real) Room effects such as reverberation and reflections impact the reproduced sound.
- Poor isolation means that outside sounds can intrude on the VR experience.

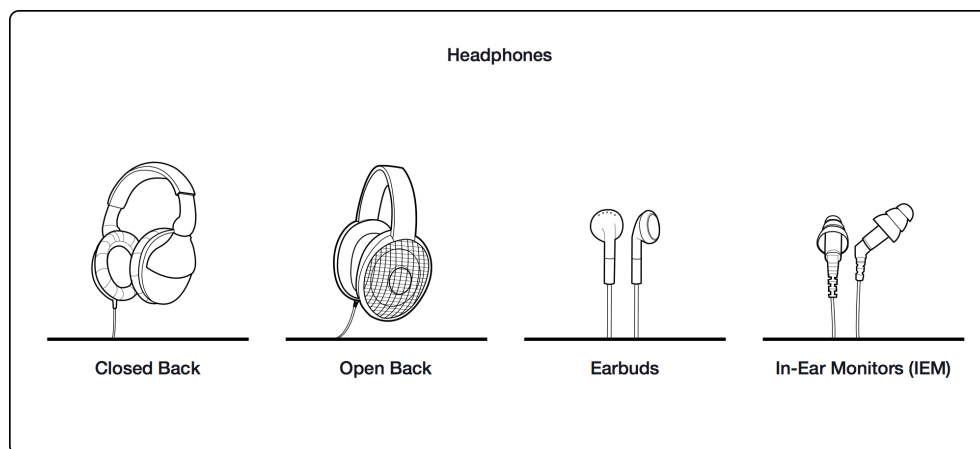
But:

- LF could be “felt” and perceived using subwoofers

Headphones

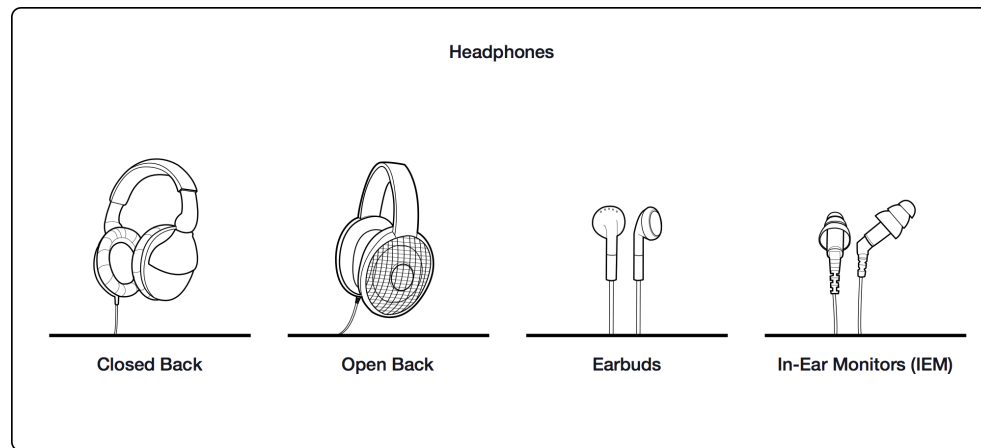
Headphones are the best option for VR:

- Move with user head
- Acoustic isolation
- Don't suffer from the “doubling down” of HRTF effects, i.e. sounds being modified from the simulated HRTF, and again by the listener's actual body geometry



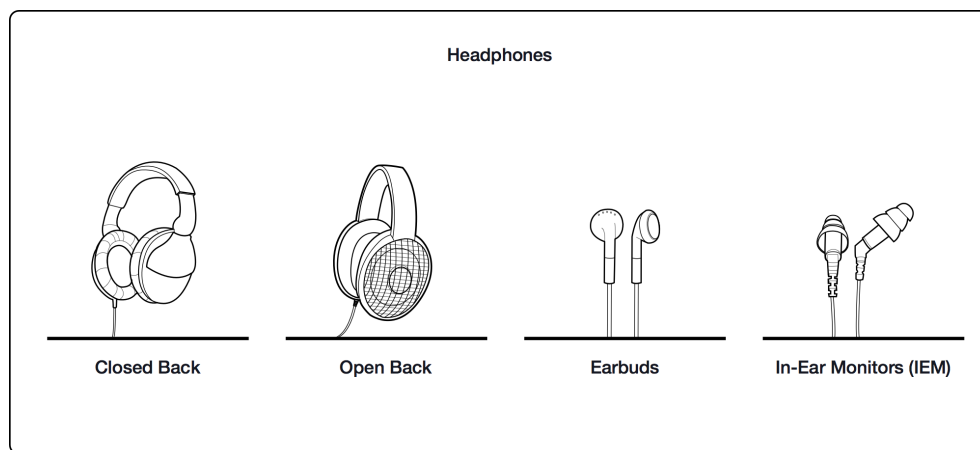
On-ear Headphones

- On-ear headphones reproduce both LF and HF
- LF are less “felt” than subwoofers, which can be used in combination with them
- Provide isolation (not optimal though)
- Outer-ear distortion and HRTF double-down effect slightly present



In-ear Headphones

- In-ear headphones / earbuds poorly reproduces LF
- Better isolation
- Outer-ear distortion and HRTF entirely removed



Try it yourself

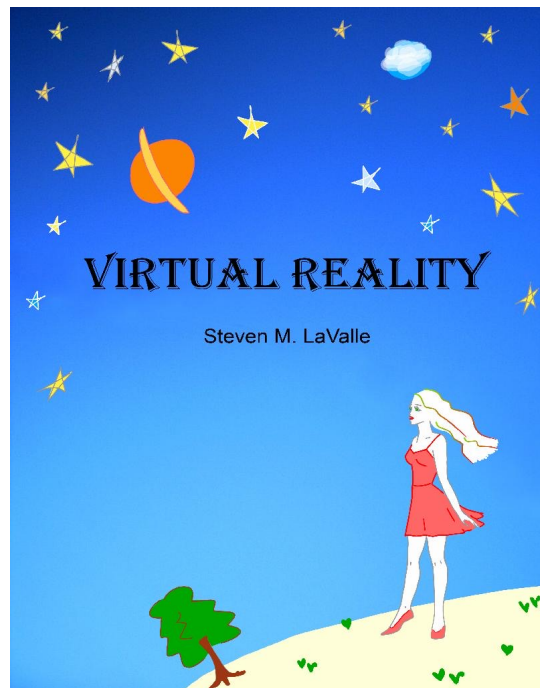
<https://cdn.rawgit.com/resonance-audio/resonance-audio-web-sdk/master/examples/birds.html>

References

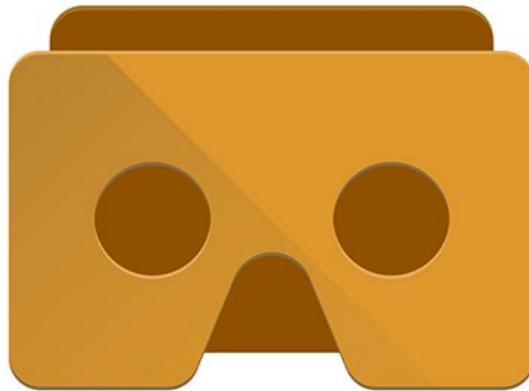
Steven La Valle, Virtual Reality, Cambridge Press

<http://lavalle.pl/vr/>

Book Available (free) online on the page of the author



Spatialization libraries and SDKs



Spatialization libraries and SDKs

There exist several spatialized audio libraries:

- Oculus Audio SDK

 - <https://developer.oculus.com/learn/audio-intro-localization/>

 - HRTF
 - No world model

- Microsoft VR

 - HRTF

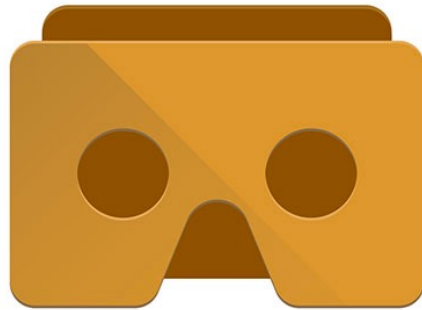
 - <https://github.com/microsoft/spatialaudio-unity>

 - Wave acoustic propagation with Azure cloud computation with Project Acoustic
<https://docs.microsoft.com/en-us/gaming/acoustics/what-is-acoustics>

Spatialization and Google VR: Resonance Audio

Two components:

1. how audio propagates in the environment
2. how it's perceived



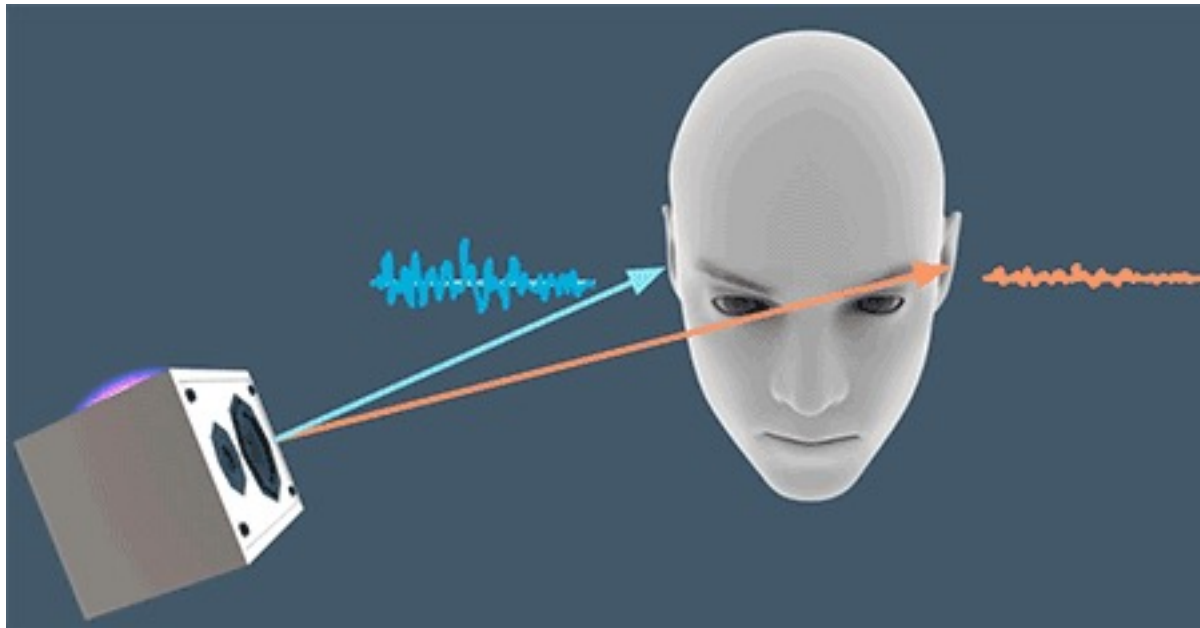
Resonance Audio SDK simulates both audio propagation and perception.

<https://resonance-audio.github.io/resonance-audio/>

Resonance Audio and HRTF

Simulates main audio cues used for localization:

- Interaural time differences
- Interaural level differences
- Spectral filtering done by outer ears



Resonance Audio and HRTF

Simulates main audio cues used for localization:

- Interaural time differences
- Interaural level differences
- Spectral filtering done by outer ears

For higher frequencies, humans are unable to discern the time of arrival of sound waves. When a sound source lies to one side of the head, the ear on the opposite side lies within the head's acoustic shadow.

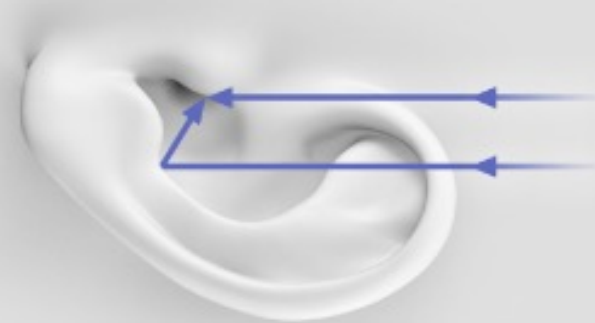
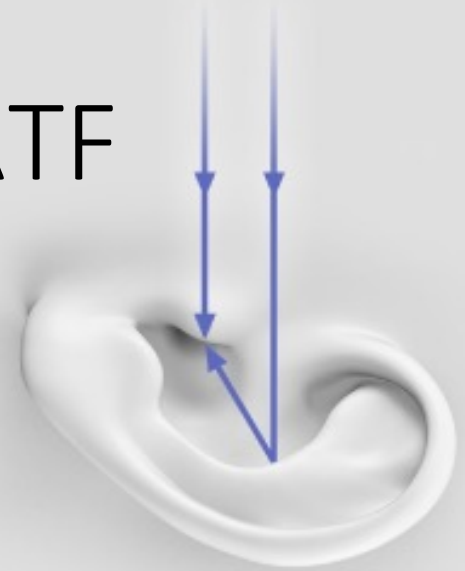
Above about 1.5 kHz, we mainly use level (volume) differences between our ears to tell which direction sounds are coming from.

Resonance Audio and HRTF

Simulates main audio cues:

- Interaural time differences
- Interaural level differences
- Spectral filtering done by outer ears

Sounds coming from different directions bounce off the inside of the outer ears in different ways. The outer ears modify the sound's frequencies in unique ways depending on the direction of the sound. These changes in frequency are what humans use to determine the elevation of a sound source.



Resonance Audio systems surrounds the listener with a high number of *virtual loudspeakers* to reproduce sound waves coming from any direction:

- Denser the array, higher the accuracy
- Computed using HRTF (and to simulate cues)
- Support Ambisonics



Ambient is modelled as an **AUDIO ROOM**

- Box-model – only size and material is specified for each material
- Reflection and scattering is simplified (real-time)

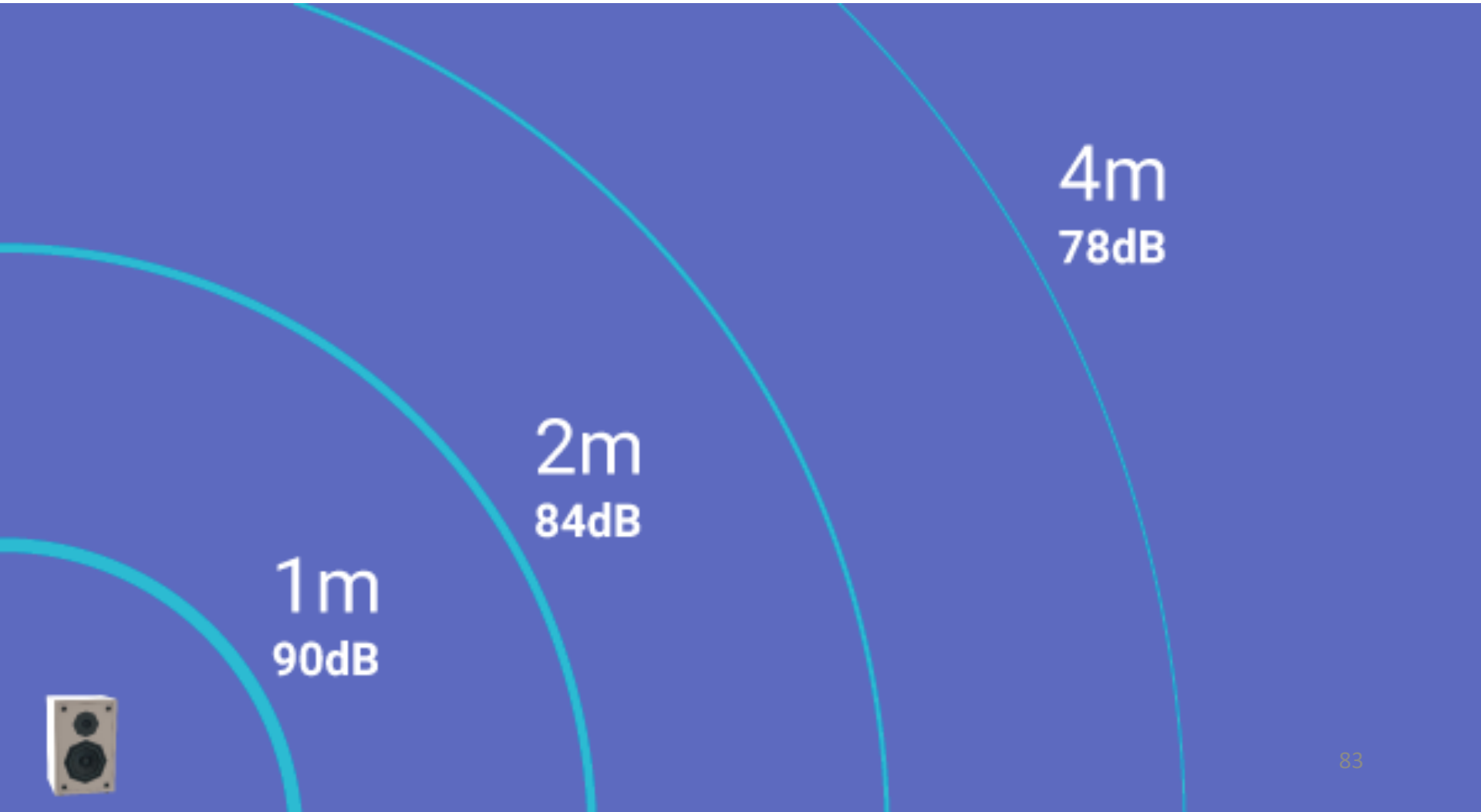
Occlusion blocks/transmits differently low and high frequency



Resonance Audio and Propagation

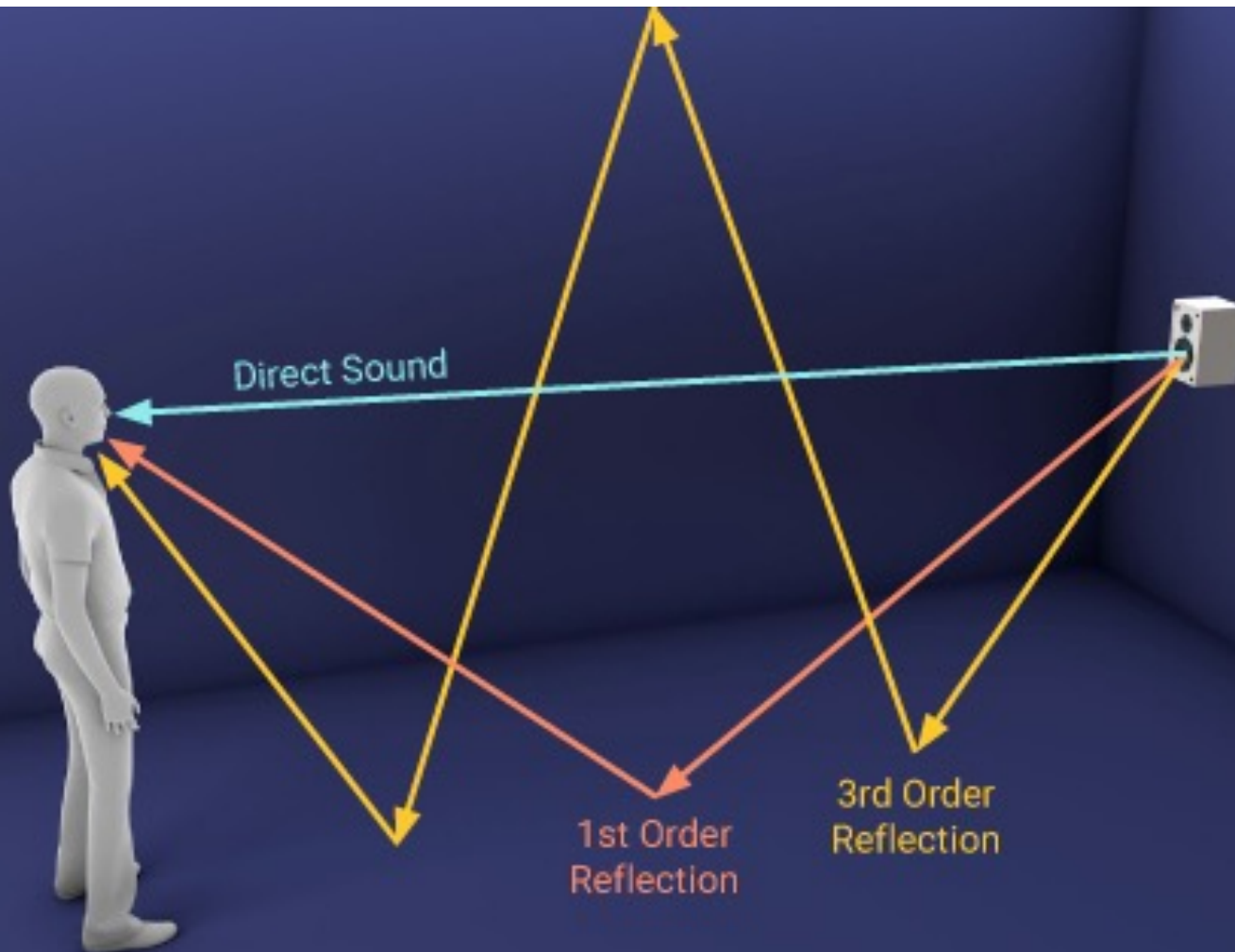
Propagation is modelled as three components

- Direct sound
- Early reflection
- Late reverb



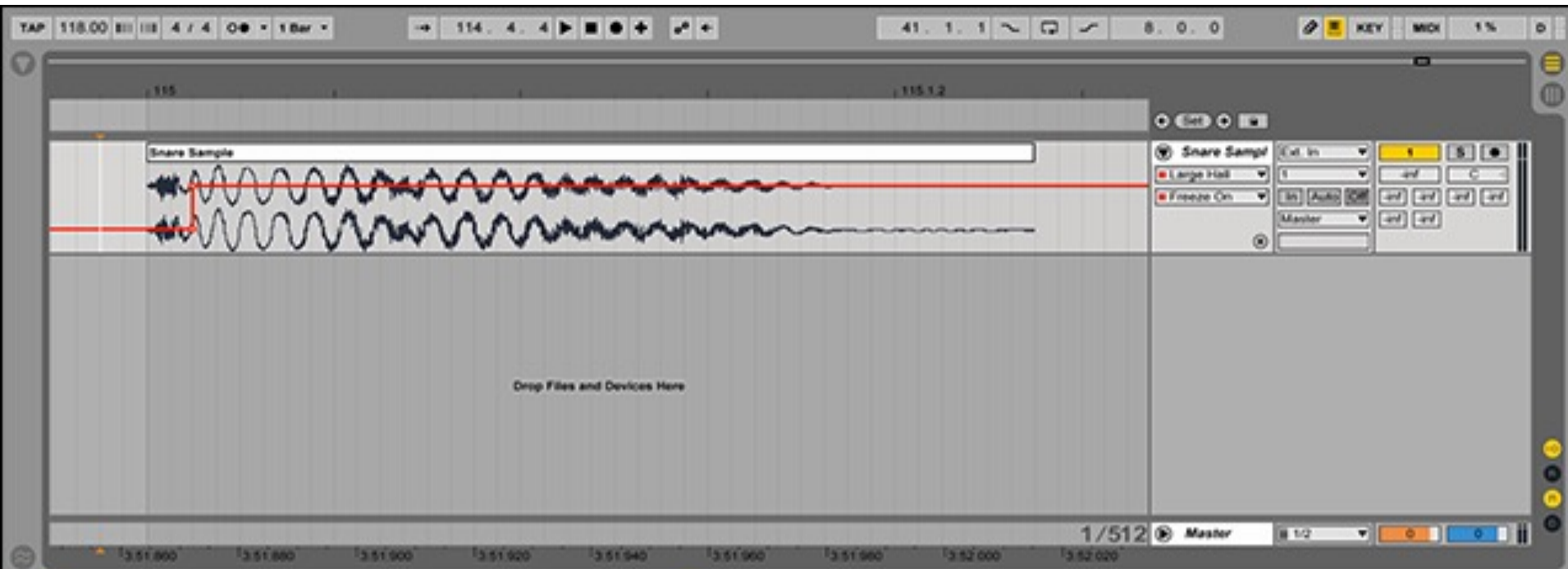
Propagation is modelled as three components

- Direct sound
- Early reflection
- Late reverb



Propagation is modelled as three components

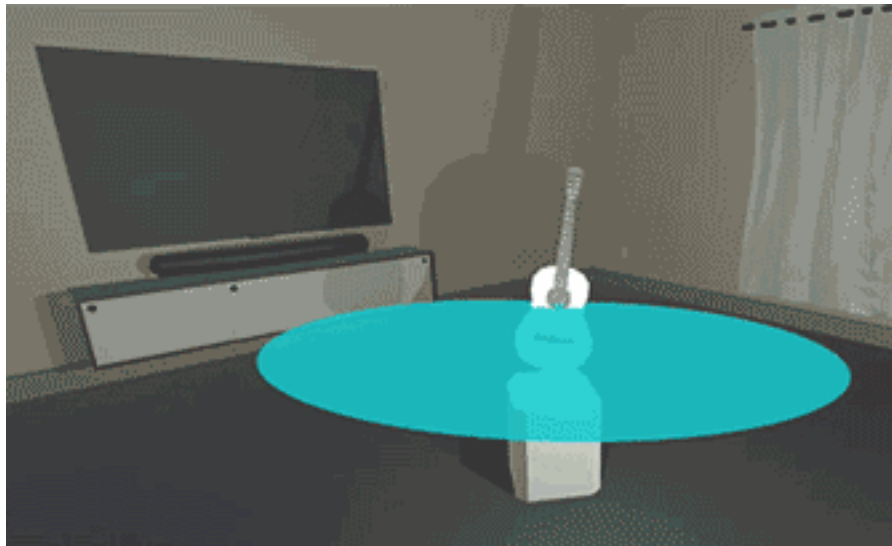
- Direct sound
- Early reflection
- Late reverb



Directivity and Head Movements

Each object has a **directivity pattern** – a model which can be used for characterize the object

Head movements are used to change time level and frequency cues for improving localization.



Directivity and Head Movements

Each object has a **directivity pattern** – a model which can be used for characterize the object

Head movements are used to change time level and frequency cues for improving localization.



Design tips

Spatialized audio is a combination of:

- Panning
- Reverb
- Distortion

But most of the work is done automatically by the engine (for Google VR) or by the player (e.g. YouTube 360)...

...but you can play with it to obtain what you want (as is done with music production – a sound that in „normal“ listening seems realistic is not necessary better for VR)



Design tips

VR audio is somehow “dual” of music recording.

- A music track is recorded usually with different microphones, all independent and close to the sound--source
- Songs are created by mixing / compressing / panning tracks together in order to simulate, in the final stereo mix, the feeling of the user listening to the track as being in the room

While in VR audio the signal processing is taken care by the environment, so audio signal should be less processed and more similar to the original source.

Try to keep the entire spectrum of sequences within a sound for VR – something that you usually avoid when the audio is recorded/mixed for listening

**FIRMWARE
UPDATE
SUPPORTS
GOOGLE
JUMP VR**

[Download Now »](#)



Ambisonic files should be used for background sounds (e.g. trees, ocean waves, wind).

Normal stereo/mono sounds can be used for standard audio sources (e.g. a guitar, television, speech).

If you want to record ambisonic tracks, easiest thing is a

Zoom H2n
(we have one if you want to try it)

<https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h2n-handly-recorder>

Design tips



Zoom makes also a (little bit more) expensive microphones designed for recording ambisonic sounds for VR

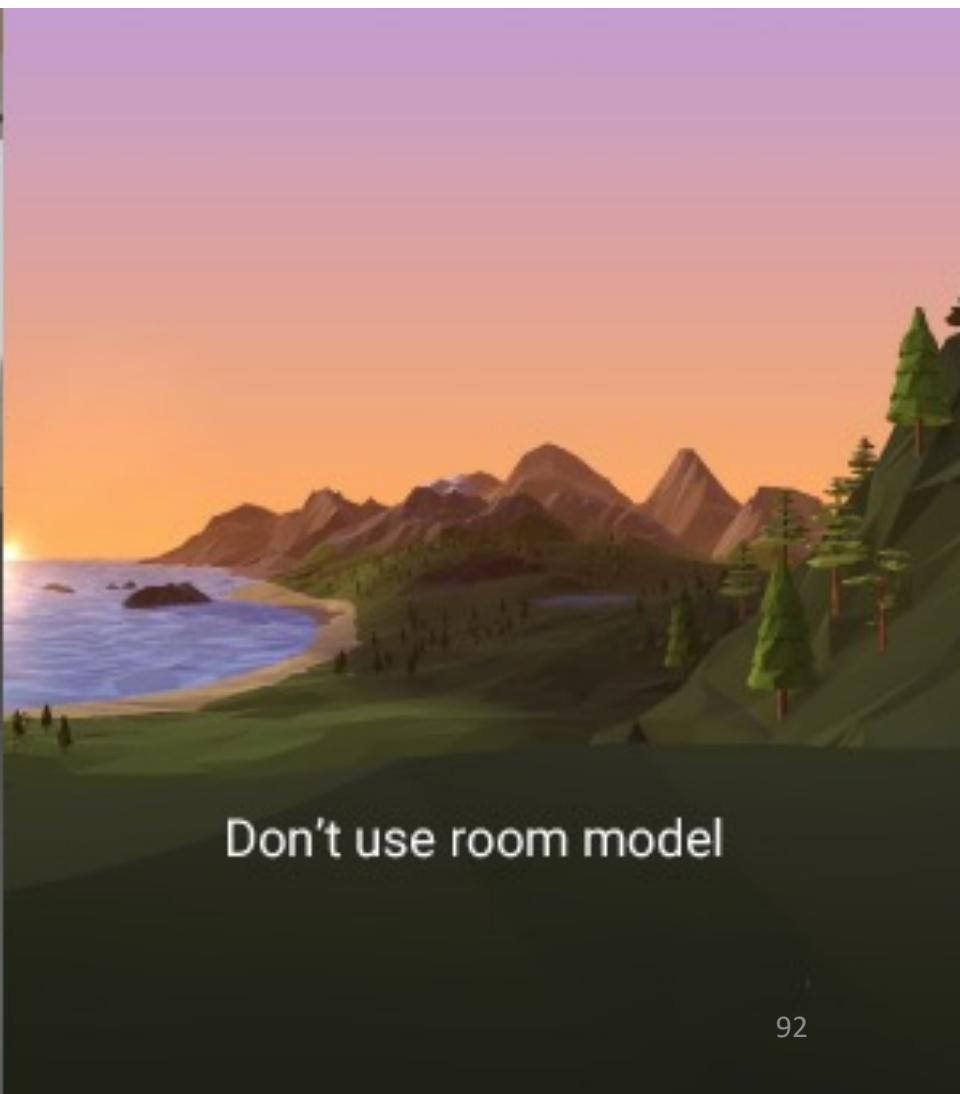
ZOOM H3-VR (<250 euro)
(Zoom H2n is around 100/150 euro)

Audio Rooms provide early reflections and reverb, which help make the sound more realistic when there are nearby walls or structures.

They are - not surprisingly - most useful when your scene takes place in an actual room. For outdoor scenes, an Audio Room can feel less natural, because you may have only one reflective surface (the ground).



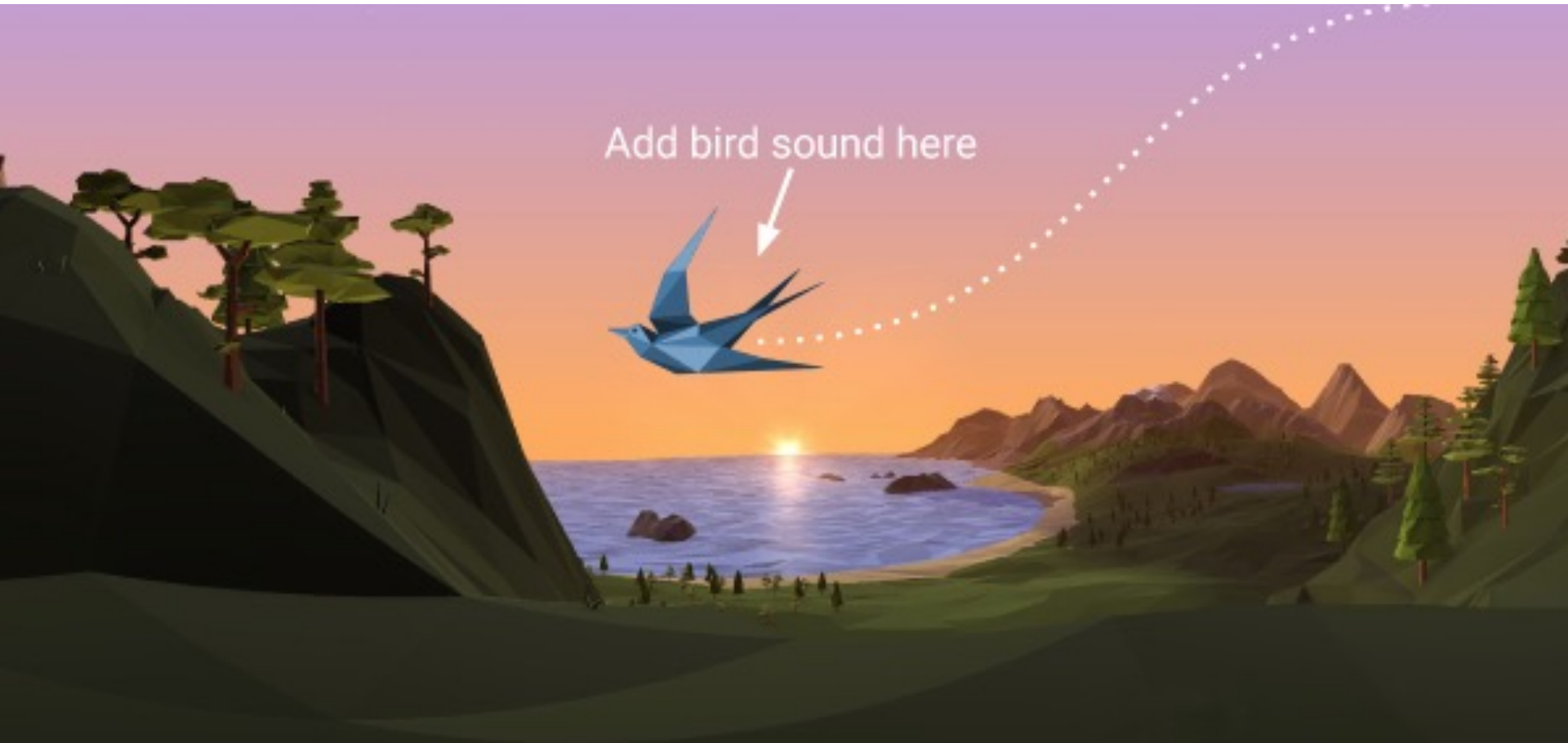
Use room model



Don't use room model

Use carefully Resonance Audio Sources

- mobile audio source moving around the listener can be very useful
- animated sound source can also be used



For background sounds use Resonance SoundField Sources and – if possible – ambisonic sounds



Repeat the sound

If users cannot localize it the first time, they can move the head accordingly to disambiguate



Use more complex sounds – uncompressed, raw, a lot of different sources – even noisy but “full” and “rich” samples



Remember how the intensity of a stimulus is correlated to the probability of its detection

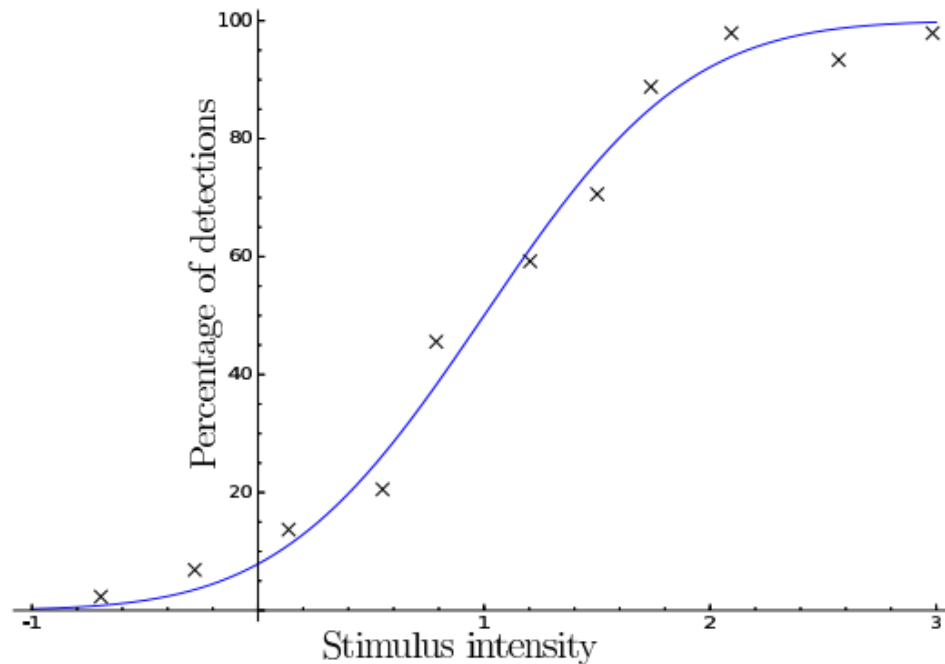


Figure 2.21: The most basic *psychometric function*. For this example, as the stimulus intensity is increased, the percentage of people detecting the phenomenon increases. The point along the curve that corresponds to 50 percent indicates a critical threshold or boundary in the stimulus intensity. The curve above corresponds to the cumulative distribution function of the error model (often assumed to be Gaussian).

Design tips

Avoids “produced” audio track as it can be perceived as less realistic;

As an example, most music is mixed in stereo.

- Since VR uses stereo headphones, it’s tempting to play stereo sounds without spatialization.
- The drawback is these stereo sounds will not be positioned in the virtual world and will not respond to head tracking.
- This makes the sounds appear “head locked”, as they follow the users head movements rather than feeling grounded in the virtual world. This can detract from the spatial audio experience and should generally be avoided when possible.
- For original compositions it’s best to mix to ambisonics which can be rotated and won’t be headlocked. If that is not an option then try to be mindful of how the music impacts the spatial audio.

Resonance Audio supports:

- Android Audio
- iOS
- Unity
- Unreal
- Web
- +
- Wwise
- FMOD

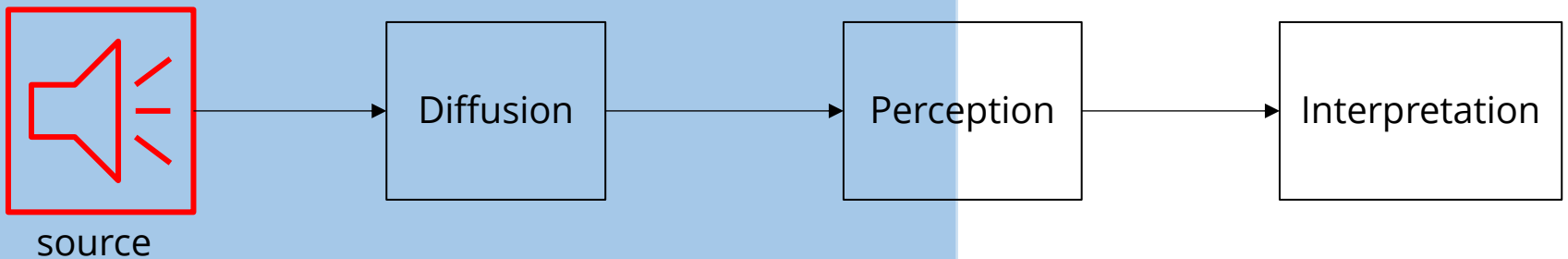
Demo – SDK for Web

<https://cdn.rawgit.com/resonance-audio/resonance-audio-web-sdk/master/examples/hello-world.html>

Try the example on your own to experience how HRTF and room model work on your own browser and the differences between standard and spatialized audio

Summing Up

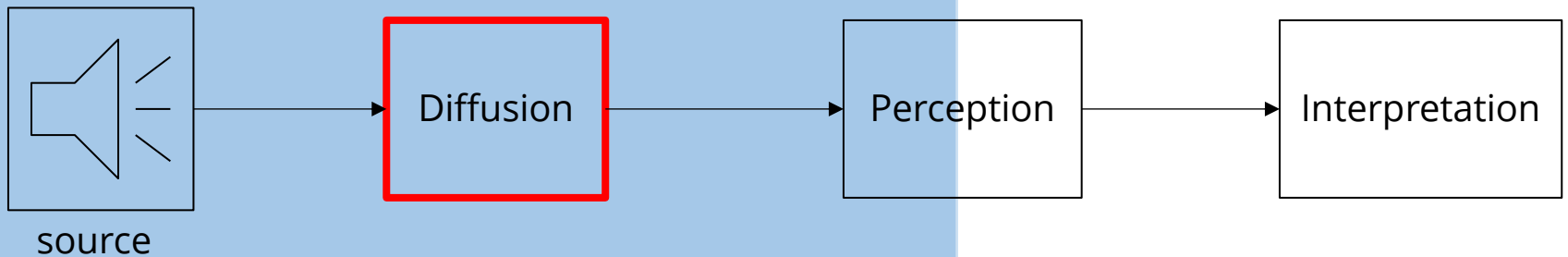
- Select carefully audio sources
- Use rich and uncompressed / unprocessed sounds
- If possible, use ambisonics
- Remember that only large-scale features impact sound (so for audio modeling details are not needed)



ARTIFICIAL WORLD - VR

Summing Up

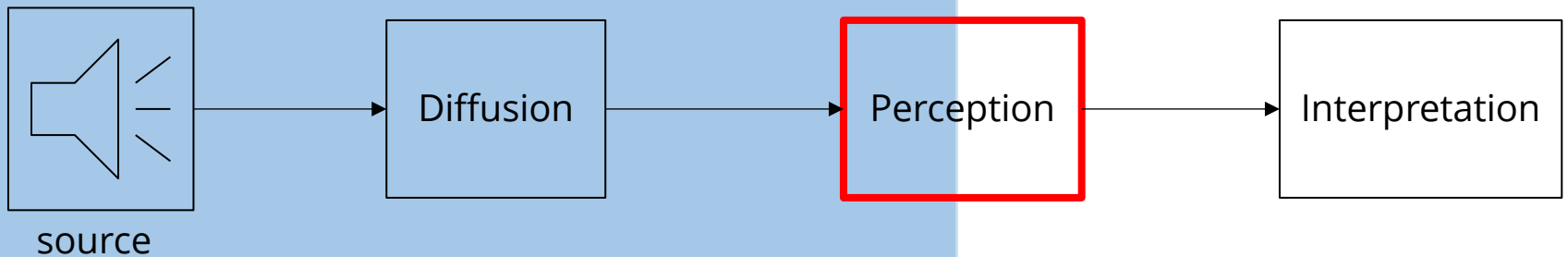
- Repeat sounds
- Model how the environment replicates/attenuates sound (LF vs HF) with room models
- Place sound sources on moving objects and with directionality



ARTIFICIAL WORLD - VR

Summing Up

- Use HRTF
- Use headphones
- Exploit the fact that those techniques allows the user to effectively (and precisely) localize the sound sources



ARTIFICIAL WORLD - VR