Spatial Audio L2

Matteo Luperto
Manuel Pezzera

matteo.luperto@unimi.it
manuel.pezzera@unimi.it

UNIVERSITÀ DEGLI STUDI DI MILANO

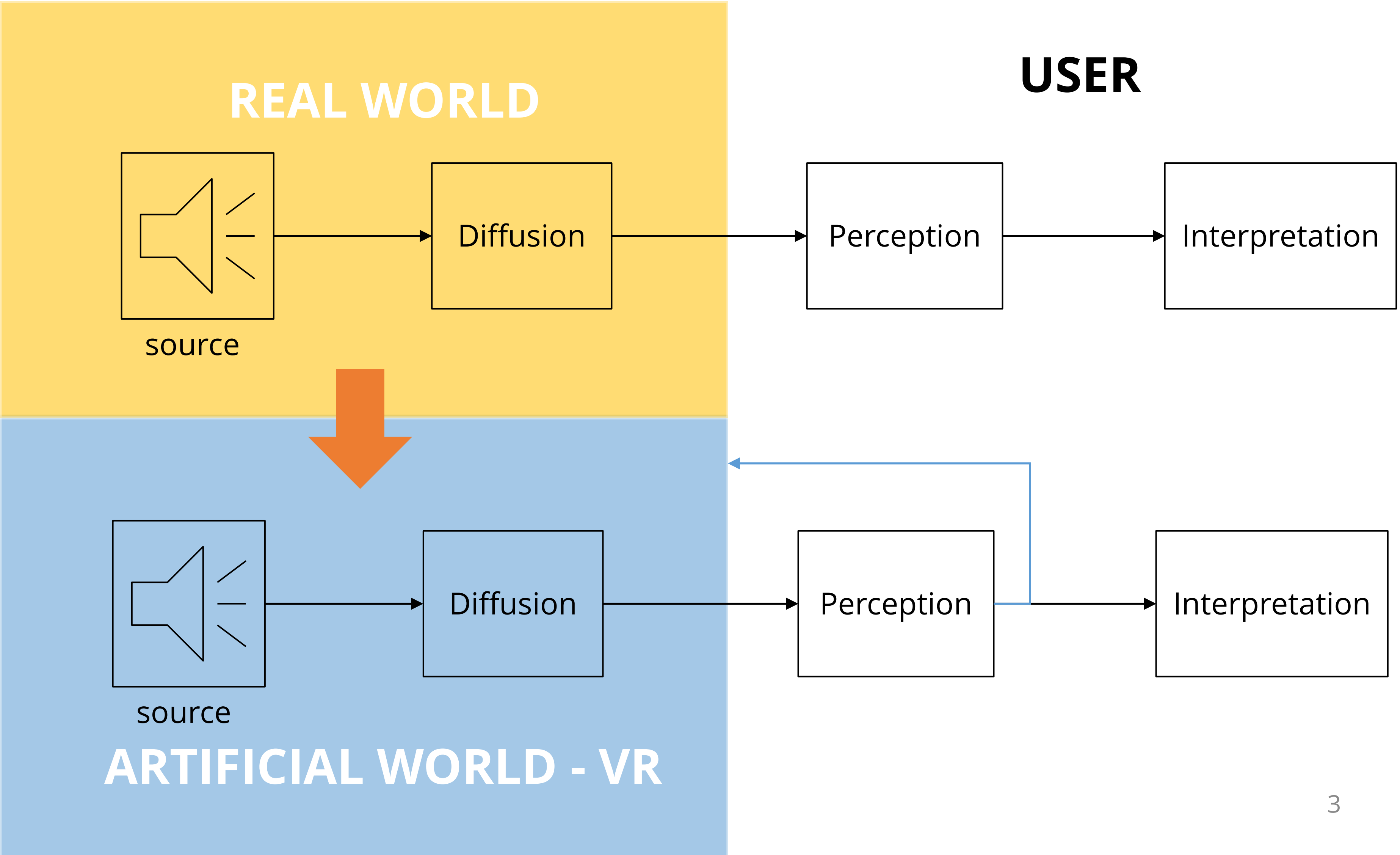
Lab 10

Realtà Virtuale 2020/2021

Today's outline

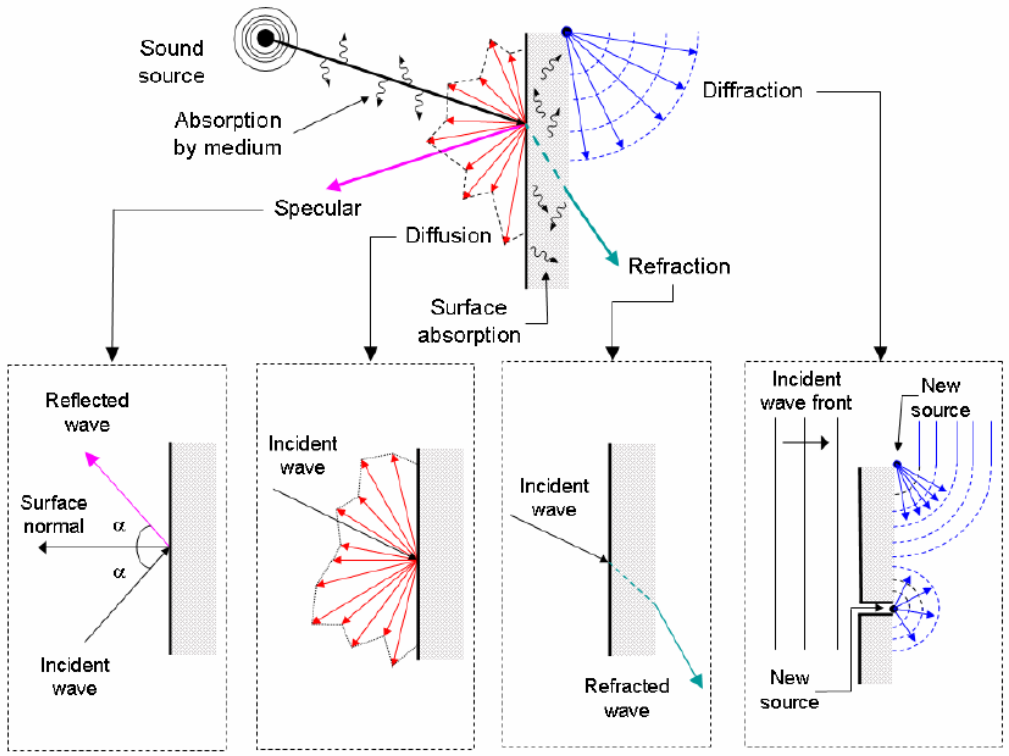
- Wrap up of Audio for VR
- SDK for spatial audio + design tips
- Unity and Spatial Audio

Audio modelling for VR



Audio modelling for VR: summing up

- Audio is emitted by a (non linear) source
- Audio waves are slow and are absorbed / reflected / refracted… by the media they encounter
- Eventually, we perceive several (distorted and delayed) replicas of the original sound



[Kapralos et al, 2008]

Audio modelling for VR: summing up

- High and Low Freqs are very different;
- LF are less subject than HF to different phenomena;
- This evenually results into the fact that we perceive disorted replicas.

speed (340 m/s)

wavelength

$$\lambda = \frac{v}{f}$$

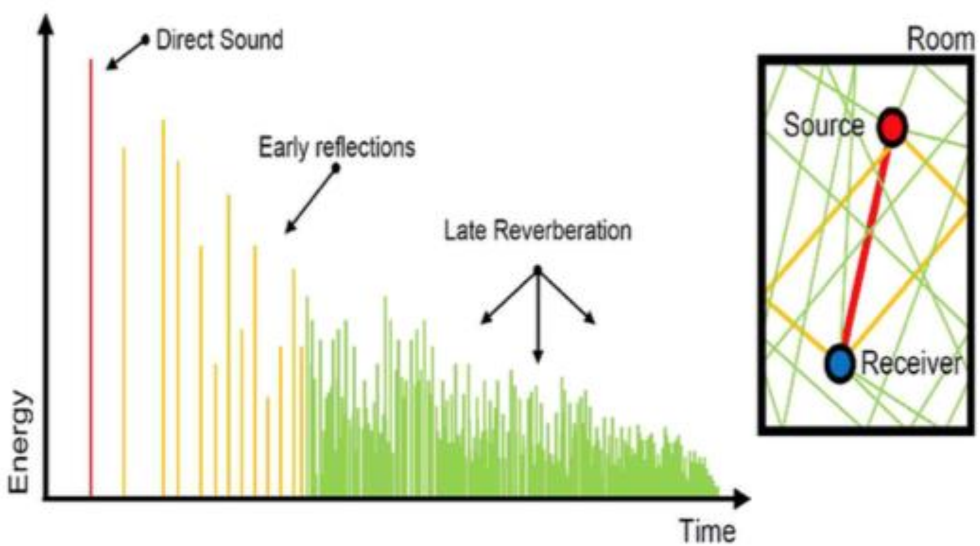
frequency

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

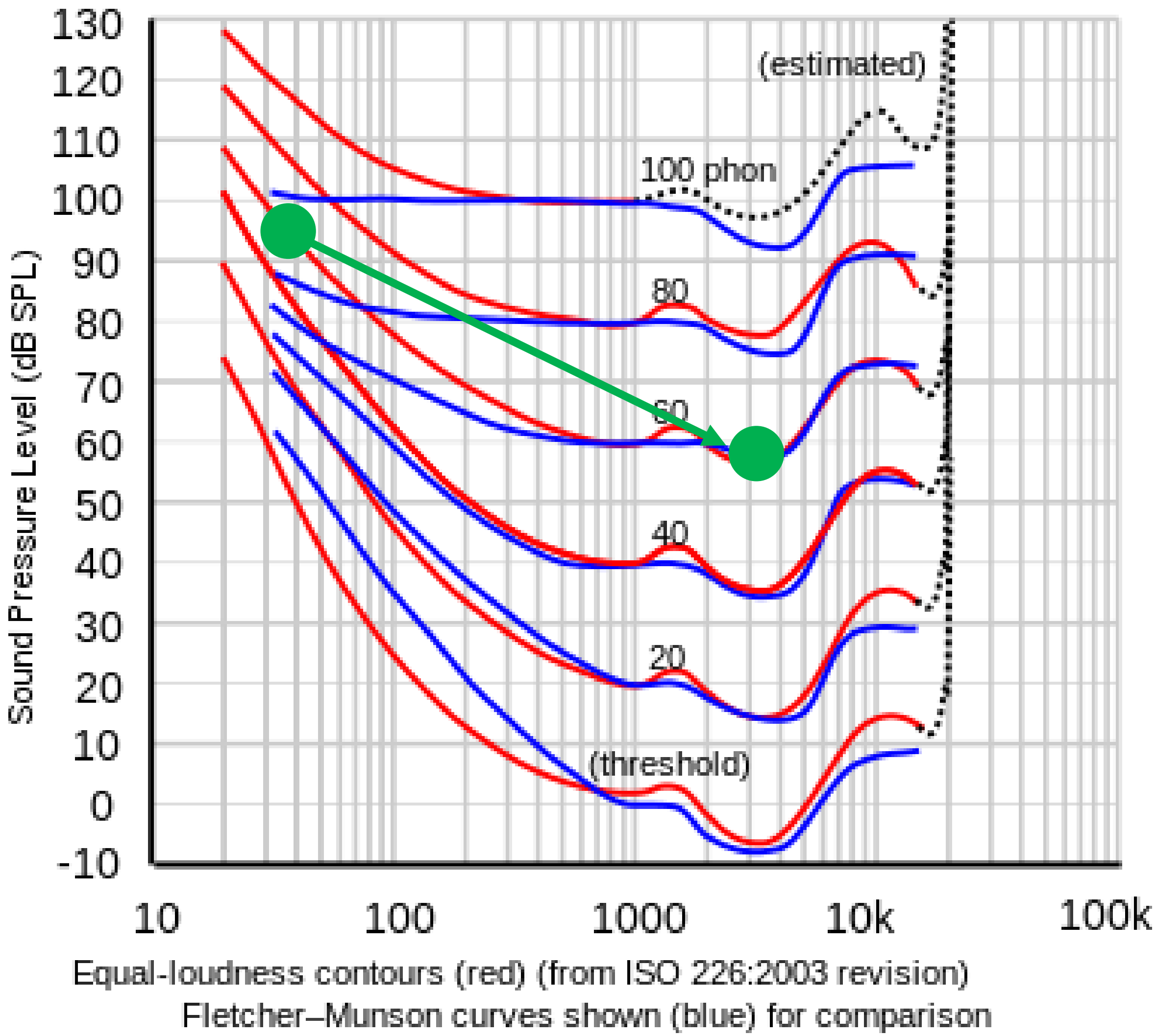
$$f = 25000 \ \texttt{Hz} \rightarrow \lambda = 13.6 \ \texttt{mm}$$

Audio modelling for VR: summing up

- High and Low Freqs are very different;
- LF are less subject than HF to different phenomena;
- This evenually results into the fact that we perceive disorted replicas.



Equal loudness contour curves

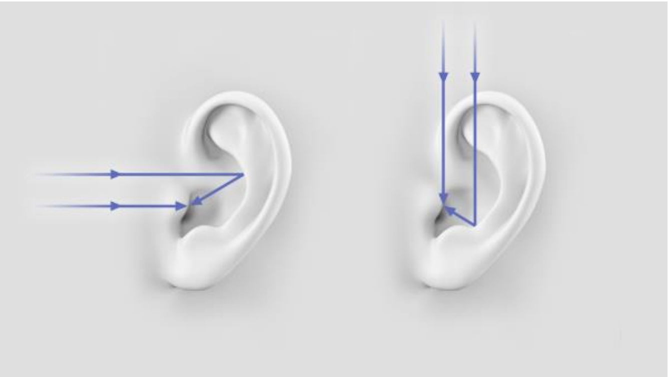


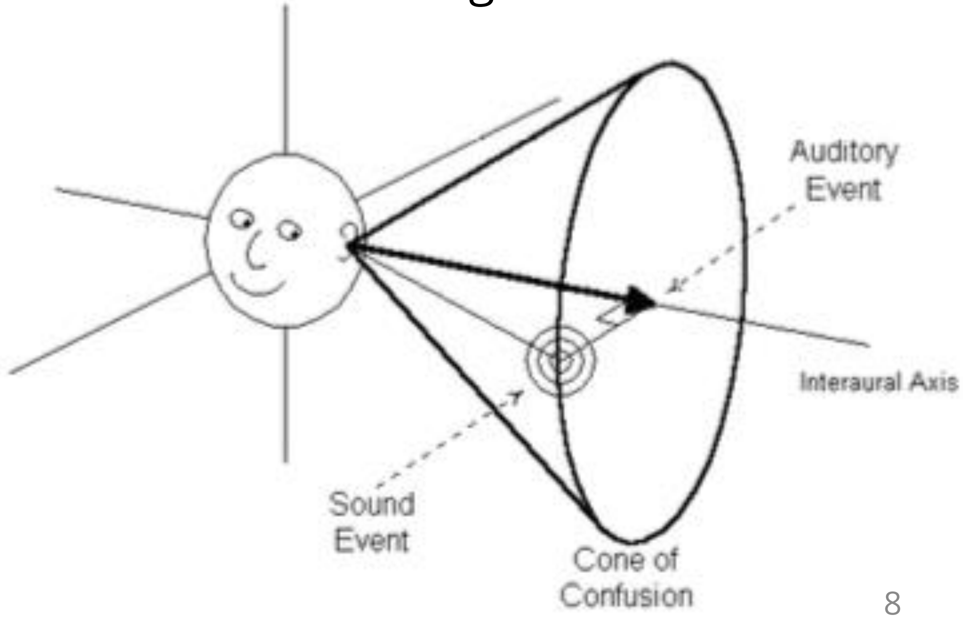
We also have a non-linear perception of different sound components.

If we *perceive* a LF and a HF at the same volume, the LF is actually at a much higher volume.

Audio perception for VR

- Our brain perceives all those replicas, filters them (we *hear* the reconstructed original sound) and uses those info to localize the sound location
- *Monaural Cues – Binaural Cues* - *Head motion*
- We have a *cone of confusion* for the sound source that we disambiguate with head movements
- Also echolocation is non-linear wrt HF and LF and angles



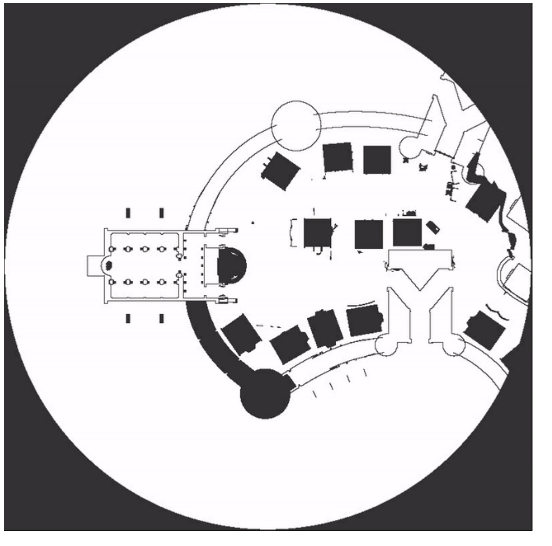


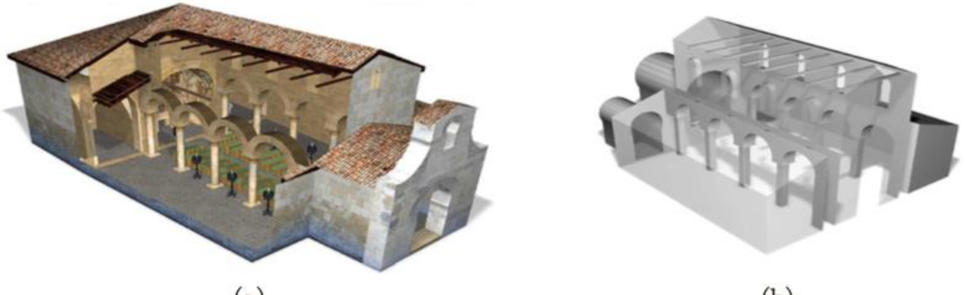
Audio perception for VR

In VR we have to replicate all those *cues* used to localize the sound so that our brain can uses them to localize virtual sounds

This is done by modelling:

- the effects due to the environment (for that we can consider only large objects and direct occlusion of the sound sources, e.g. room type and material)
- sound propagation (so how to create all those distorted replicas)
- sound perception (modelling distortion due to outer ear and micro-delays). For that we uses *Head-Related-Transfer-Functions* ***HRTF***





(a) (b)

https://www.microsoft.com/en-us/research/project/project-triton/

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

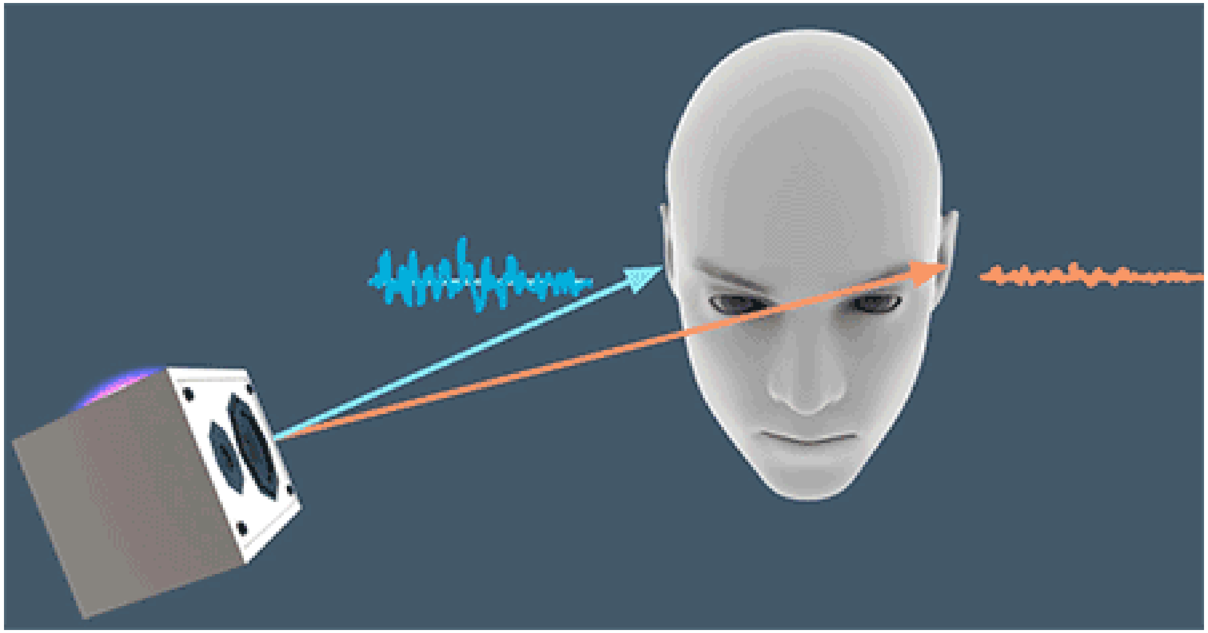
Resonance Audio SDK offers both audio propagation and perception.

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

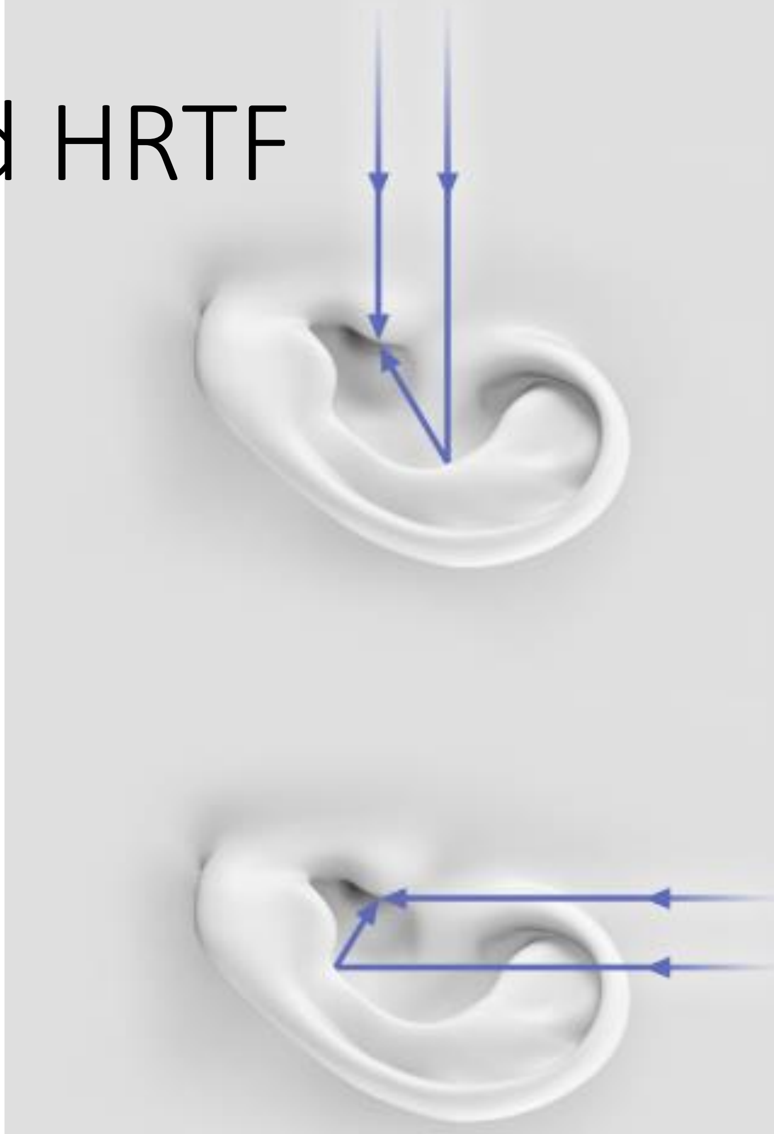
For higher frequencies, humans are unable to discern the time of arrival of sound waves. When a sound source lies to one side of the head, the ear on the opposite side lies within the head's acoustic shadow.
Above about 1.5 kHz, we mainly use level (volume) differences between our ears to tell which 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**

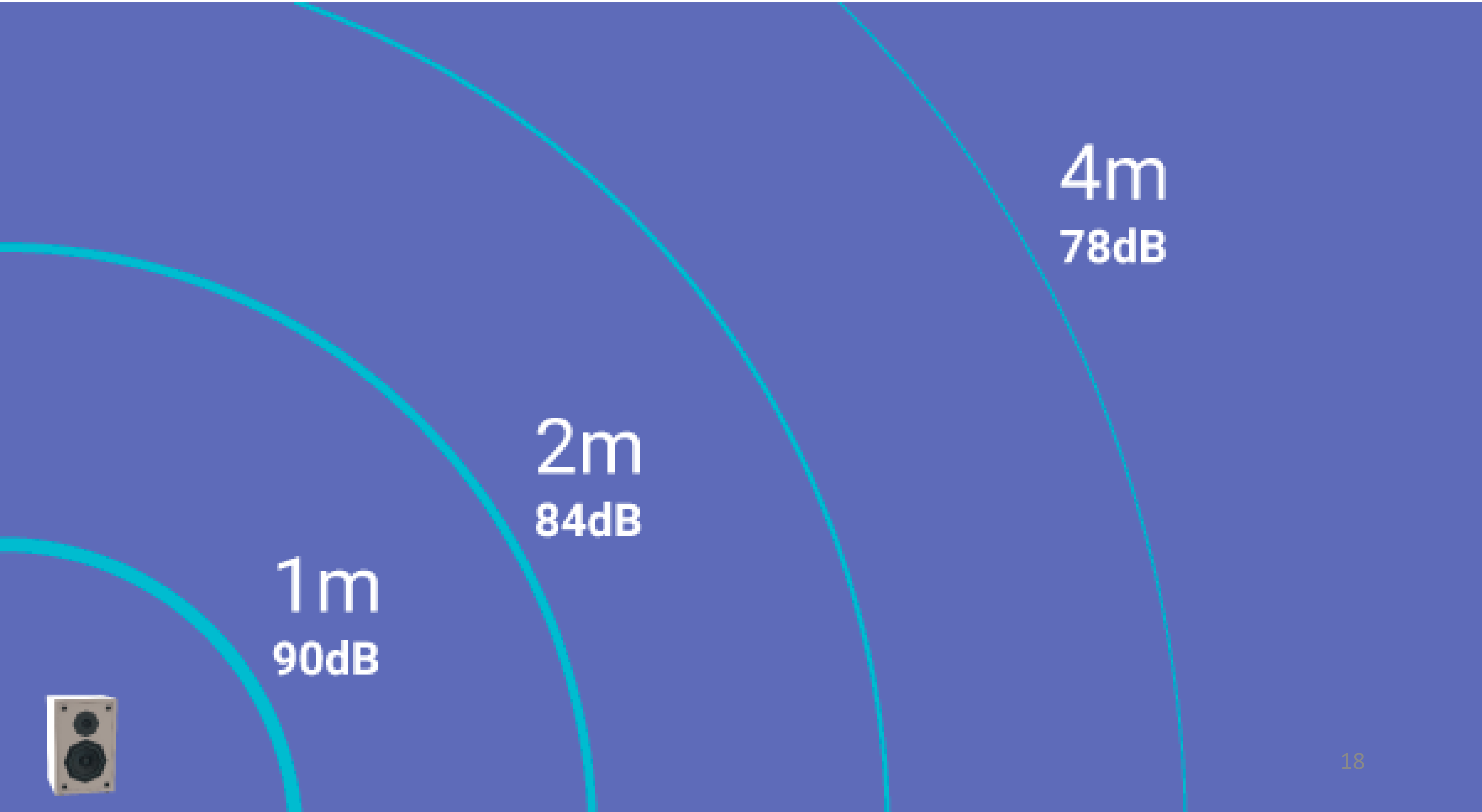
- Box-model – only size and material is specified for each material
- Reflection and scattering is simplified (real-time)

Occlusion blocks/transmits differently low and high frequency

Resonance Audio and Propagation

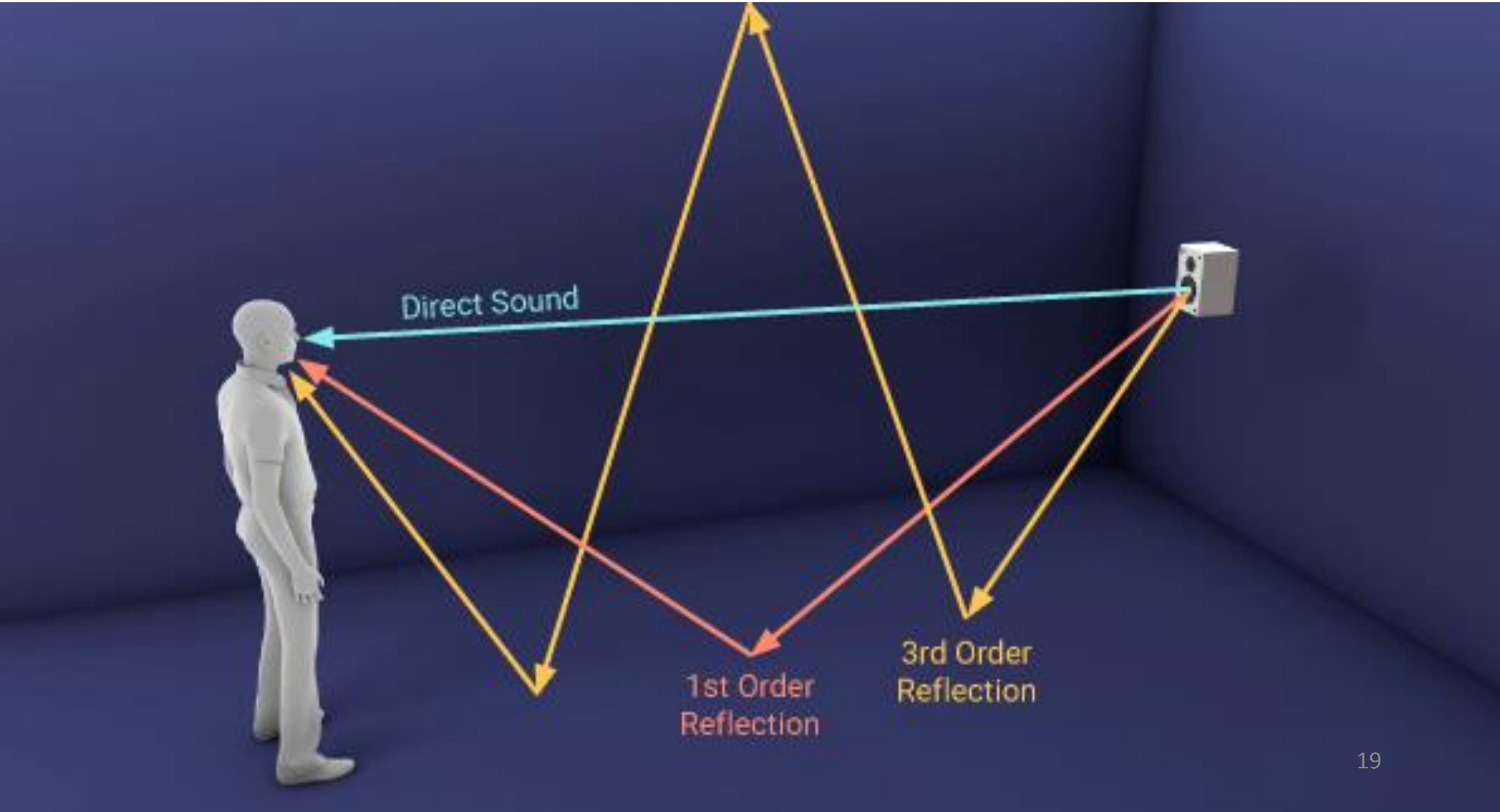
Propagation is modelled as three components

- <u>Direct sound</u>
- Early reflection
- Late reverb



Propagation is modelled as three components

- Direct sound
- <u>Early reflection</u>
- Late reverb



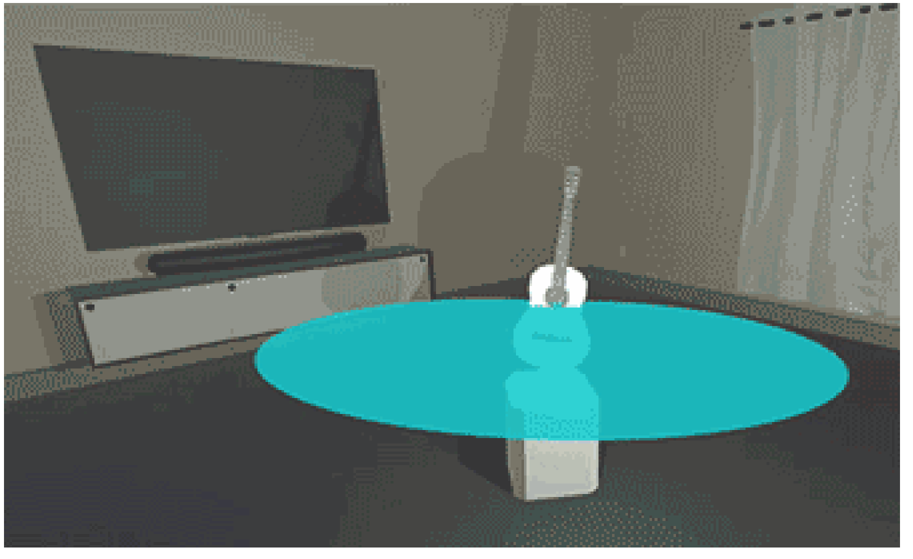
Propagation is modelled as three components

- 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 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 alone and close to the 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

Ambisonic files should 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-handy-recorder

Design tips

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

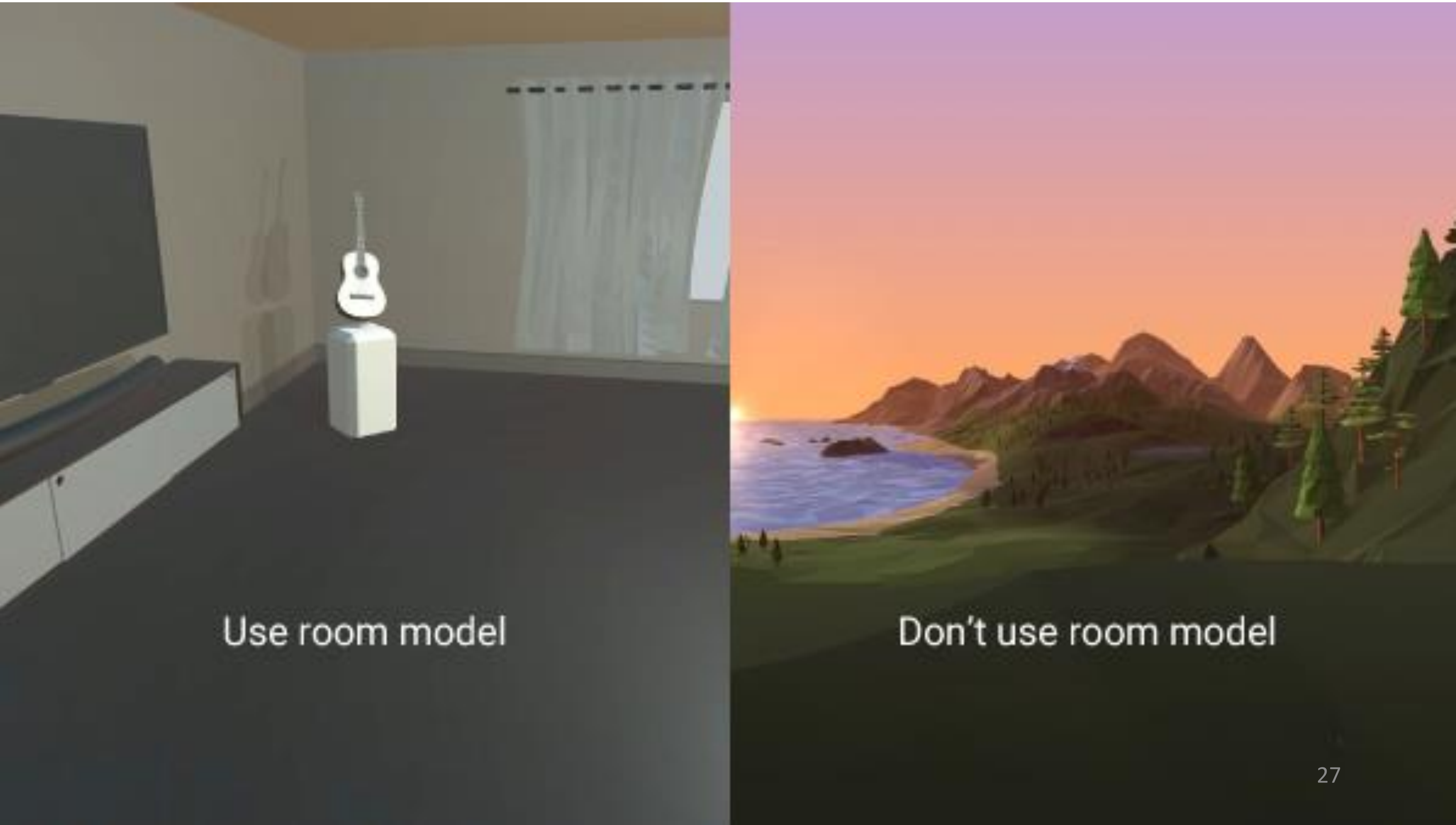
Normal stereo/mono sounds can be used for standard audio sources (e.g. a guitar, television, speech).

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

ZOOM H3-VR (<400 euro)
(Zoom H2n is approx 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 carefully Resonance Audio Sources

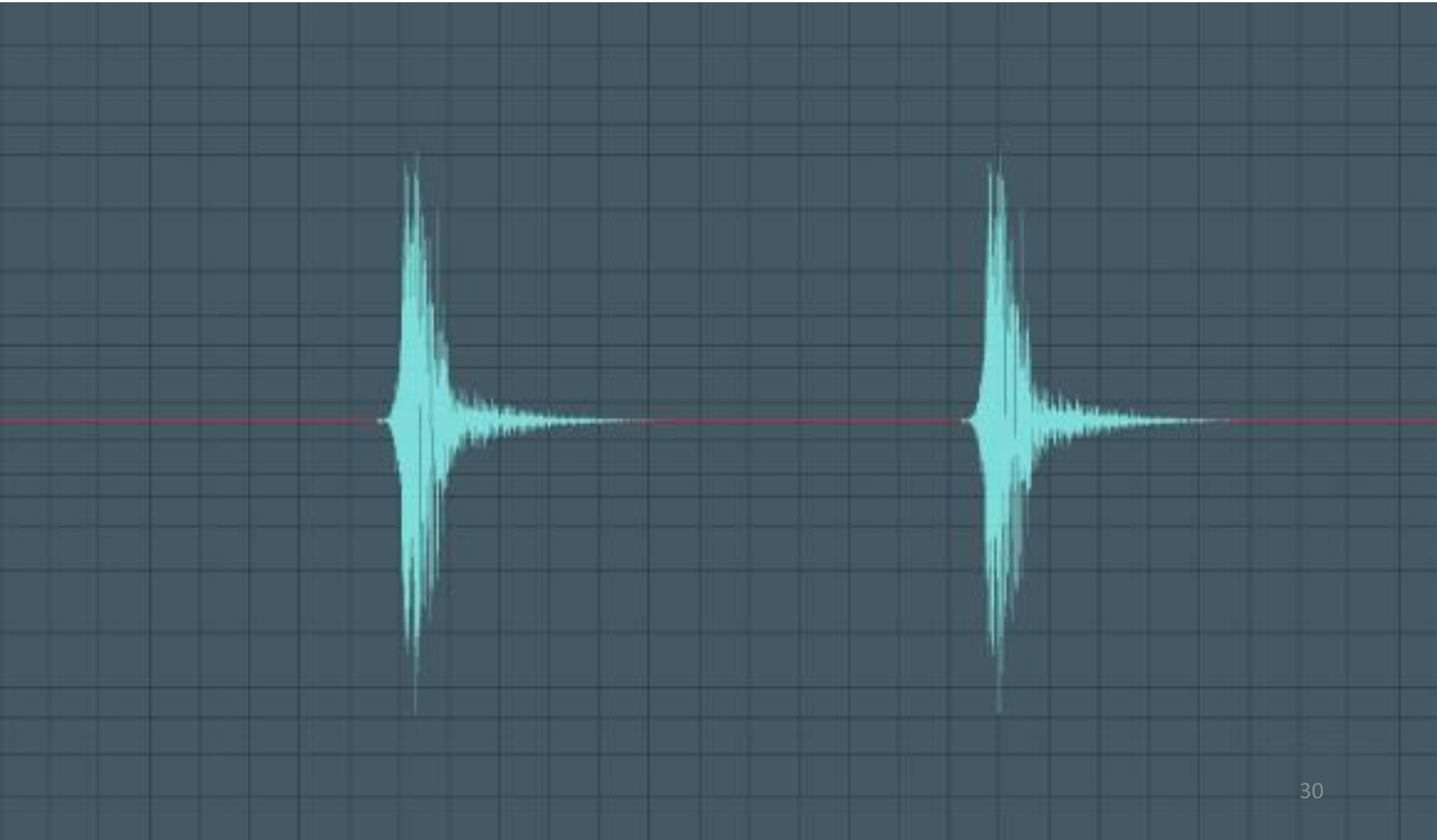
– mobile audio source moving around the listener can be very useful

– 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

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 is using 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

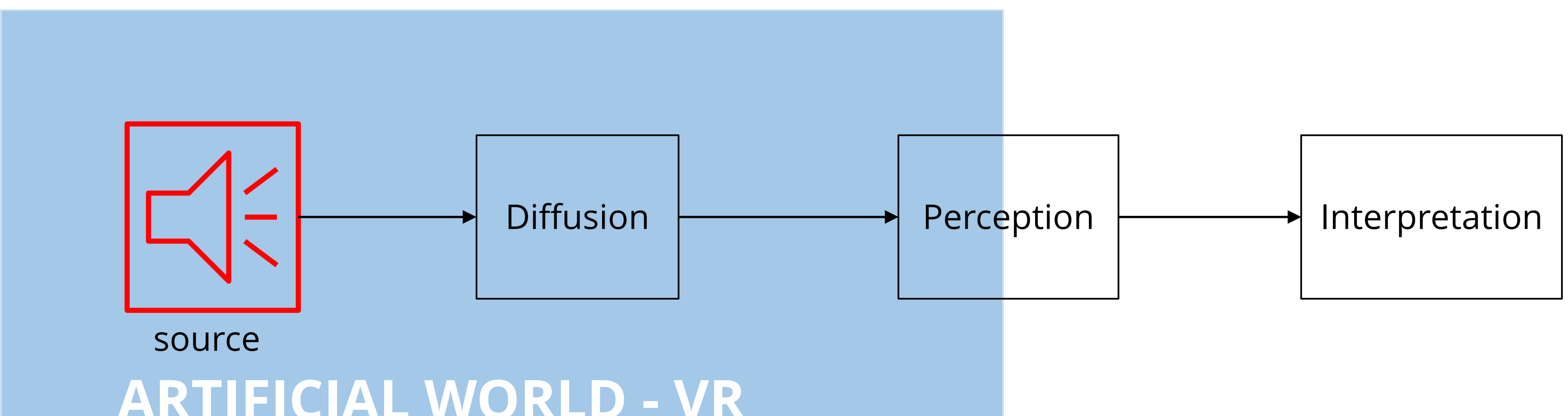
Demo – SDK for Web

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

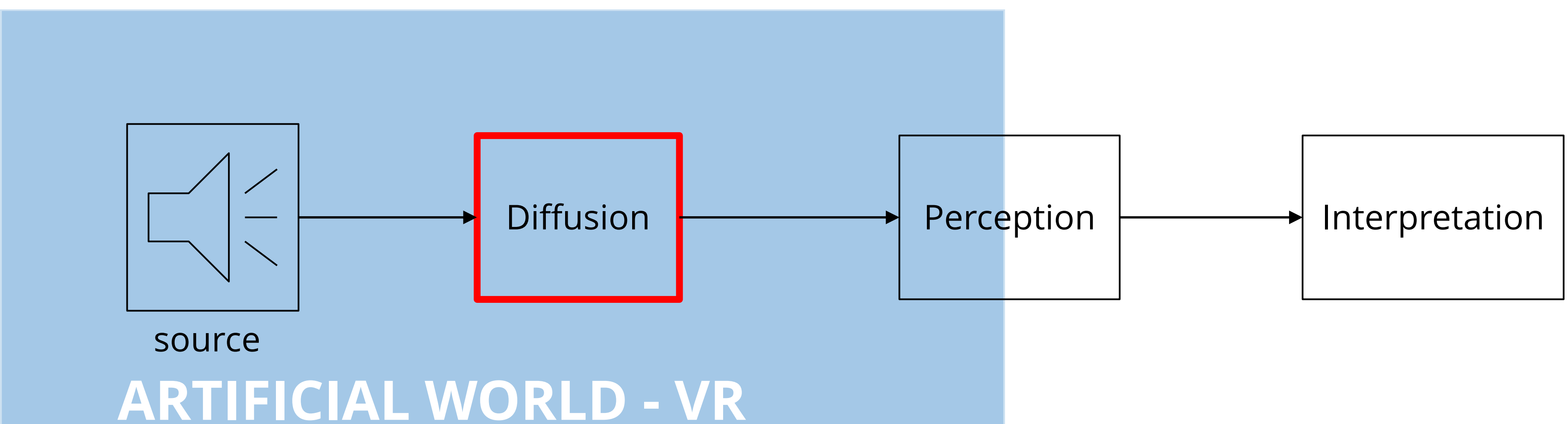
Summing Up

- 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)



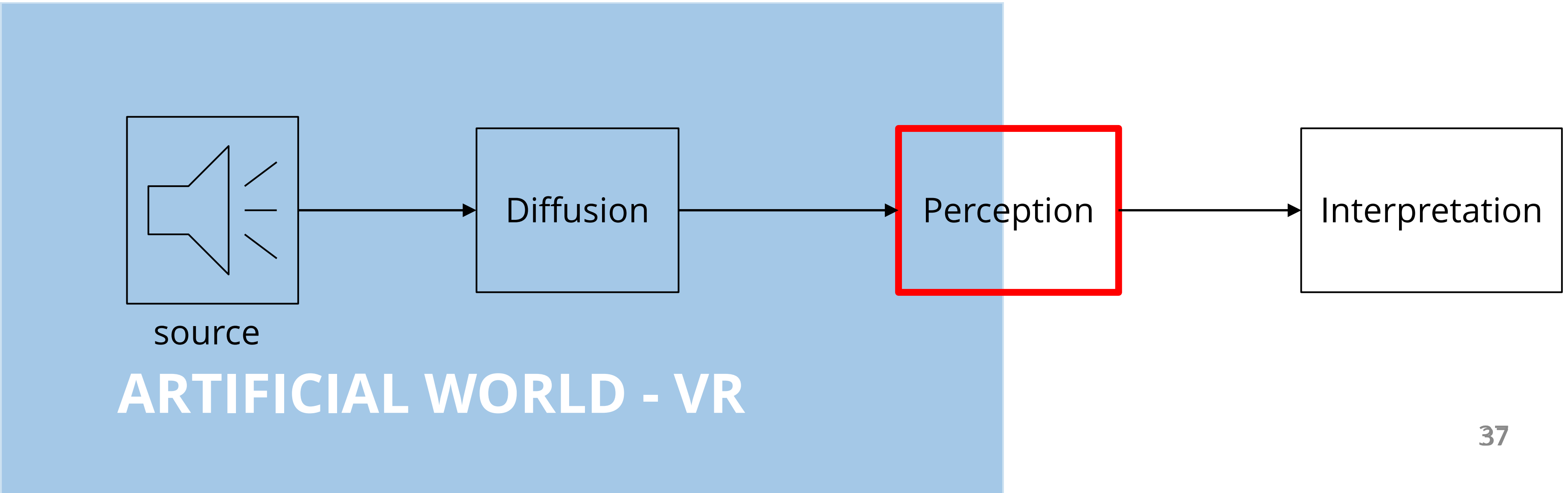
Summing Up

- Repeat sounds
- Model how the environment replicates/attenuates sound (LF vs HF) with room models
- Place sound sources on moving objects and with directionality
- Model stuff that changes the sound (e.g., occlusion)



Summing Up

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



References

Steven La Valle, Virtual Reality, Cambridge Press

http://lavalle.pl/vr/

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