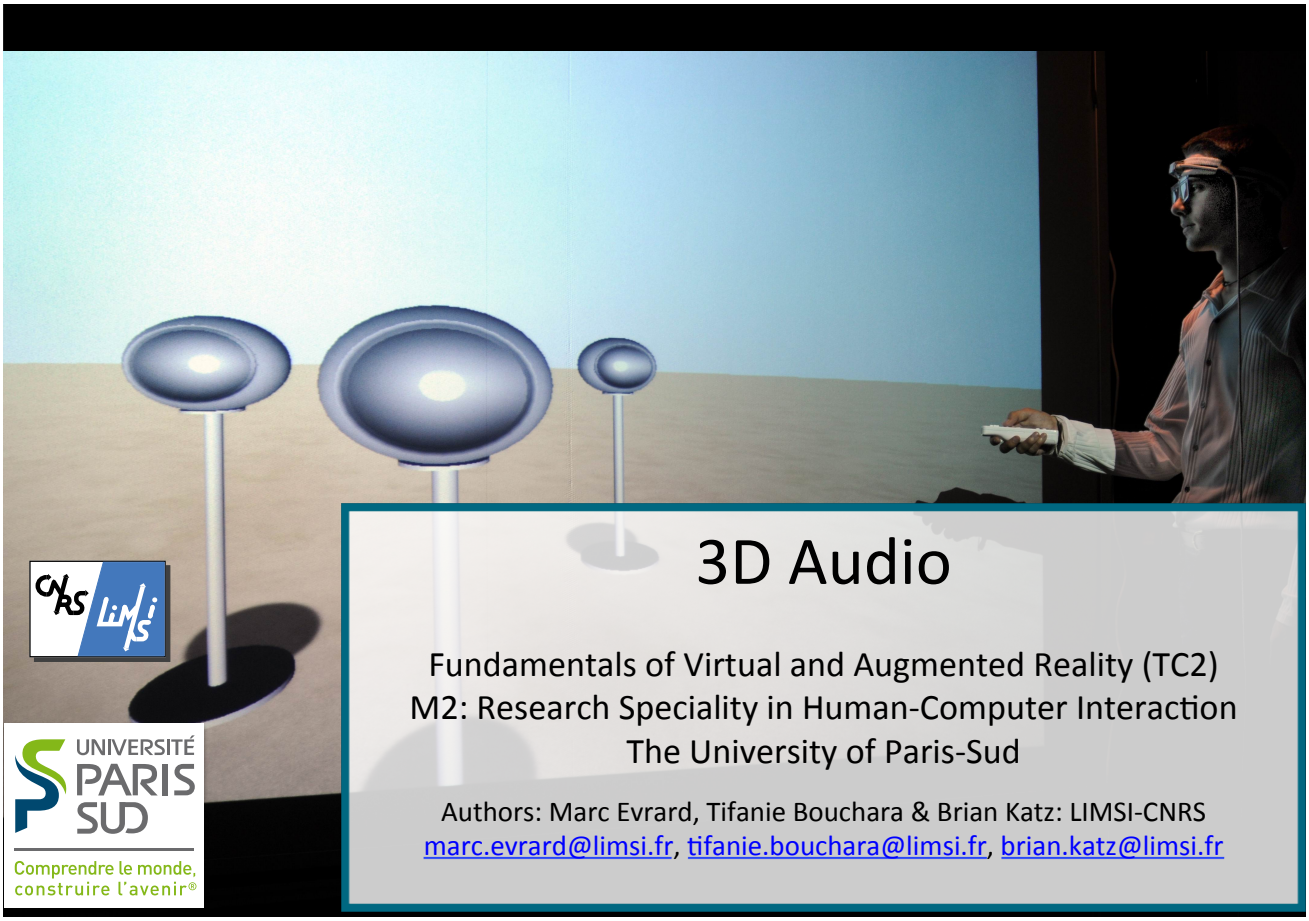


3D Audio

Fundamentals of Virtual and Augmented Reality (TC2)
M2: Research Speciality in Human-Computer Interaction
The University of Paris-Sud

Authors: Marc Evrard, Tifanie Bouchara & Brian Katz: LIMSI-CNRS
marc.evrard@limsi.fr, tifanie.bouchara@limsi.fr, brian.katz@limsi.fr

Introduction

Brainstorming

What is the role of sound in everyday conditions?

=> Role of sound in interfaces?

Introduction

⇒ Immersion (presence)
⇒ Environment (information about the location)
⇒ Spatial information beyond visual
⇒ Aesthetics / art
⇒ Communication (e.g. speech)
⇒ Prevention (e.g. alarm)
⇒ Bringing additional information through sound properties

Different types of sound:
Music / speech / environmental sounds

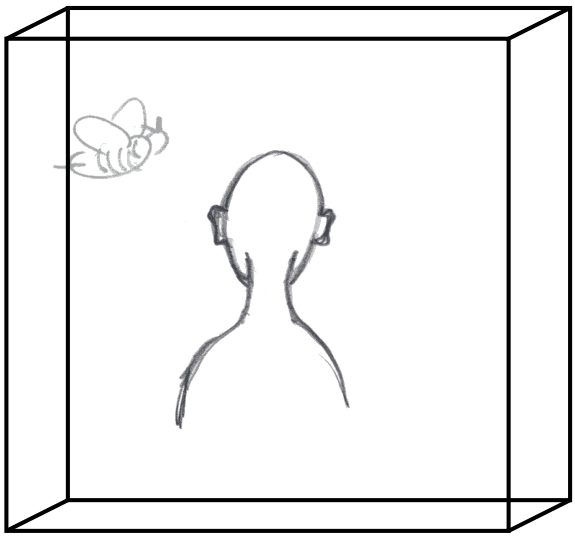
Introduction

Audio in immersive systems → notion of space

Sound objects
- Audio content
- Spatial location
- Spatial behavior

Environment
- Geometry
- Acoustics

Listener, **listening** and **reproduction**
- Listening conditions
- Spatial position



Outline

1. Sound and acoustics
2. Hearing (psychoacoustic)
3. Room acoustics
4. Sound interfaces for virtual/augmented reality (VR/AR)
5. 3D audio techniques
6. Sonification and auditory displays

Sound and acoustics

Sound is a vibration but:
- "Sound" is used to describe audible vibrations in air
- "Vibration" is used to describe tactile vibrations (touch)

Vibration is a periodic phenomenon

- Sound propagates through a material, made of particles
- Particles must be able to move – vibrate – oscillate
- Particles do not flow, only vibration does – energy flows
- Propagating energy of sound is called the acoustic wave

Acoustics is an interdisciplinary science:
- study of mechanical waves in gases, liquids and solids
- Includes: vibration, sound, ultrasound and infrasound

(Wikipedia. For more info see Beranek, *Acoustics*, 1993 edition)

Physical and perceptual parameters

Physical parameters of the sound wave:
- Frequency (1/period, velocity/wavelength)
- Amplitude
- Duration

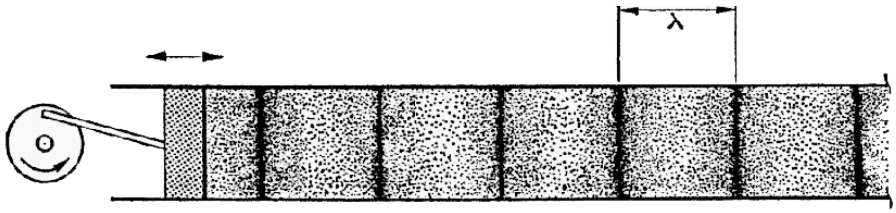
Perceptual parameters of the sound wave:
- Pitch
- Loudness (dB logarithmic scale)
- Length, rhythm

Wave motion and sound speed

Vibration in air translates to pressure changes

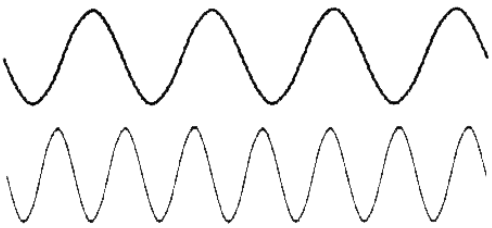
Frequency: $f = \frac{c}{\lambda} \quad \frac{\left[m/s\right]}{\left[m\right]} = \frac{\left[Cycles\right]}{\left[s\right]} = \left[Hz\right]$

Length of 1 cycle = wavelength (λ)



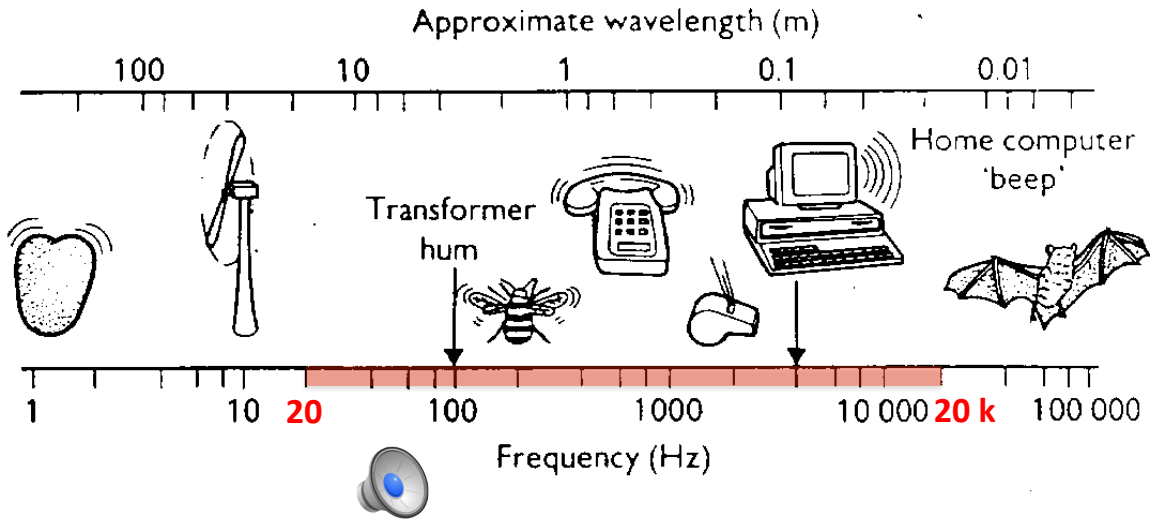
500Hz

1 kHz



Frequency

Frequency is related to **pitch** sensation
Frequency is often shown using the **linear unit** (Hz)
on a **logarithmic scale**



Sound pressure level of typical sounds

Loudness and dynamic range are expressed in a logarithmic unit:
The decibel (dB)
$10\times$ the power = 1B (bel) = 10dB ≈ "Twice the loudness"
Decibel is always a **level ratio** between two **power** values
If the ratio is calculated according to a fixed reference,
decibels then describes an **absolute level**

Ex.: Sound Pressure Level (SPL)

$$dB_{SPL} = 20 \times \log_{10}\left(p / p_{ref}\right), \quad p_{ref} = 2\times10^{-5}\,Pa$$ (Pressure level reference)

Why 20? $B_{SIL} = \log_{10}\left(I / I_{ref}\right)$ $dB_{SIL} = 10 \times \log_{10}\left(I / I_{ref}\right)$

$$I \propto p^2 \qquad dB_{SPL} = 10 \times \log_{10}\left(p^2 / p_{ref}^2\right)$$

Sound pressure level of typical sounds

Permissible noise exposure

Duration per day	Loudness
8 hours	90 dB_{SPL}
4 hours	95 dB_{SPL}
2 hours	100 dB_{SPL}
1 hours	105 dB_{SPL}
30 min	110 dB_{SPL}
15 min	115 dB_{SPL}
7 min	120 dB_{SPL}

Recommended by the OSHA (US)
(Occupational Safety & Health Administration)

Ex. of typical SPL values
(expressed in A-weighted dB scale)

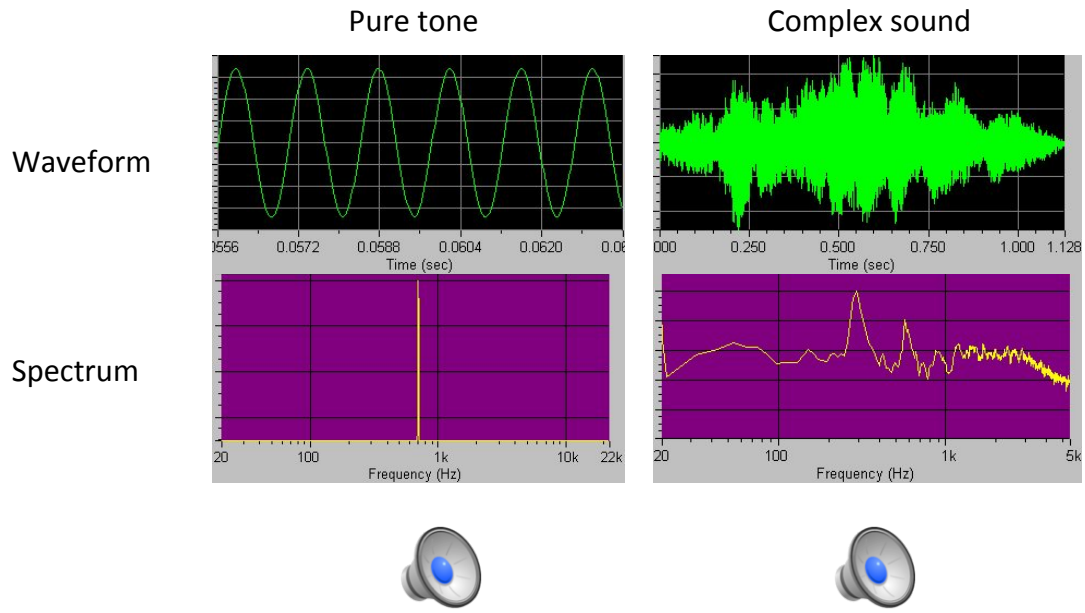
Power @1kHz	Pressure @1kHz	Loudness	Sound
10^{8}	200 kPa	200 dB_A	≈ 1 atmosphere
10^{2}	200 Pa	140 dB_A	Threshold of pain
10^{0}	20 Pa	120 dB_A	Jet takeoff @100m
10^{-2}	2 Pa	100 dB_A	Disco/nightclub
10^{-4}	200 mPa	80 dB_A	Vacuum cleaner @3m
10^{-6}	20 mPa	60 dB_A	Conversation @1m
10^{-8}	2 mPa	40 dB_A	Quiet private office
10^{-10}	200 µPa	20 dB_A	Soft whisper @2m
10^{-12}	20 µPa	0 dB_A	Threshold of hearing

Wave motion and speed of sound

The speed of the acoustic wave depends on the conductive material

Material	*Approx. Speed of Sound*	
	(ft/sec)	*(m/sec)*
Air	1 130	345
Lead	4 000	1 220
Water	4 630	1 410
Brick	9 850	3 000
Concrete	11 150	3 400
Wood	11 150	3 400
Glass	13 450	4 100
Aluminum	16 730	5 100
Steel	17 060	5 200

Sound representation



Typical sound

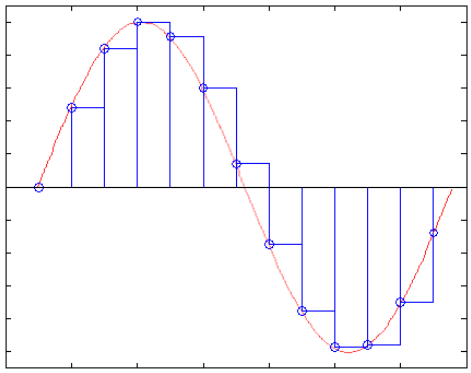
- Pure tone
- Sweep
- Burst
- Speech
- White noise
- Pink noise
- Red noise (brown)
- Grey noise

From sound to digital audio

Sampling: isolation of the audio signal

The **sampling frequency** (f_s): the **highest** possible **frequency** (f_{max}) of the signal

Shannon-Nyquist theorem: $f_{max} = f_{smpl}/2$



The **Bit depth** also set the quality, ex. for a CD:

- f_s = 44'100 Hz
- f_{max} = 22'050 Hz
- Bit depth = 16 bits => SNR ≈ 16*6 = 96 dB (*Signal to Noise Ratio*)
- Bit Rate = 16*44'100 = 706 kbps
  ≈ 5 MB/min (mono)

Sampling frequency and bit depth set the audio system performance

Better quality => more processing

Voice 44 kHz 16bit
Voice 8 kHz 8bit
Voice 22 kHz 16bit
Music 44 kHz 16bit
Voice 8 kHz 16bit
Music 8 kHz 8bit

Digital audio

Record and reproduce

Synthesis

Pure synthesizing

Sampling (synthesizing with recorded samples)

Transformation

Frequency domain filtering

Temporal domain filtering

The sound object

The sound object

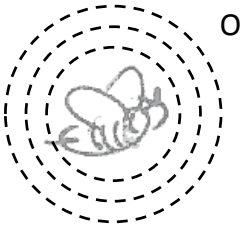
Audio content
- Pre-recorded file
- Synthetic parameters

(Only one channel)

Spatial location
- Geometrical location in the scene
- Orientation

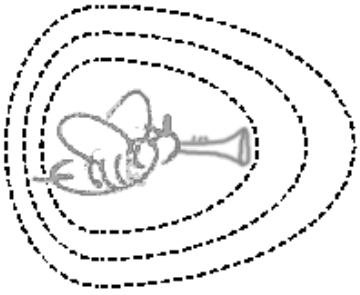
Spatial behavior
- Directivity (may depend on frequency)
- Speed

Variations due to spectral / temporal changes
- Pitch, timbre
- Echo
- …



Omnidirectional

Directional



The listener

Hearing

Hearing is a **multidimensional** process

- Time
  Rhythm, tempo, variation
- Frequency
  Pitch, harmony, modulation

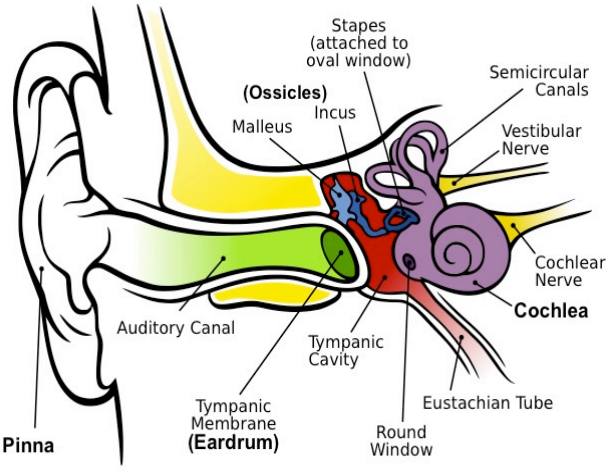
Perception depends on the individual

- Space
  Distance, angle, elevation
  In all directions
  (Left / Right, Front / Back, Top / Below)
- Segregation (separation) of several sources

Space	Frequency
Onset time	Timbre …

Ears

The auditory system captures acoustical vibrations and brings the information to the brain



The ear is composed of 3 parts:

Outer ear (Pinna + auditory canal):
Collects sound – amplifies acoustic vibrations

Middle ear (Eardrum and ossicles):
transforms air vibrations into liquid vibrations (lymph fluid) => impedance matching

Inner ear (Cochlea):
Hair cells in the cochlea transform mechanical vibrations of the lymph liquid into an electric signal that will be transmitted to the brain by auditory nerve

Ears: more comments

In the cochlea, signals are processed by frequency bands (**tonotopic coding**) before being combined into sound objects later in the brain

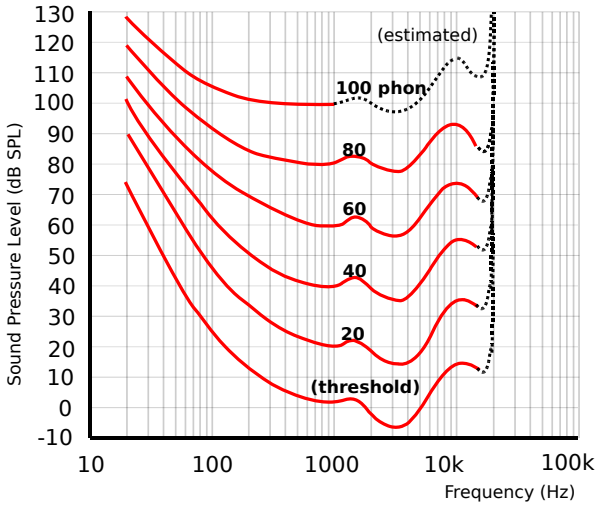
Some nerve cells compare signals from both ears:
binaural processing

Ears are sensitive to:
average sound pressure over a 0.5 to 1s interval
frequency via the cochlea

Ear response is **logarithmic**

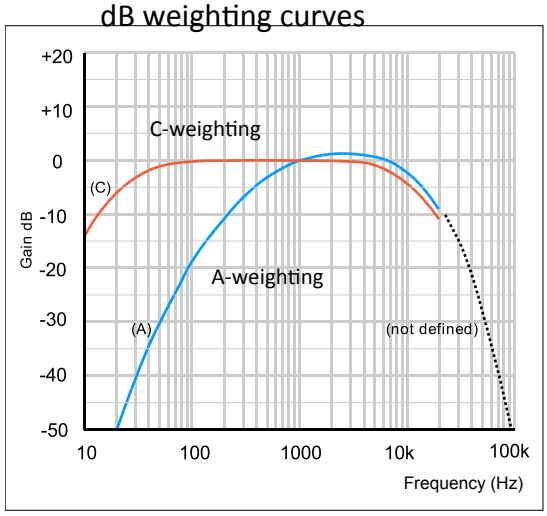
Ears: sensitivity

Isosonic curves



Equal-loudness contours (IOS 226:2003 revision)

dB weighting curves

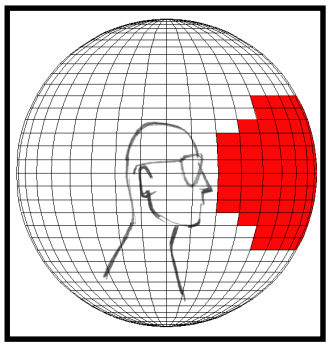


Greater sensitivity for upper-mid frequencies: between 2 kHz – 6 kHz

Less sensitivity for low frequencies

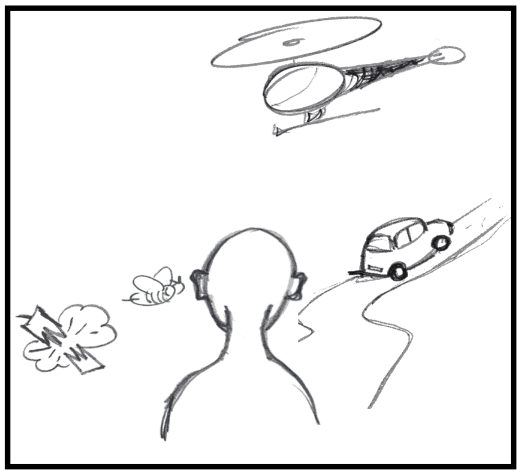
The higher the level, the more homogeneous the sensitivity

Spatial hearing



Passive listening (automatic and unconscious):

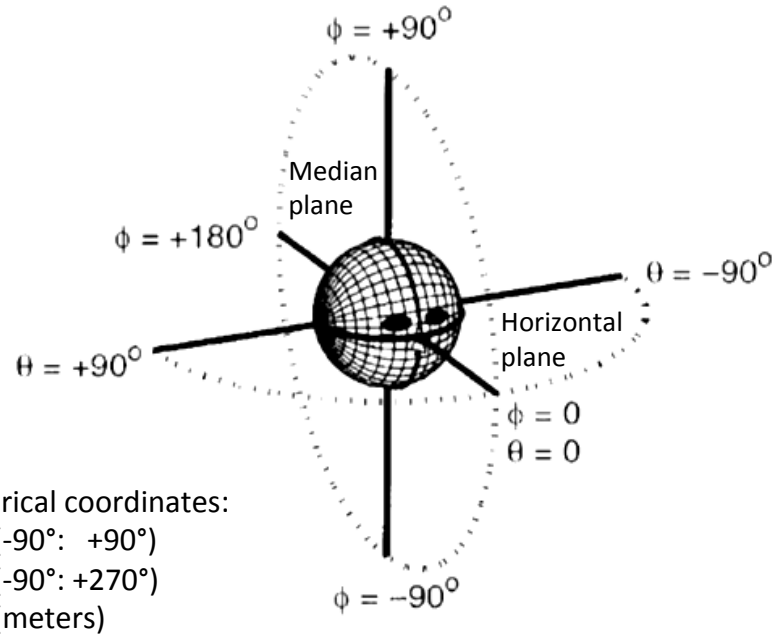
- Accuracy
  - ±5° on average
  - Perception of distance
  - Perception in 360° in elevation and azimuth (only ≈120°x70° in vision)
- Examples of use:
  - Object discrimination
  - Speaker / other voice (noise) discrimination
  - Environmental organization
  - Awareness and focus changes



Active listening (requires focus):

- ±1° accuracy possible in front of listener
- Examples of use:
  - Threat localization and identification
  - Machine analysis and diagnosis

Spatial hearing



Interaural-spherical coordinates:

θ *Azimuth* (-90°: +90°)

Φ *Elevation* (-90°: +270°)

r *Distance* (meters)

Spatial hearing: localization cues

Localization cues:
direction and distance wise

1) Direction wise (angle)

- **ITD**: Interaural Time Differences (mainly for lower frequencies)
- **IID**: Interaural Intensity Differences (mainly for higher frequencies)
- **DDF**: Direction Dependent Filtering

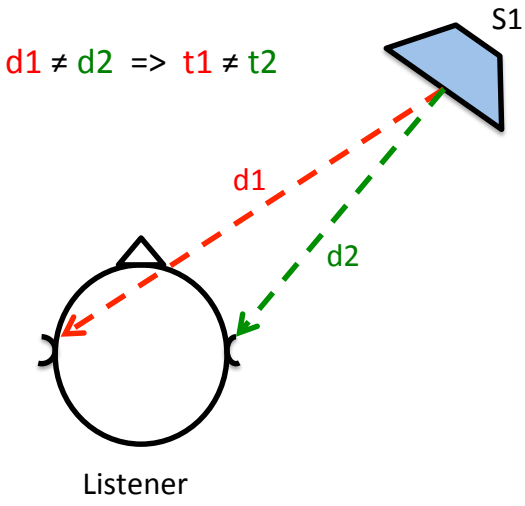
Spatial hearing: ITD

ITD = **Interaural Time Difference**

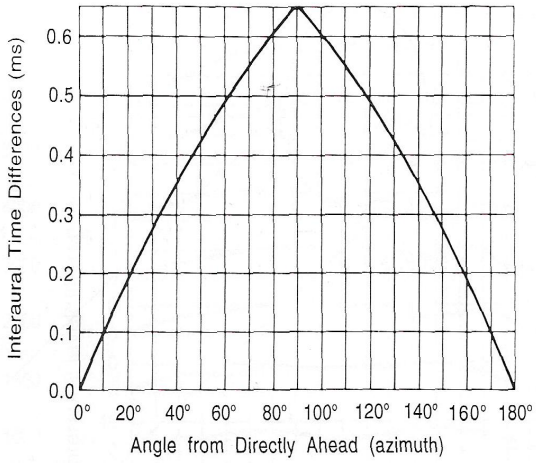
Difference in the arrival time of a signal at the two ears

Does **not** dependent on frequency,

but limited to a $f_{max} \approx 1.5$ kHz (spatial aliasing)



ITD as a function of azimuth

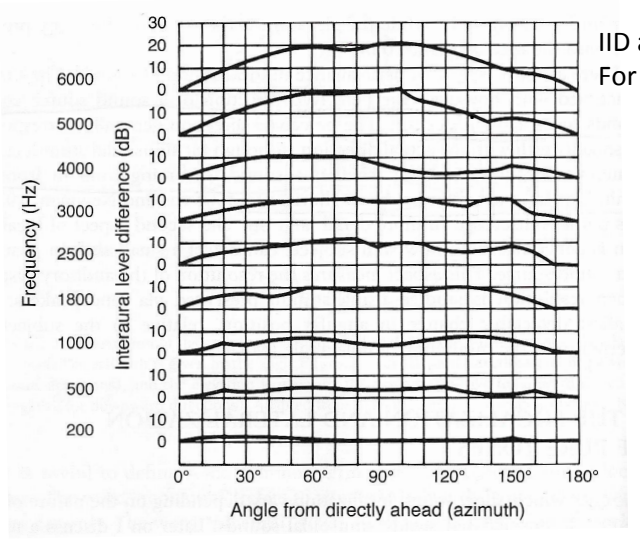


Spatial hearing: IID

IID = **Interaural Intensity Difference**

Intensity difference between the two ears

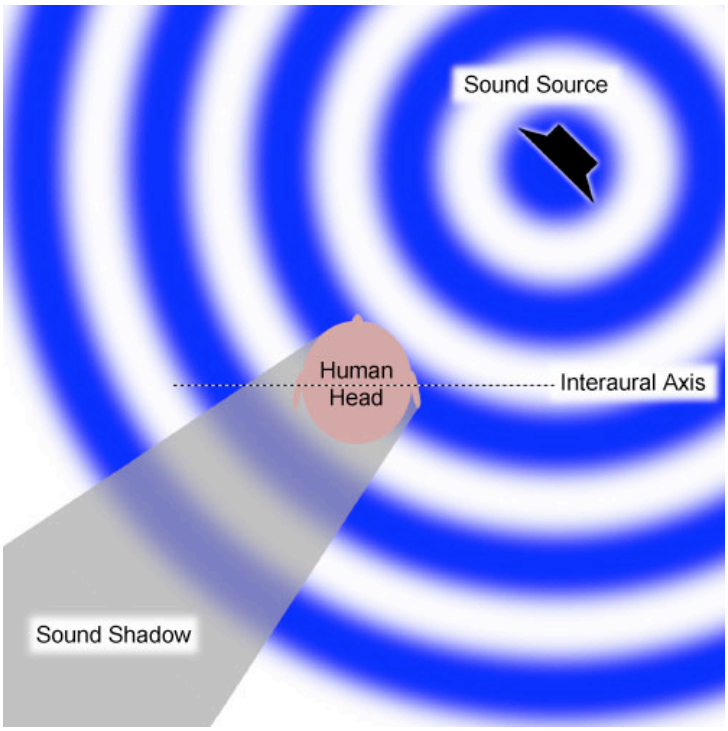
Frequency dependent parameter



IID as a function of the azimuth
For various frequencies

[*Feddersen*, 1957]

Spatial hearing: IID

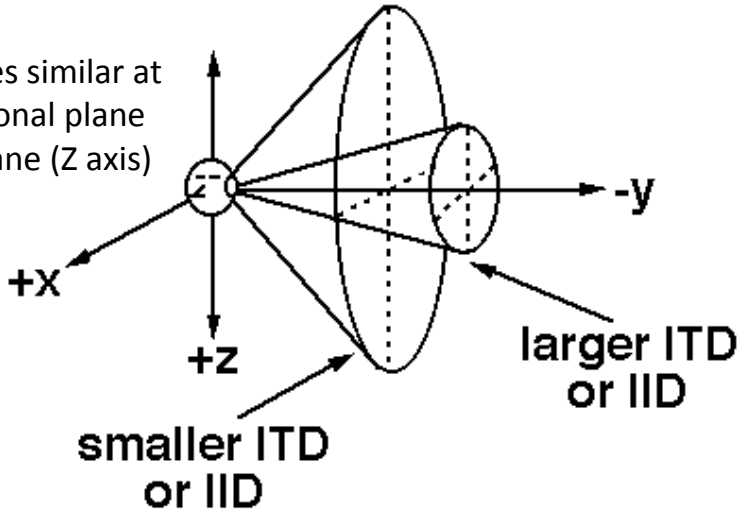


Cone of confusion & inversion

Cone of confusion

Time and intensity differences similar at opposite locations in the coronal plane (X axis) and in the vertical plane (Z axis)

Front / Back inversion



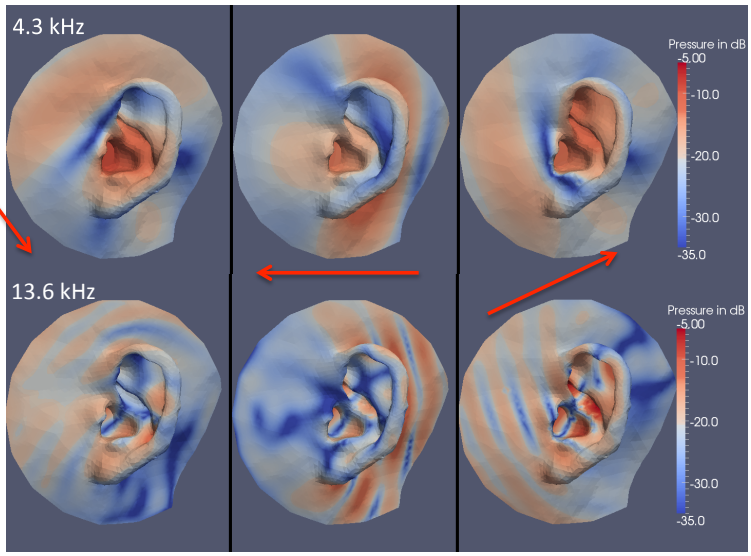
The cone of confusion is resolved by filtering achieved by the outer ears and head movements

Spatial hearing: DDF

DDF = **Direction Dependent Filtering**

Filtering performed by the external ear, the head and the shoulders

Depends on the sound direction of incidence

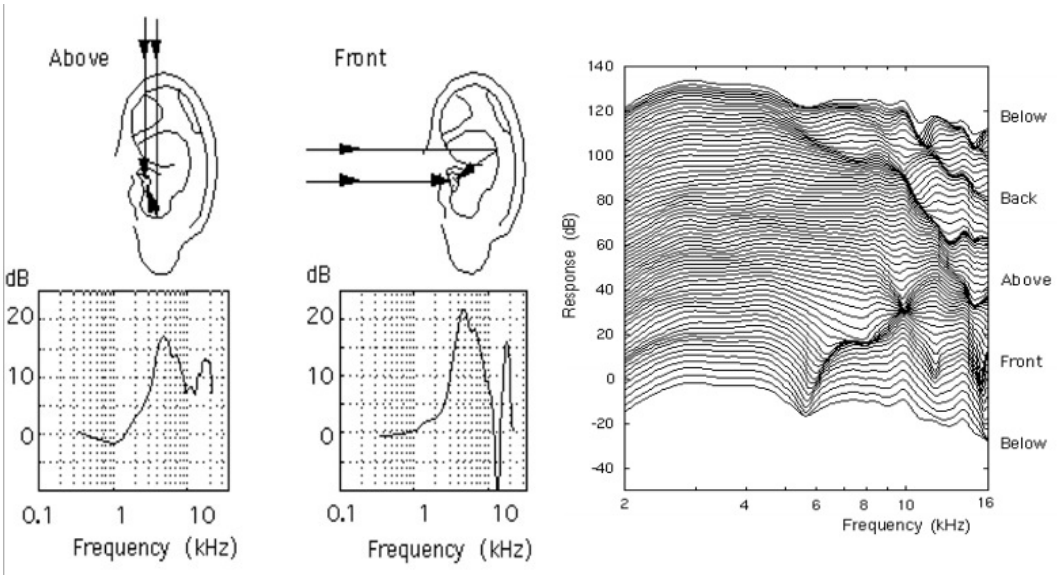


[Makoto 2010]

Spatial hearing: DDF

Simplified model for the very high frequencies: pure reflections and sum

In practice: diffusion and diffraction occur



Spatial hearing: HRTF

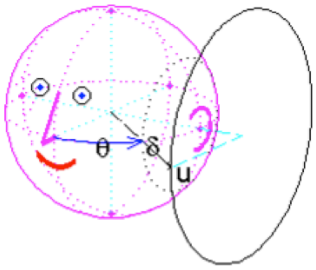
Head-Related Transfer Function (HRTF)

Transfer function describing:

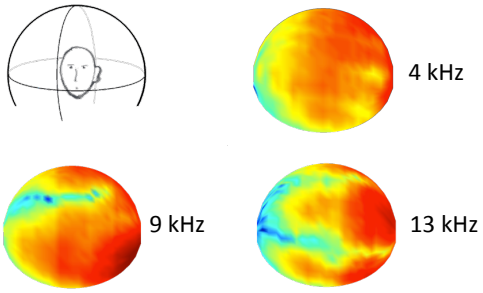
- all temporal and frequency filtering
- from various locations around the listener

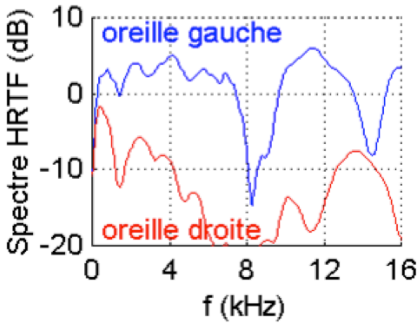
Depends on:

- body, head, ear/pinna geometry
  - => HRTFs are individual



VIDEO



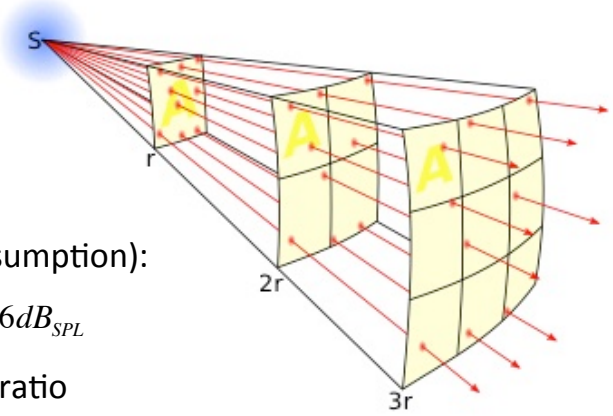


Spatial hearing: distance cues

Localization cues:

direction and distance wise

2) Distance wise



(Non-exhaustive list)

- Sound intensity (free field assumption):

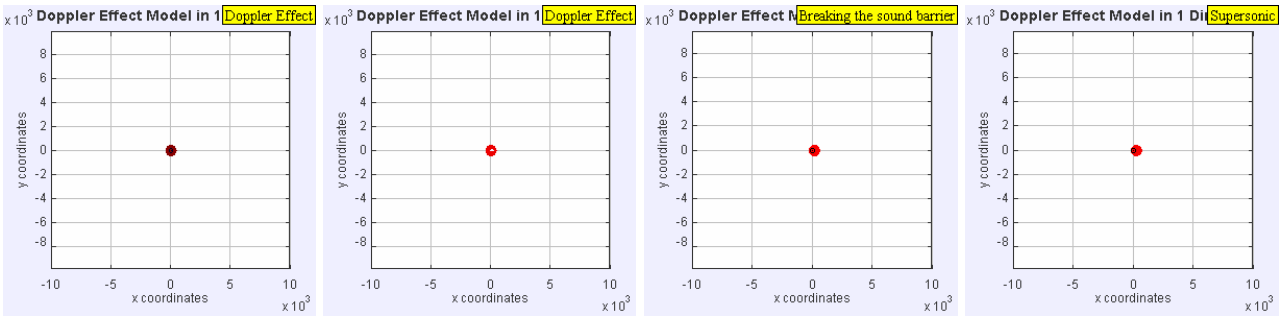
$$I \propto 1/r^2 \Rightarrow \quad r \times 2 \rightarrow \quad -6dB_{SPL}$$

- Direct to reverberant energy ratio
- Attenuation of high frequencies when distance increases
- Attenuation of source details (absence of low intensity sounds)

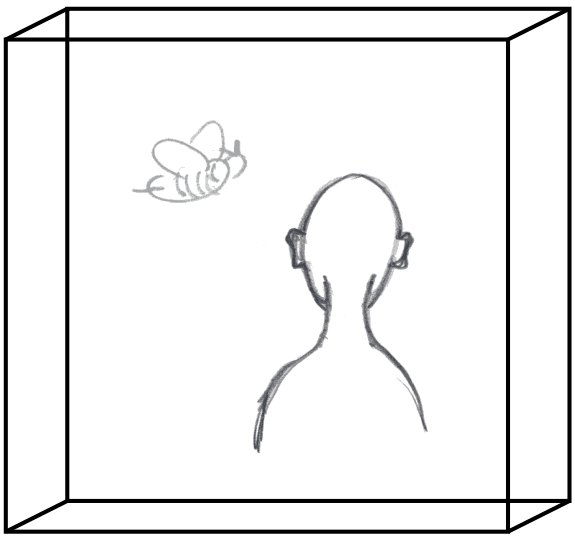
Motion perception: the Doppler Effect

Change in frequency of a wave
for a source moving relative
to an observer

0 km/h < 1250 km/h 1250 km/h > 1250 km/h



The environment



The environment

Audio example: Voice Alarm

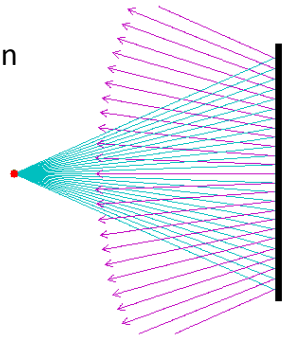
Voice alarm alone (anechoic recording)

Voice alarm in a room (anechoic recording + virtual reverb)

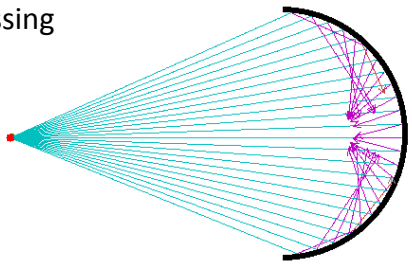
Voice alarm in a room + background noise
(mix: anechoic + reverb + background noise)

Acoustics and surface

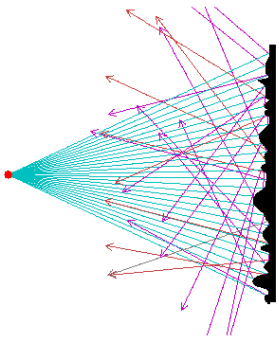
Basic reflection



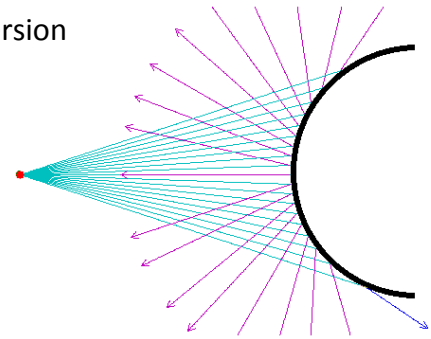
Focussing



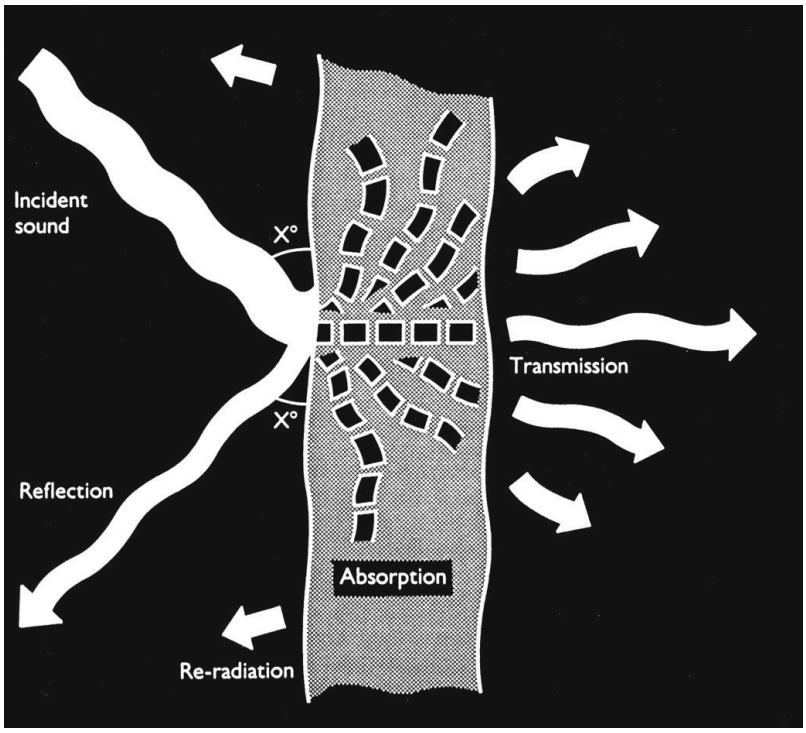
Diffusion



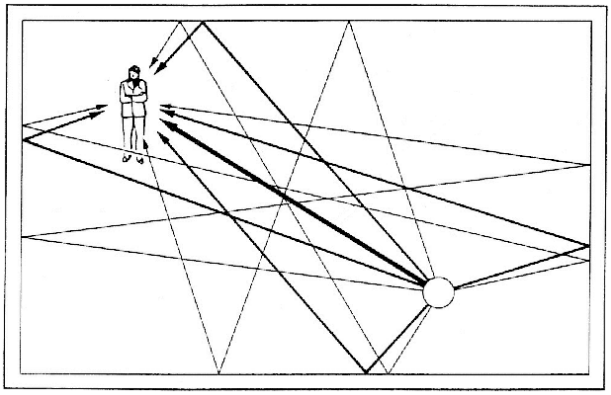
Dispersion

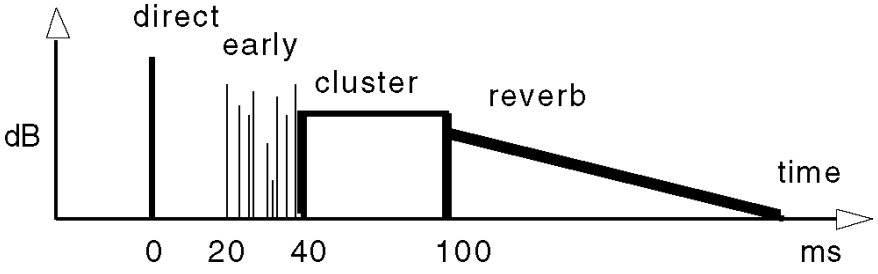


Acoustics and materials

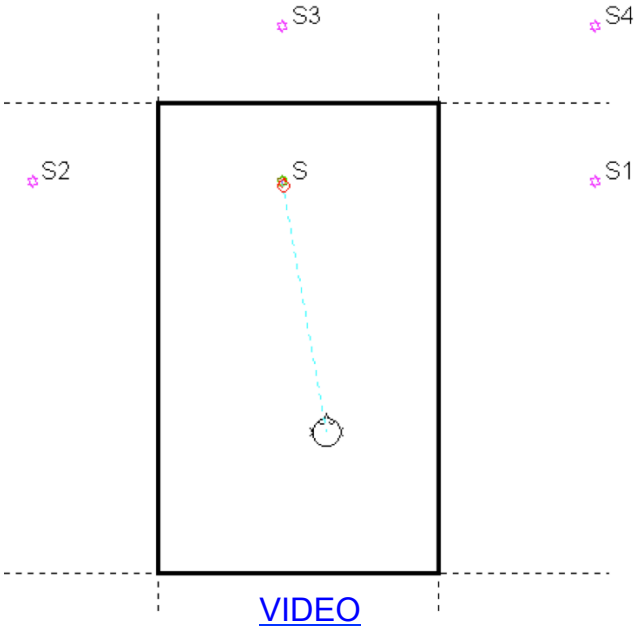


Reverberation time



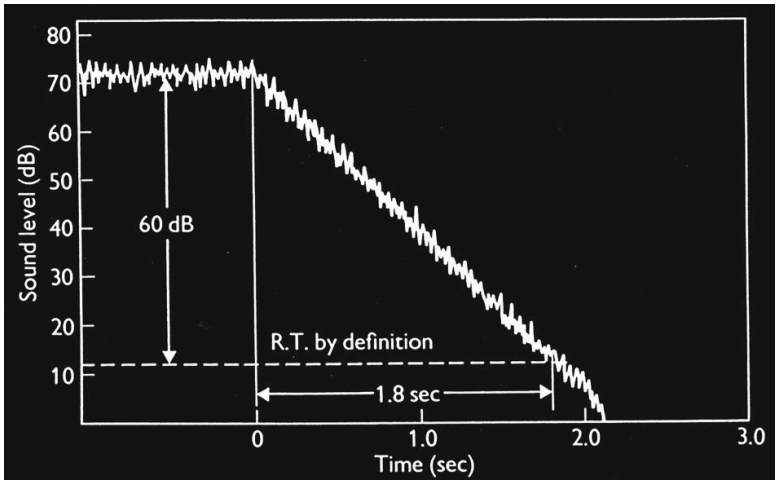


Reverberation time



VIDEO

Reverberation time: examples



Piano

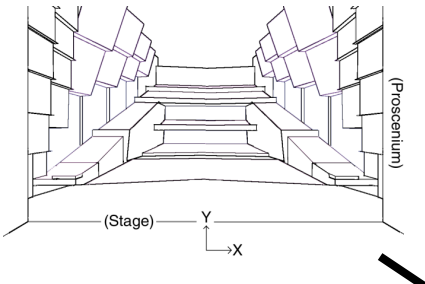
Speech

0 1 3 6 RT in seconds

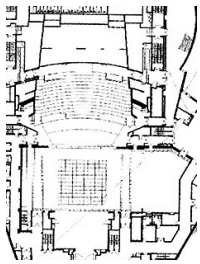
Room geometry effect: flutter echo

Geometric simulation

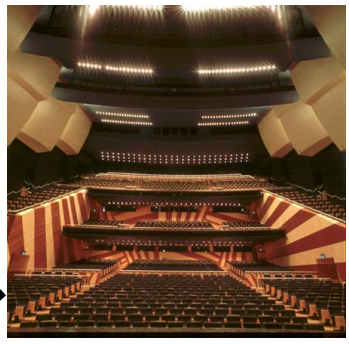
3D acoustical model



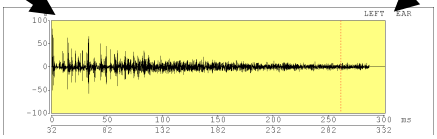
Architectural maps



Real room



Calculated *or* **measured**



Impulse Response

How to?

1) **Create a geometrical model** with the acoustic properties of materials as an approximation of an existing room (or not)

2) Get the 3D **impulse response** to filter the signal

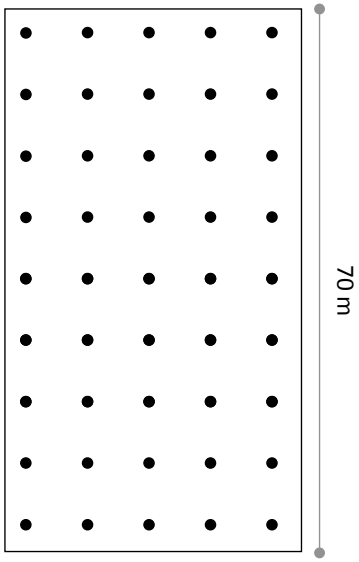
→ High cost processing → Not for real-time

VIDEO

Calculating geometric simulation

Parameters of the room

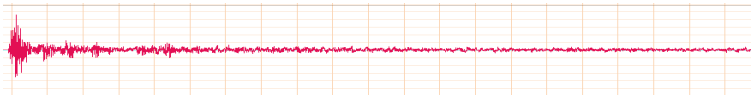
Freq	125	250	500	1000	2000	4000
RT	17.7	14.7	12.5	9.9	6.9	3.9
RT nearfield	-	15.3	11.9	9.0	5.4	3.1
Clarity	-6.0	-6.8	-6.1	-1.3	2.7	4.3



Impulse response of the room

Geometric simulation by convolution

Impulse response of the room



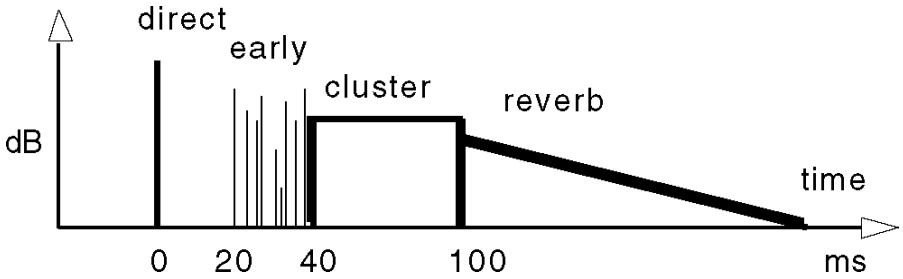
Voice alarm alone (raw)

Voice alarm in the room

Examples of a high hat

Big hall 1 Big hall 2 Big hall 3 Mini Cave

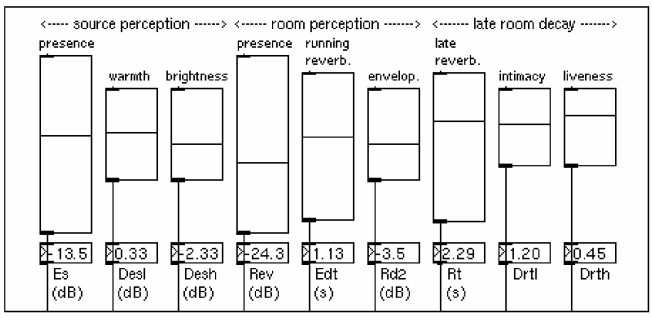
Perceptual simulation



- Direct control of the filter parameters
- Control of perceptual parameters associated to the room effect (presence / intelligibility / reverb …)

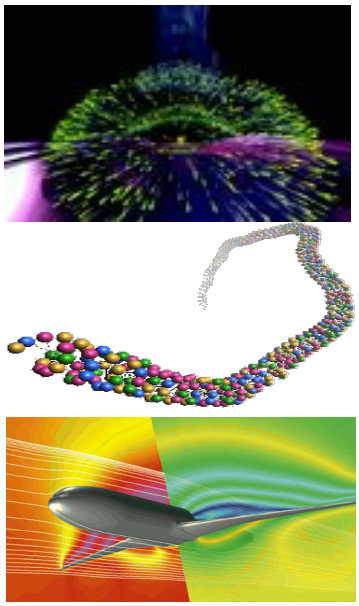
– Less linked to real geometry

+ Faster in terms of processing (real-time possible)



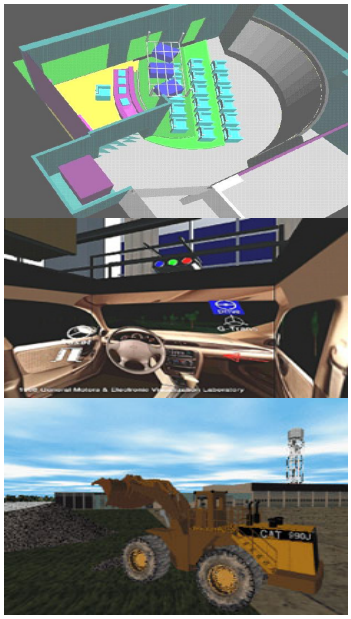
Applications

Complex system

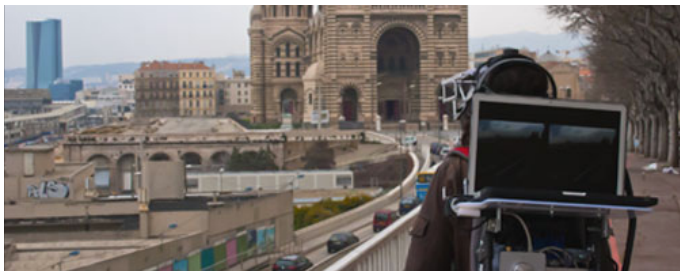


Abstract scenes

Real environment simulation



Auditory augmented reality: links



Art

http://www.a-reality.org/
"Reality is that which, when you stop believing in it, doesn't go away"

Auditory graphs

http://sonify.psych.gatech.edu/research/auditorygraphs/auditory_graphs_explained.html

Augmented aquaria for blind people

http://sonify.psych.gatech.edu/research/aquarium/index.html

3D sound reproduction

Some questions to ask oneself when starting a project:

How many people are to be immersed?

Real-time or not?

Audio only or other modalities? Are there equipment conflicts?

Immersive virtual environment needed?

What are the precision requirements?

Listening and reproduction

- Listening parameters are linked to the reproduction mode
- Choice of the **reproduction system**
  - Stereo
  - Loudspeaker array (several types)
  - Binaural (=> headphones)
- Factors
  - Number of simultaneous listeners / users
  - Number of simultaneous channels
  - Environment to be reproduced
  - Limits of audio equipment
  - Level of details / quality
  - Quality / immersion / intrusion compromise
  - Audio vs. vision compromise

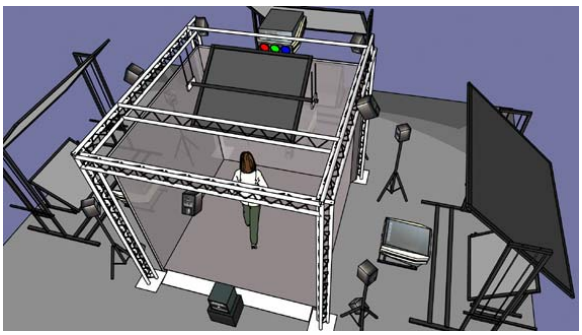
Listening and reproduction

How to choose the reproduction system?

Compromise between:

Quality / immersion / intrusiveness

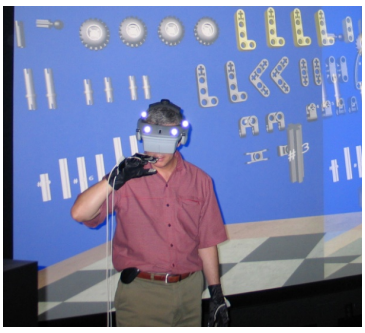
Audio vs. vision



Competition: screen vs. loudspeakers

Conflict: visual / audio

Acoustical control is challenging



Intrusive displays

Strong immersion, but

Restrained interaction

Reproduction: headphones & binaural

Simulates binaural hearing

Separated channels: one per ear

Usually provided by a pair of headphones

Implementing HRTF filters

Problem:

- HRTFs are not universal => variations between subjects
- Elevation discrimination is hard to perceived precisely
- Cone of confusion (esp. front-back) are typical:
  - if HRTFs are **generic** (instead of individualized)
  - if **no** head motion tracking provided

...

Binaural: HRTF

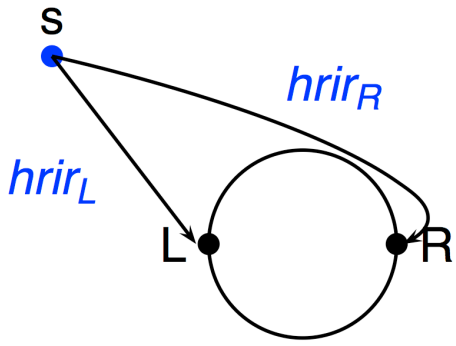
Head-Related Transfer Function

- Transfer functions that describe temporal and frequency filtering from a source in a precise location to both ears

- Describe filtering due to geometry of: **body, head and ear (pinnae)**

- $HRTF_{individual}(\rho,\theta,\varphi,f)$ depends on:
  the individual, the position, and the frequencies

- HRTF Measurement:
  - 2 microphones inside the ears
  - Pure synthetic tones (sine-sweep)
  - At different positions

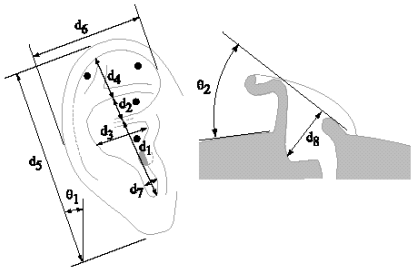


Individualization of HRTF

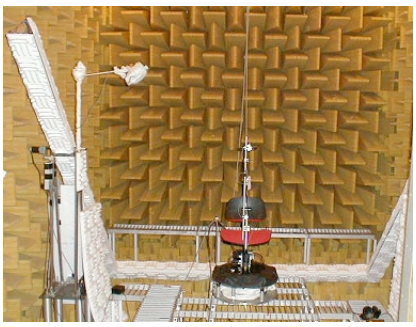
Individual adaptations are required for a good rendering

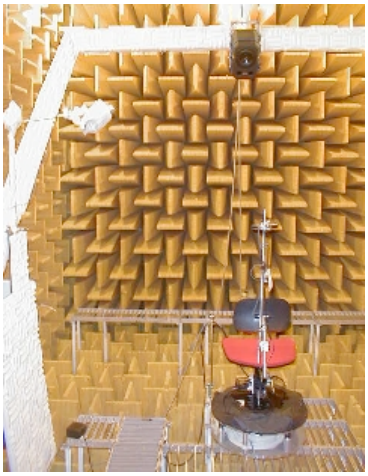
E.g. head size is one of the parameters
(as distance between eyes for stereoscopy)

Some commercial solutions exist: microphones allowing "artificial" binaural recording (Dummy Head)



HRTF measurements



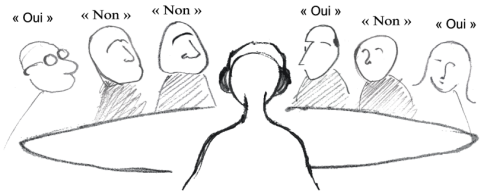


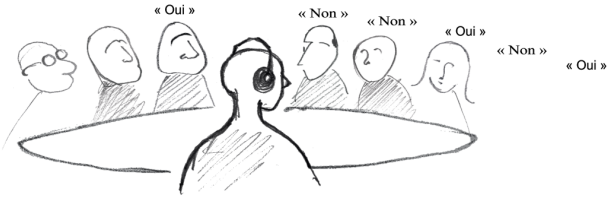
http://recherche.ircam.fr/equipes/salles/listen/system_protocol.html

Immersive binaural

Immersive situation => tracking

- **Head motion**: the sonic scene follows the headphones' orientation
  but we want it to be fixed!





- **Tracking system** in real-time is mandatory in order to fix the virtual audio sources to an absolute position in space
- This increases the **immersion** sensation
- And reduces the degradation due to non-individualized HRTF (front-back confusion, etc.)

Refreshment rates limit the perception of motion speed for the user

Reproduction: stereo pair of loudspeakers

Panning laws

Intensity panning ΔI: level difference between speakers

Phase panning ΔΦ: phase difference between speakers

Delay panning ΔT: onset time difference between speakers

→ Possible to have more width
but listening position is not precise nor stable

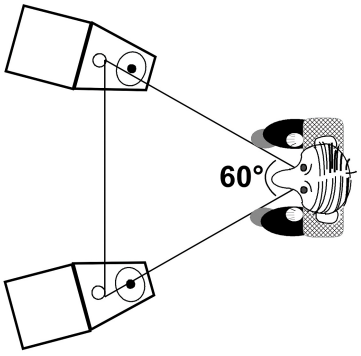
Independent of listener orientation

Optimized for only one source (small *sweet spot*)

Maximum performance in front without elevation:

±30° with **ΔI**

±70-80° **ΔI and ΔT**



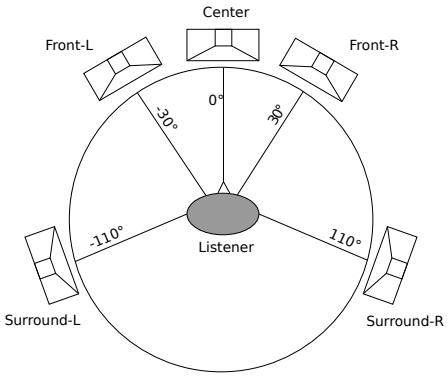
Multichannel panning

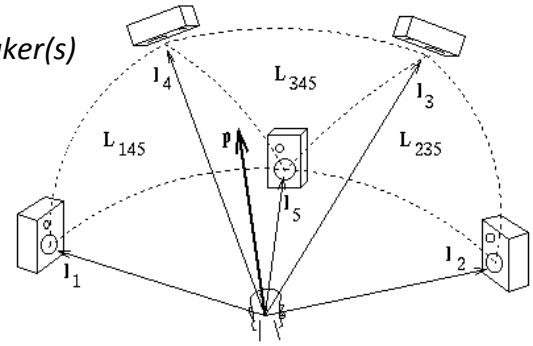
Stereo sound generalization

Multichannel panning laws

- "Surround Sound": 5.1

- Vector Based Amplitude Panning (VBAP)
  - N-channels
  - Source emitted by *the 3 closest speaker(s)* of the virtual position



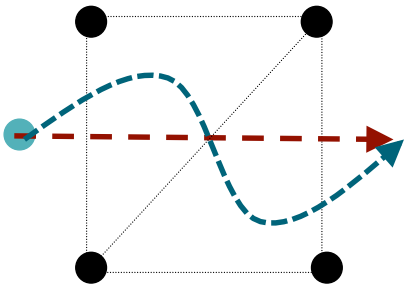


Multichannel panning

Jump between loudspeakers

Accuracy depends on number / density of loudspeakers

N° of loudspeakers => localization accuracy



What we want

What we get

Ambisonics

RECORDING

- 1st order microphone is available since the 70s
- 2nd, 3rd and 4th orders are now also available
- Possible to record real 3D scene for virtual environments

SYNTHESIS

- Possible to create very complex and rich 3D scenes

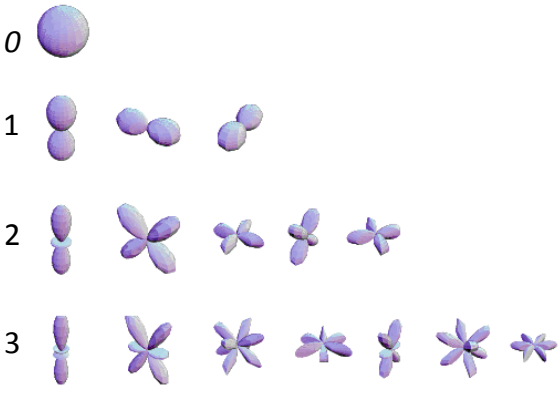
REPRODUCTION

- Listening zone is more stable and larger than with conventional panning techniques
- Multi-users is thus possible
- **Virtual sound object reproduction zone only possible behind the loudspeakers**

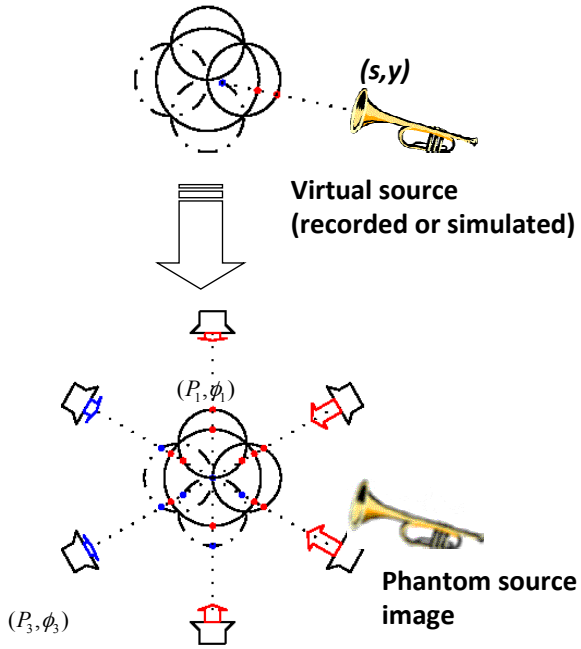
Ambisonics

Encoding and decoding: Spherical **harmonic decomposition**

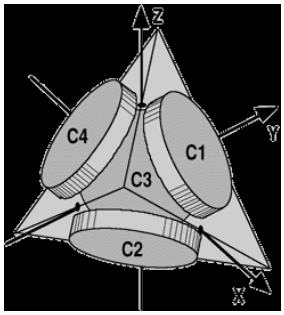
Spherical Harmonic Order

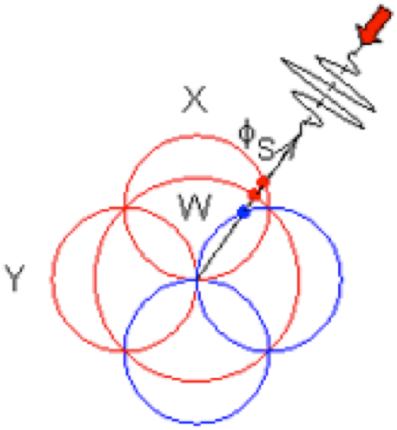


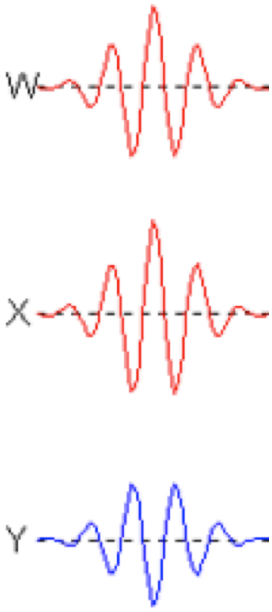
Number of channels { (2*order + 1) in 2D ; (order + 1)² in 3D



Ambisonics







VIDEO

Ambisonics

ENCODING

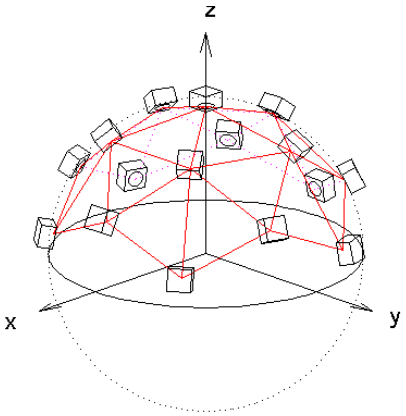
3D (X, Y, Z) via spherical harmonics

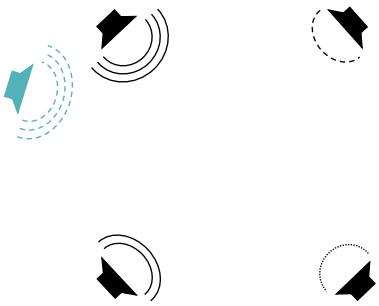
The entire environment (global)

DECODING

3D → N-channels

Depends on the number and position of the loudspeakers





Ambisonics

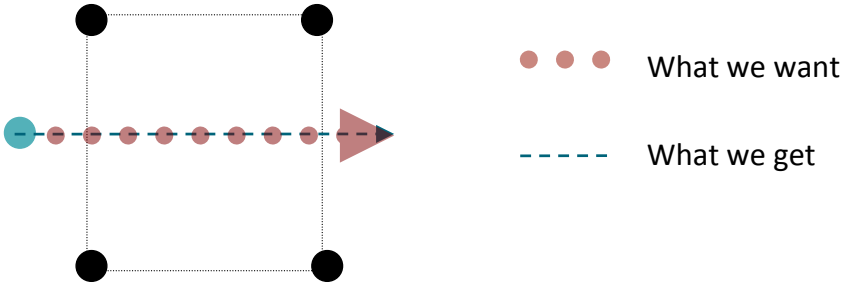
Better impression of space, regular motion between speakers

but

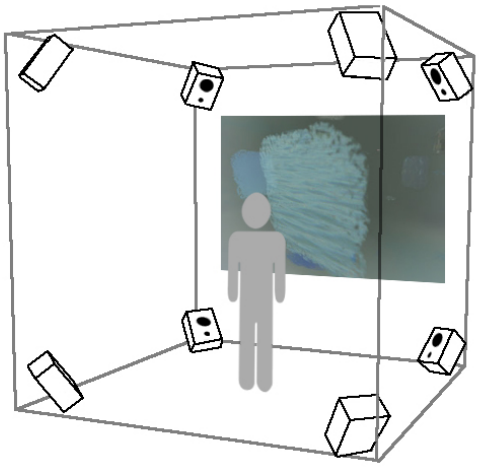
Still inaccurate localization if first order ambisonic system are used

⇒ HOA (Higher Order Ambisonics) are required

- Many channels => expansive microphone solutions
- Hard to keep the system stable, with a good SNR (Signal to Noise Ratio)



Ambisonics: 1st order 3D example



Ambisonic spatialization using a cubic array of 8 loudspeakers

http://cabezal.com/~inguz/cube/

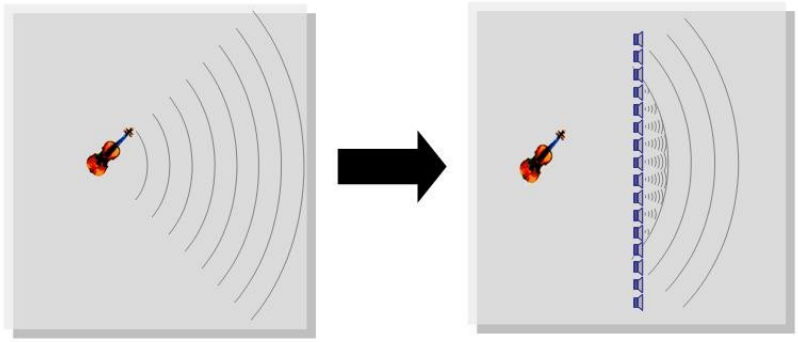
Wave Field Synthesis (WFS)

Holophony (≈ holography)
Wave-Field Syntheses: reproduction of the ‘‘real’’ sound field

Huygens principle:

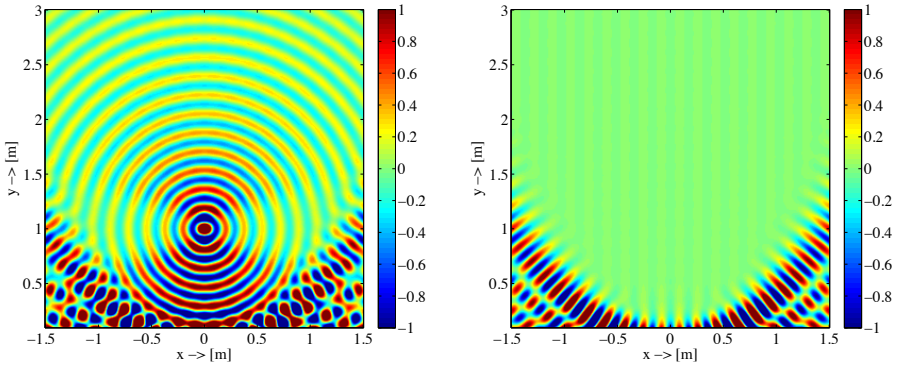
- Any wave front can be regarded as a superposition of elementary spherical waves
- The sum of these secondary waves determines the form of the wave at any subsequent time

It is possible to simulate the acoustic wave produced by a real source with an array of individual loudspeakers (elementary waves) VIDEO



Wave Field Synthesis (WFS)

- Spectral limit occurs according to the distance between loudspeakers: spatial aliasing: 10 or 20 cm are typical => 1.7kHz or 850Hz $\left(f_{\max} = \frac{c}{2\Delta x}\right)$



[Spors 2009]

- Accurate reproduction zone depends on the array width
- Can recreate sound object at the back **and** in front of the array

Wave Field Synthesis (WFS)

Performance

- Large number of speakers: usually not processed for elevation => **2D**
- Multiusers, precise motions, large listening zone

  But high cost processing, esp. for real-time processing => several CPU

Listening zone has a higher spatial fidelity and a bigger sweetspot than for "panning" and ambisonic techniques

The array is like an acoustic "window"

All visible objects in the window are well reproduced

WFS system does not depends on the number and the position of the listeners

To increase the size of the listening zone

it is necessary to extend the speaker array

Wave Field Synthesis (WFS)

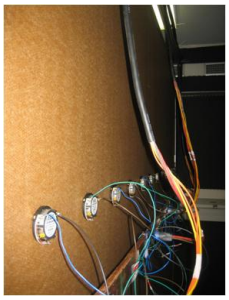
Multi speaker system

IRCAM

Multi Actuator Panel System

LIMSI-CNRS, Sonic Emotion

SMART-I²



Other Examples

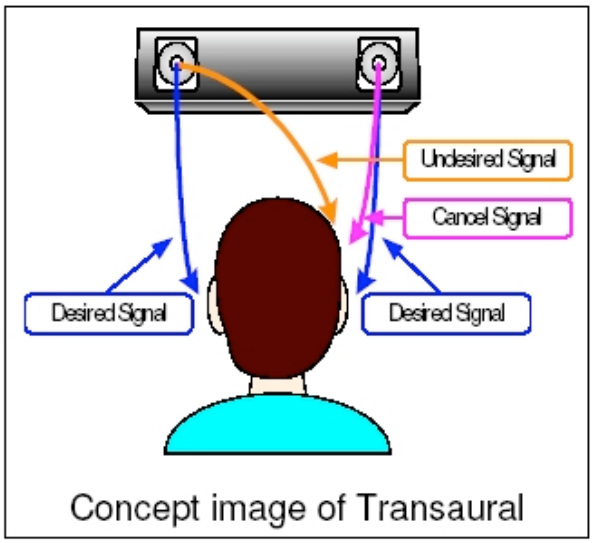
http://www.four-audio.com/en/References/wave-field-synthesis.html

Transaural

Binaural (HRTF) reproduction with loudspeakers

Cross-talk cancellation:

Each ear needs to hear its own channel only



Concept image of Transaural

Comparison: binaural vs. …

Transaural

Comparison: binaural vs. …

Ambisonics

Comparison: binaural vs. …

WFS

Implementation: basis

- Virtual reality system
  - Positions and orientations of sources (X, Y, Z)
  - Position and orientation of the listener (X ,Y , Z)
  - Global environment (volume, geometry, materials)
- Acoustic system
  - Environmental transformation / positions (3D impulse response)
  - Filtering of the source signal with the impulse response
  - 3D signal transformation => N-audio-channels

Binaural implementation

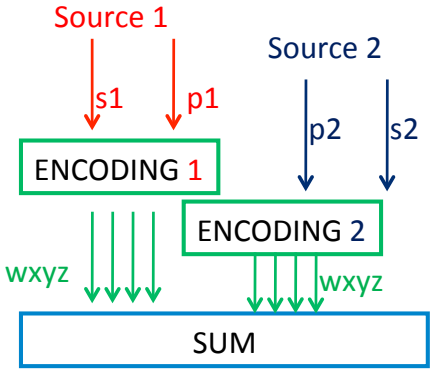
- There is one HRTF associated with one HRIR (H-R Impulse Response) in the time domain, for each specific angle (every 5-15°) for both ears
- The source signal is convolved with the HRIR corresponding to the position (for each ear)
- The obtained signals for each sources, the results of their convolution are added
- If the user is moving, the appropriate HRIR change
  => all source convolutions need to be reprocessed
- If the source is moving, the appropriate HRIR change
  => that source convolution needs to be reprocessed

Rem: reflections and reverberation excluded (no global environment)

Ambisonic implementation

Encoding / B-format conversion

- A position in the scene is taken as a reference
- The source signal is multiplied by spherical harmonics according to the source position from the reference position
- Sum of all encoded sources

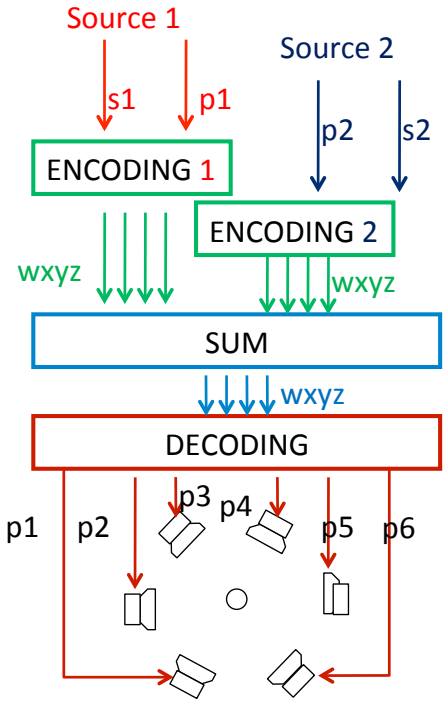


Rem: reflections and reverberation excluded (no global environment)

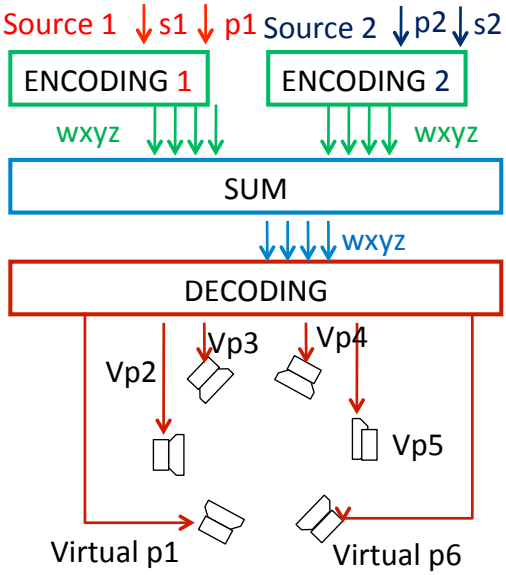
Ambisonic implementation

Decoding

- From the N-channels, spherical harmonics are combined according to loudspeaker positions
- This transformation is fixed if loudspeakers do not move
- If the listener moves, nothing changes
- If the reference position of the source changes, multiplication coefficients change



Virtual Ambisonics



Concept:
After decoding, each output is spatialized at the virtual position of the associated loudspeaker through binaural transformation (HRIR)

Gain:
- HRIR (binaural) convolution processing for $n_{loudspeakers}$ sources
  Instead of $n_{sound\_objects}$ sources
  (interesting if source number >>)
- Sources are fixed (speakers)
- Available on headphones => portable

Summary

	Binaural	Transaural	VBAP	WFS	Ambisonics
Material - configuration	– Headph. + LFE – Headtracking	– 2 / 4 speakers – Headtracking	Sphere of loudspeakers	Large number of very small loudspeakers	Group of loudspeakers
Processing requirement	Complex (HRIR)	Complex (HRIR)	Simple	Complex (many speakers > many channels)	Less Complex (less channels, no user tracking)
Audience size / Personalized soundscape	Large* / Possible	Single* / Possible	Large but limited sweet spot / not possible	Large / not possible, but realistic scape	Large but limited sweet spot
Precision / motion	High precision if personalized HRTF provided	High precision if personalized HRTF provided	Precise but motion is not well reproduced	Highest precision & motion reprod. (horizontal plane)	Relatively precise & good motion reprod. if H.O.A.
Intrusiveness / Space requirement	Intrusive / No particular space required	Not intrusive / Few space required	Not intrusive / Large and fixed arrangement required	Not intrusive / Very Large num. of speakers in a linear array	Not intrusive / Large but flexible arrangement possible
Particularities	HRTF needed	HRTF needed Reprod trough stereo speakers	Only object based	Only object based mix, spatial aliasing	Possibility of decoupling rec. & reproduction

Your summary

Methods / Cues						
Reproduction system type						
Principles						
Listening zone						
Number of listeners						
Accuracy						
Motion						
Intrusion						
Material Costs						
Processing Costs						

3D audio methods: main differences

Based on sound field synthesis (panning, Ambisonics, WFS)	**Binaural**
• Global space • Not very precise (WFS is precise)	• Individual space • Precision (if own HRTF) • Needs to create own HRTF
• Sound objects can **only** be reproduced **behind** the loudspeakers (except WFS)	• Near field sources possible
• Simple to process (except WFS)	• High cost for processing
• Independent of user motion	• Tracking is mandatory => potential delay
• Depends on the number of loudspeakers	• Intrusive (headphones)